SPRINGER NATURE
Reference

Hirak Ranjan Dash
Pankaj Shrivastava
J. A. Lorente
Editors

Handbook of DNA Profiling

Springer

Handbook of DNA Profiling

Hirak Ranjan Dash • Pankaj Shrivastava •
J. A. Lorente
Editors

Handbook of DNA Profiling

With 173 Figures and 95 Tables

Springer

Editors
Hirak Ranjan Dash
School of Forensic Science
National Forensic Sciences University
Delhi Campus
New Delhi, India

Pankaj Shrivastava
DNA Fingerprinting Unit
State Forensic Science Laboratory
Sagar, Madhya Pradesh, India

J. A. Lorente
Laboratory of Genetic Identification
Department of Legal Medicine
Toxicology and Physical Anthropology
Faculty of Medicine
University of Granada
Granada, Spain

GENYO
Centre for Genomics and Oncological Research (Pfizer / University of Granada / Andalusian Regional Government)
Granada, Spain

ISBN 978-981-16-4317-0 ISBN 978-981-16-4318-7 (eBook)
ISBN 978-981-16-4319-4 (print and electronic bundle)
https://doi.org/10.1007/978-981-16-4318-7

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*A tribute to **Dr. Tapaswini Tripathy**, a young dedicated COVID warrior and a great academician who lost her battle of life to COVID in 2021, leaving behind her family and kids in ocean of sorrows.*

Preface

The advent of DNA fingerprinting technology has revolutionized the process of criminal investigation. With the help of DNA report, it has become much easier for the investigating agencies, law practitioners, and honorable judges to conclude/opine in severely complicated cases. The beauty of this technology is that it does not only identify a culprit, but it has the potential to exonerate a wrongly convicted person. The scientific truth of DNA technology corroborating with the criminal justice system has been proven over the years throughout the globe.

Generating DNA report of a few pages requires great expertise, skilled manpower, constant brainstorming in the SOPs, and use of advanced technologies. Due to varied nature of forensic samples, it becomes immensely difficult for DNA examiners to process these samples for generating a DNA profile. In this regard, the technology has seen much advancement over the years from RFLP to DNA phenotyping by using next-generation sequencing (NGS). The adoption of continuously evolving technology helps in processing the challenged samples, generating a DNA profile in much lesser time, as well as increasing the informativeness of a single analysis.

Though law differs with countries, the technology used is the same. Thus, a comprehensive collection of advanced DNA technologies and case works has been included in this volume. This will be immensely useful for technology transfer and knowledge sharing among the examiners and other beneficiaries. Additionally, proper understanding of the technology will be ensured among the students, medical officers, investigating agencies, law practitioners, and honorable judges through real case studies. The edited volume includes the use of DNA fingerprinting technology in solving varieties of criminal as well as civil cases such as paternity dispute, identification of mutilated remains, culprit identification in sexual assault cases, and murder cases. Chapters on non-human studies are also included. The chapters are written by eminent DNA practitioners and academicians around the globe in simple English to help beginners grasp the topic with ease. This will be a huge asset not only for DNA examiners, investigating officers, law practitioners, and honorable judges but also for students of forensic science, genetics, law, and forensic medicine.

New Delhi, India — Hirak Ranjan Dash
Sagar, India — Pankaj Shrivastava
Granada, Spain — J. A. Lorente
April 2022

Contents

Volume 1

About the Editors

Dr. Hirak Ranjan Dash is currently working as an Assistant Professor of Forensic Biology and Biotechnology at the National Forensic Sciences University, Delhi Campus, India. Besides teaching and research, he is actively involved in conducting training programs to various beneficiaries of forensic science. Before joining academics, he served as a forensic DNA expert at Madhya Pradesh Forensic Science Laboratory, India, for over 6 years. He has an experience of conducting DNA examination in more than 1500 complicated cases. He has completed his Ph.D. from the Department of Life Science, National Institute of Technology, Rourkela, India, He received his M.Sc. in Microbiology from Orissa University of Agriculture and Technology, Odisha, India. His research interests include forensic microbiology, thanatomicrobiome analysis, molecular microbiology, DNA fingerprinting, and genetic manipulation. He is one of the pioneers in India to work on NGS technology-based forensic DNA analysis. He has written 8 books and published 50 research papers, 14 book chapters, 12 conference proceedings, and 4 popular science articles.

Dr. Pankaj Shrivastava received his Ph.D. in microbiology with a specialization in biotechnology from the Department of Biological Science, Rani Durgawati University, Jabalpur, Madhya Pradesh, India. He is in charge of the Forensic DNA Fingerprinting facility of the Govt. of Madhya Pradesh, India at Sagar. Dr. Shrivastava has more than 13 years of experience as a bench worker for examining and reporting a wide range of criminal cases using DNA technology and deposing the court evidence. His research interests include population DNA database, improvement of

methods in forensic DNA typing, finding rapid protocols for the technology, and microbial forensics. He is a two-time recipient of the prestigious Pt. Govind Vallabh Pant Award from the Bureau of Police Research and Development, Ministry of Home Affairs, Govt. of India, and a recipient of the Anusrajan Award from AIISECT University, Bhopal, for his authored books. Dr. Shrivastava is also a recipient of the FICCI Smart Policing award for developing a fast DNA-typing protocol. He is a visiting faculty at many universities, police training institutes, and judicial officers training institutes. Dr. Shrivastava has written 13 books and published 65 research papers in journals of repute in forensic science, 21 book chapters, 84 conference proceedings, and 15 popular science articles.

J. A. Lorente is Professor of Forensic Medicine at the University of Granada, Spain. After graduating from the Faculty of Medicine of Granada in 1985, he completed his Ph.D. (Honors) in medicine and surgery at the University of Granada, in 1989. Dr. Lorente has published over 170 peer-reviewed papers and several books and book chapters. He has a special focus on the use of DNA and its application to human rights (he launched the first-ever database to identify missing people, the Spanish Phoenix Program, back in 1999); he also created and launched the DNA-PROKIDS Program in 2004, and the DNA-Pro-ORGAN Program in 2016. His areas of interest in forensics also deal with population variability and analysis of old and ancient DNA samples and databases expansion and control. Dr. Lorente is also actively working in medical genomics, and he is the scientific director of the Center for Genomics and Oncological Research (GENYO), where his team focuses on liquid biopsy and cancer interception. Dr. Lorente is the founder and first president of the AICEF (Ibero-Latin American Network of Forensic Sciences) and an honorary member of the AFSN (Asian Forensic Sciences Network).

Contributors

Radhika Agarwal DNA Fingerprinting Unit, Forensic Science Laboratory, Bhopal, Madhya Pradesh, India

Fatima J. AlBuarki Forensic Science Laboratory, Directorate of Forensic Science, Ministry of Interior, Manama, Kingdom of Bahrain

Arash Alipour Tabrizi Legal Medicine Research Center, Legal Medicine Organization, Tehran, Iran

Noora R. Al-Snan Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Kingdom of Bahrain, Manama, Bahrain

Juan Carlos Alvarez Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

M. J. Alvarez-Cubero Department of Biochemistry and Molecular Biology III, Faculty of Medicine, University of Granada, Granada, Spain

Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), Granada, Spain

Antonio Amorim Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

Institute for Research and Innovation in Health (i3S), University of Porto, Porto, Portugal

Faculty of Sciences of the University of Porto (FCUP), Porto, Portugal

Eduardo Arroyo-Pardo Laboratory of Forensic and Population Genetics, Legal Medicine Psychiatry and Pathology Department, Medicine School of the Complutense University of Madrid, Madrid, Spain

Forensic Sciences: Forensic Genetics and Toxicology Group, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos IdISSC, Madrid, Spain

Saranya Ramesh Babu Department of Forensic Sciences, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia

Ashish Badiye Department of Forensic Science, Government Institute of Forensic Science, Nagpur, Maharashtra, India

Anna Barbaro Department of Forensic Genetics, Studio Indagini Mediche E Forensi (SIMEF), President Worldwide Association of Women Forensic Experts (WAWFE), Reggio Calabria, Italy

Akanksha Behl Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, Uttar Pradesh, India

Mark Benecke Forensic Biology Unit, International Forensic Research and Consulting, Cologne, Germany

Hannah E. Berry TU Human Remains Identification Lab (THRIL), Department of Chemistry, Towson University, Towson, MD, USA

Monika Chakravarty Forensic Science Laboratory, Home Department Government of NCT of Delhi, Delhi, India

Rajendra V. E. Chilukuri SyberGENE Inc, San Jose, CA, USA

Sumit Choudhary School of Forensic Science and Risk Management, National Security and Police University of India, Rashtriya Raksha University, Gandhinagar, Gujarat, India

Pawan Kumar Chouhan Regional Forensic Science Laboratory, Indore, Madhya Pradesh, India

Beatrice Corradini Department of Biomedical, Metabolic and Neural Sciences, Institute of Legal Medicine, University of Modena and Reggio Emilia, Modena, Italy

Heitor S. D. Correa Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, Forensic Medicine Unit, University of Brescia, Brescia, Italy

Venusia Cortellini Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, Forensic Medicine Unit, University of Brescia, Brescia, Italy

E. D'Orio Bio Forensics Research Center, Angri, Italy

Natalie Damaso Biological and Chemical Technologies, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA, USA

Hirak Ranjan Dash School of Forensic Science, National Forensic Sciences University, Delhi Campus, New Delhi, India

M. C. A. De Ungria DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines Diliman, Quezon, Philippines

Program on Biodiversity, Ethnicity, and Forensics, Philippine Genome Center, University of the Philippines Diliman, Quezon, Philippines

Thilaga Dhanapal Forensic Sciences Department, DNA Division, Chennai, India

Mayank Kumar Dubey Senior Faculty-Forensic Science Department, Mody University, Laxmangarh, Rajasthan, India

Kelly M. Elkins TU Human Remains Identification Lab (THRIL), Department of Chemistry, Towson University, Towson, MD, USA

Francesca Ferrari Department of Biomedical, Metabolic and Neural Sciences, Institute of Legal Medicine, University of Modena and Reggio Emilia, Modena, Italy

Gianmarco Ferri Department of Biomedical, Metabolic and Neural Sciences, Institute of Legal Medicine, University of Modena and Reggio Emilia, Modena, Italy

Lorenzo Franceschetti Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, Forensic Medicine Unit, University of Brescia, Brescia, Italy

G. Francione Bio Forensics Research Center, Angri, Italy

X. Gálvez Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

Anubha Gang Regional Forensic Science Laboratory, Indore, MP, India

Cláudia Gomes Laboratory of Forensic and Population Genetics, Legal Medicine Psychiatry and Pathology Department, Medicine School of the Complutense University of Madrid, Madrid, Spain

Forensic Sciences: Forensic Genetics and Toxicology Group, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos IdISSC, Madrid, Spain

Ulhas Gondhali Jindal Institute of Behavioural Sciences, O.P. Jindal Global University, Sonipat, India

Christian Haarkötter Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

Samuel D. Hodge Temple University, Philadelphia, PA, USA

E. B. Jimenez National Institutes of Health, University of the Philippines Manila, Manila, Philippines

Jaskirandeep Kaur Jossan Department of Forensic Science, Punjabi University, Patiala, Punjab, India

Neeti Kapoor Department of Forensic Science, Government Institute of Forensic Science, Nagpur, Maharashtra, India

Dan E. Krane Biological Sciences, Wright State University, Dayton, OH, USA

Naresh Kumar Home Department, Government of NCT of Delhi, New Delhi, India

Sachil Kumar Department of Forensic Sciences, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia

Vikash Kumar Centurion University of Technology and Management, Bhubaneswar, Odisha, India

R. K. Kumawat DNA Division, State Forensic Science Laboratory, Jaipur, India

Pushpesh Kushwaha Dr. APJ Abdul Kalam Institute of Forensic Science and Criminology, Bundelkhand University, Jhansi, India

DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home (Police), Government of Madhya Pradesh, Sagar, India

Pietro Liò Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

Ana María López-Parra Laboratory of Forensic and Population Genetics, Legal Medicine, Psychiatry and Pathology Department, Medicine School, Complutense University of Madrid (UCM), Madrid, Spain

Grupo de Ciencias Forenses: Genética y Toxicología forenses, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain

J. A. Lorente Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

GENYO, Centre for Genomics and Oncological Research (Pfizer / University of Granada / Andalusian Regional Government), Granada, Spain

B. Lorente-Remon Program International Liaison Officer (ILO), Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, DNA-PROKIDS, Granada, Spain

Donato Madalese Histocompatibility and Immunogenetics Laboratory, UOSD Cryopreservation and Ba.S.C.O, Oncohaematology Department, Pausilipon Hospital, A.O.R.N. Santobono-Pausilipon, Naples, Italy

Nithyanandam Mahalaxmi Forensic Sciences Department, Chennai, Tamil Nadu, India

Uthandaraman Mahalinga Raja Forensic Sciences Department, DNA Division, Chennai, India

Giovanna Maisto Flow Cytometry Laboratory, UOSD Cryopreservation and Ba.S. C.O, Oncohaematology Department, Pausilipon Hospital, A.O.R.N. Santobono-Pausilipon, Naples, Italy

M. Mangione Bio Forensics Research Center, Angri, Italy

L. J. Martínez-González Department of Biochemistry and Molecular Biology III, Faculty of Medicine, University of Granada, Granada, Spain

Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), Granada, Spain

M. I. Medina-Lozano Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

John Meehan Beasley School of Law at Temple University, Philadelphia, PA, USA

Kelly A. Meiklejohn Department of Population Health and Pathobiology, North Carolina State University, Raleigh, NC, USA

Heather Miller Coyle Forensic Science Department, Henry C. Lee College of Criminal Justice and Forensic Sciences, University of New Haven, West Haven, CT, USA

Amarnath Mishra Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, Uttar Pradesh, India

Indresh Kumar Mishra Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, Uttar Pradesh, India

Forensic Science Laboratory, Rohini, Government of NCT, Delhi, India

Aditi Mishra School of Forensic Science and Risk Management, National Security and Police University of India, Rashtriya Raksha University, Gandhinagar, Gujarat, India

Alka Mishra Regional Forensic Science Laboratory, Gwalior, Madhya Pradesh, India

P. Montagna Bio Forensics Research Center, Angri, Italy

Iman Muharam HID Professional Services, Genetic Sciences Division, Thermo Fisher Scientific, Scoresby, VIC, Australia

Usharani Munuswamy Forensic Sciences Department, DNA Division, Chennai, India

D. S. Negi Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India

Mahalakshmi Nithyanandam Forensic Sciences Department, DNA Division, Chennai, India

Robert O'Brien National Forensic Science Technology Center, Largo, FL, USA

Carla Paintner HID Professional Services, Genetic Sciences Division, Thermo Fisher Scientific, Singapore, Singapore

Sara Palomo-Díez Laboratory of Forensic and Population Genetics, Legal Medicine, Psychiatry and Pathology Department, Medicine School, Complutense University of Madrid (UCM), Madrid, Spain

Grupo de Ciencias Forenses: Genética y Toxicología forenses, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain

International University of La Rioja (UNIR), Logroño, Spain

Prateek Pandya Amity Institute of Forensic Sciences, Amity University, Uttar Pradesh, India

Inusha Panigrahi Genetic Metabolic Unit, Department of Pediatrics, APC, PGIMER, Chandigarh, India

Roberta Penta de Vera d'Aragona Histocompatibility and Immunogenetics Laboratory, UOSD Cryopreservation and Ba.S.C.O, Oncohaematology Department, Pausilipon Hospital, A.O.R.N. Santobono-Pausilipon, Naples, Italy

M. Katherine Philpott Dept. of Forensic Science, Virginia Commonwealth University, Richmond, VA, USA

Elena Pilli Department of Biology, University of Florence, Laboratory of Anthropology - Forensic Molecular Anthropology Unit, Florence, Italy

Nadia Pinto Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

Institute for Research and Innovation in Health (i3S), University of Porto, Porto, Portugal

Centre of Mathematics of the University of Porto, Porto, Portugal

Cristoforo Pomara Department of Medical, Surgical and Advanced Technologies "G.F. Ingrassia", University of Catania, Catania, Italy

Pooja Puri Amity Institute of Forensic Sciences, Amity University, Noida, India

Avinash Chand Puri Regional Forensic Science Laboratory, Indore, Madhya Pradesh, India

Rajshree Raghunath Forensic Sciences Department, DNA Division, Chennai, India

Manisha Rana Department of Forensic Science, Guru Ghasidas University, Bilaspur, India

DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home (Police), Government of Madhya Pradesh, Sagar, India

Kashiya R. Reese TU Human Remains Identification Lab (THRIL), Department of Chemistry, Towson University, Towson, MD, USA

James M. Robertson Research Support Unit, Federal Bureau of Investigation Laboratory Division, Quantico, VA, USA

Shipra Rohatgi Former Faculty/ Research Scholar, Amity Institute of Forensic Science, Amity University Gurugram/Noida, Noida, India

Subhasish Sahoo HOD DNA Division, State Forensic Science Laboratory, Bhubaneswar, Odisha, India

María Saiz Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

Monica Salerno Department of Medical, Surgical and Advanced Technologies "G.F. Ingrassia", University of Catania, Catania, Italy

Rashmita Samal DNA Division, State Forensic Science Lab, Rasulgarh, Bhubaneswar, India

Samreen S. Sayed Forensic Science Laboratory, Directorate of Forensic Science, Ministry of Interior, Manama, Kingdom of Bahrain

Monica Scali Department of Life Science, University of Siena, SIENA, Italy

Federica Schiano di Tunnariello Manipulation and Cryopreservation Laboratory, UOSD Cryopreservation and Ba.S.C.O, Oncohaematology Department, Pausilipon Hospital, A.O.R.N. Santobono-Pausilipon, Naples, Italy

Andrei Semikhodskii Medical Genomics Ltd., Tver, Russian Federation

Francesco Sessa Department of Clinical and Experimental Medicine, University of Foggia, Foggia, Italy

Vivek Kumar Shrivastav Regional Forensic Science Laboratory, Indore, MP, India

Pankaj Shrivastava DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

Ritesh Kumar Shukla Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, Gujarat, India

Enrico Silingardi Department of Biomedical, Metabolic and Neural Sciences, Institute of Legal Medicine, University of Modena and Reggio Emilia, Modena, Italy

James Simpson Victoria University of Wellington, Tirohanga, New Zealand

Rajinder Singh Department of Forensic Science, Punjabi University, Patiala, Punjab, India

Sourabh Kumar Singh Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, India

J. H. Smith Forensic Services of the South African Police Service, Pretoria, South Africa

E. V. Soniya Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala, India

Priyanka Srivastava Genetic Metabolic Unit, Department of Pediatrics, APC, PGIMER, Chandigarh, India

Pradnya Sulke Government Institute of Forensic Science, Nagpur, Maharashtra, India

U. Suresh Kumar Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala, India

Parag M. Tamhankar Centre for Medical Genetics, Mumbai, India

Vasundhara P. Tamhankar Centre for Medical Genetics, Mumbai, India

Tugba Unsal Sapan Institute of Addiction and Forensic Sciences, Uskudar University, Istanbul, Turkey

Kamayani Vajpayee DNA Fingerprinting Unit, Forensic Science Laboratory, Bhopal, Madhya Pradesh, India

School of Arts and Sciences, Ahmedabad University, Ahmedabad, India

Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, Gujarat, India

Lakshmi Vasudevan Centre for Medical Genetics, Mumbai, India

Andrea Verzeletti Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, Forensic Medicine Unit, University of Brescia, Brescia, Italy

Rita Vignani Department of Life Science, University of Siena, SIENA, Italy

Serge-genomics, SIENA, Italy

Part I

Principles of Forensic DNA Profiling

Forensic DNA Investigation

1

J. H. Smith

Contents

Abstract

Forensic DNA evidence is assisting in resolving crime and is now commonly used in most countries. It assists in exonerating the innocent, identification of perpetrators of crime, establishes paternity, and identification of human remains. When processing and testing exhibit material, the crime scene and laboratory examiners must be mindful of the evidential value of DNA physical evidence. There has been no other technology that has been so voraciously scrutinized for

J. H. Smith (✉)
Forensic Services of the South African Police Service, Pretoria, South Africa
e-mail: smithjoe@saps.gov.za

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_57

acceptance in the courts as forensic DNA evidence. Since the introduction of forensic DNA evidence to the courts, there have been many reported incidences where errors have occurred during the collecting and testing phase of DNA exhibit material and the interpreting of the results. The use of forensic DNA evidence is based on reliable technology: provided that the necessary quality assurance and quality control measures are consistently applied and the value of forensic DNA findings is not overstated. Forensic databases are assisting investigators in the investigation of cases where serial and multiple offenders are involved. Verified forensic DNA investigative lead reports that are generated from the comparison searches conducted on the forensic DNA databases can either link a known person on the database to crime scenes or link various crime scenes where the person's forensic DNA profiles are not yet uploaded to the database. Cross-forensic investigative lead products can also be used to check for certain common attributes of value that is shared between different types of forensic investigative leads to aid with the resolution of linked cases.

Keywords

Forensic DNA · Crime scene · Exhibit material · DNA databases · Short term repeat · DNA findings · Erroneous DNA findings · Evidential value · Probative value · Value of DNA Evidence · Physical evidence · Forensic investigative leads · Forensic DNA investigative leads · Cross-forensic investigative leads · Erroneous forensic DNA findings · Best practices using forensic investigative leads as effective · Serial murder cases · Serial rape cases · Modus operandi

Abbreviations

DNA	Deoxyribonucleic acids
IBIS	Integrated ballistic identification system
IMEI	The unique International Mobile Equipment Identity number on cell phones
ISO	International Organization for Standardization
SMS	Small message service available on cell phones
STR	Short tandem repeat

Definitions

- **Competence:** the demonstrated ability to apply knowledge and skills.
- **Chain of Custody:** a physical log(s) for a single piece of evidence material that documents who had possession and when the evidential material was in his or her possession during the different stages of processing the evidential material (e.g., from the collection, storage, transport, receipt, and analysis phases and the disposition thereof and delivered to court).

- **Defendant's fallacy:** the fallacy by arguing that because many people in the population may share the forensic DNA profile, thus the DNA evidence is worthless; or arguing that if a profile frequency is reported as 1 in 1 million, then it is expected that the number of people in the population with a matching profile is 100 and the probability that the defendant is the culprit is only 1 in 100 (1%).
- **Low copy number (LCN):** particularly for current short tandem repeat (STR) typing, refers to the analysis of any sample that contains less than 200 pg. of template DNA.
- **LCN typing:** simply can be defined as the analysis of any DNA sample where the results are below the stochastic threshold for reliable interpretation.
- **Modus Operandi:** the style or method that a perpetrator uses in the committing of his or her crime.
- **Prosecutor's fallacy:** the statistical reasoning typically used by a prosecutor overstates DNA evidence, for example, match probability and the probability that the defendant is innocent is confused. The prosecutor uses the match probability and states that there is 1 in x probability/chance or occurrence that the genetic material (DNA) is from the defendant. The "prosecutor's fallacy" has compromised the use of DNA evidence for a fair trial. This fallacy suggests that the rarity of a profile is interchangeable with the probability that the defendant is innocent (e.g., the rarity of a one in a million match produces the false conclusion that the chance of the defendant being innocent is one in a million).
- **Stochastic effects:** the observation of intra-locus peak imbalance and/or allele dropout resulting from random, disproportionate amplification of alleles in low-quantity template analytical thresholds should establish a balance between allele preservation and artifact editing, resulting in confident, accurate allele calls.
- **Restriction enzyme:** an enzyme with endonuclease activity that cuts DNA at specific sequences.
- **Restriction fragment length polymorphism (RFLP):** the detection of length or sequence variation after cleaving DNA with a restriction enzyme and detection using a small section of DNA within the sequence between two restriction sites. RFLP testing was in the detection of minisatellites.
- **Short tandem repeat (STR):** polymorphic region of DNA where alleles differ in the number of tandemly arranged core repeats. STR allele typically ranges in size between 100 bp and 400 bp. Also known as microsatellites.
- **Southern blot:** a technique that is used to transfer DNA from a gel onto a nylon membrane.
- **Variable number tandem repeat (VNTR):** polymorphic region of DNA where alleles differ in the number around 500 bo to over 20 kb. Also known as minisatellites.
- **Forensic investigative leads:** Verified outcome linking different crime scenes or linking a person to a crime scene after performing a comparison search on a forensic database such as DNA database or a fingerprint database or a ballistic biometric database.

Introduction

Edmond Locard (1877–1966) is considered to be one of the main persons that laid the foundations for the development of modern forensic science. He is well-known for his "Locard Exchange Principle" that is based on the fact that there is an interaction between different physical entities when they come into contact. He made the inference "that every contact leaves a trace value" (Saferstein 2011).

There is a good likelihood that the genetic material (DNA) originating from a perpetrator will be left behind on the crime scene due to him or her coming into direct contact touch with an object at the crime scene Physical evidence, such as genetic material (DNA), found at crime scenes has proven to be a powerful tool in the fight against crime (Goodwin et al. 2011). The genetic material (DNA) can identify perpetrators, convict the guilty, and exonerate the innocent (Meintjes-Van der Walt 2010).

Forensic DNA profiling has undergone significant development since the first case where it was applied as an investigative tool to assist with the identification and conviction of a murderer and rapist (Butler 2010). Initially, it was only possible to obtain forensic DNA from large visible stains such as blood or semen. Today, forensic DNA profiles are derived from trace invisible amounts of DNA, such as skin cells left from touch transfer (Van Oorschot and Jones 1997). The early DNA profiling techniques were labor intensive and time-consuming compared to the available techniques deployed in the modern forensic laboratory.

The forensic multiplex chemistries used in the last decade have seen a significant increase in the number of markers deployed to provide more discrimination power in forensic DNA profile identification, to facilitate international sharing of forensic DNA profiles, and to minimize incidental outcomes from comparison searches on forensic DNA databases (Li 2015). The commercial kits available today utilize multiplex analysis of multi-allelic short tandem (STR) markers located on different chromosomes for high power of discrimination, and many countries are routinely using STRs with at least 20 loci (Goodwin et al. 2011) Additional DNA techniques that can be useful under the appropriate circumstances have been developed, such as analysis of mitochondrial DNA and the Y-chromosome (Butler et al. 2002), as well as low copy number (LCN) DNA testing (Goodwin et al. 2011).

The main questions that a forensic DNA examiner is asked to address are (Saferstein 2011; Manamela et al. 2015):

- Who left the genetic material (DNA) at the crime scene?
- From what body fluid did the genetic material originate from?
- How was the genetic material deposited?
- Has the necessary quality assurance and quality control been applied during the processing and analyses of the exhibit material containing the genetic material?
- Have the results been objectively and fairly interpreted and reported?

The interpretation of a single-source forensic DNA profile obtained from exhibit material is simple and can provide powerful scientific evidence either to exclude or

to include any person as a possible source of that genetic material (DNA) (Balding 2005; Goodwin et al. 2011).

When exhibit material contains genetic material (DNA) from more than one contributing person, the interpretation of who is the source of this DNA can become difficult or even impossible to attribute to with confidence (Balding 2005; Goodwin et al. 2011).

Technological improvements in forensic DNA analysis have resulted in the ability to analyze even smaller quantities of DNA with increased discrimination power. The advent in the establishment of forensic DNA databases for comparison searching has assisted forensic examiners in collaboration with law enforcement to resolve many casework linking serial (see Fig. 1: Example of DNA conviction and mass screening case) and multiple offenders (Meintjes-Van der Walt 2010; Smith 2020).

Value of Forensic DNA Investigations

Today, some wrongly convicted people have been exonerated because of forensic DNA evidence (Gould et al. 2013). Furthermore, in casework, persons are excluded routinely. Forensic DNA has proved to be a powerful investigative tool in:

- Solving crime where evidence is matched with a perpetrator
- In deciding who the father in paternity testing is
- Mass disaster identification of victims in natural disasters (such as the 2004 Indian Ocean Tsunami that killed about 225,000 people)
- Terrorism acts (e.g., almost 3000 people were killed in the attack on the World Trade Center on September 11, 2001)
- Identification of unidentified remains in structural building collapses (e.g., on September 12, 2014, a guesthouse located within a church premise in Nigeria completely collapsed to the ground and killing 115 persons)
- Identification of the “unknown” soldier
- Helping in working out inheritance claims

The physical evidence will not be accepted in court if the exhibit integrity is compromised due to a shortcoming in one of the aforementioned aspects.

Biological contamination is the compromise of the integrity of the exhibit where the unwanted transfer of genetic material (DNA) from an unknown source to a piece of physical evidence. In the process, the original condition of the exhibit material is changed and therefore also the future analysis result to be obtained. The adding or removal of extraneous material from the exhibit material (i.e., contamination) is usually unknown to the investigator or forensic examiner. Therefore, the influence on the analysis result will be unknown as well. Measures to prevent contamination must be in place when processing a crime scene and when handling exhibits. Contamination can be prevented by following basic contamination prevention steps. For example, the wearing of personal protective

clothing and the correct use of packaging material for exhibit material collected are critical contamination preventive measures. The consequences, if contamination does occur with exhibit material, will negatively impact on the resolution of the case in question.

In 1987 Colin Pitchfork was the first person to be convicted using forensic DNA profiling. It was also the first case where mass screening identification was applied to a population to assist identification in a criminal investigation.

On 21st November 1983, a 15-year-old Lynda Mann's body was found on a remote footpath after she did not return from visiting her friend. The post-mortem revealed that she had been sexually assaulted and strangled. A vaginal sample was retrieved from her body. Three years later, another 15-year-old girl, Dawn Ashworth's body was found in a wooded area. She was also been raped and strangled.

Richard Buckland, a seventeen-year-old who had a learning disability was arrested for the murder of Ashworth. He had some knowledge of the crime that was not in the public domain. Whilst being questioned, Buckland would confess to the crime, but he later withdrew his confession. He, however, would consistently deny that he committed the first murder.

Dr. Peter Gill from the Forensic Science Service in the United Kingdom developed a preferential extraction method to separate spermatozoa from vaginal cells. This method was used to extract the spermatozoa from the vaginal samples that were collected from the bodies of the two girls. Subsequently, Professor Alex Jeffreys from the University of Leicester applied his method of using distinguishable patterns in a person's DNA to help solve the cases.

Professor Jeffreys discovered that certain regions of DNA contained DNA sequences that were repeated many times next to each other. He observed that the number of repeated sections could differ from one person to the next. These DNA repeat regions became known as a variable number of tandem repeats. His technique visualised these repeated sections as distinguishable patterns (resembling that of a barcode). Using the restriction digestion length polymorphism (RFLP) the first DNA profile was created by Jeffreys. Professor Jeffreys *first performed restriction digestion on the DNA extract using REase and separated the various DNA fragments by making use of agarose gel electrophoresis. To perform southern hybridization, the separated DNA fragments were transferred to a nylon sheet. The detection of the various fragments was made possible by hybridizing the radio-labelled probes to the various fragments. This original RFLP method resulted in many bands per sample on an autoradiogram, so giving the resemblance of a barcode. Unfortunately, his method also required a relatively large quantity of source DNA to generate a forensic DNA profile (Goodwin et al. 2011).*

It was clear that the same person had raped and murdered both the girls. By doing so, it was evident that Buckland's DNA was different and he was exonerated by using this novel forensic DNA profiling technique of Jeffreys.

The Leicestershire police and the Forensic Science Service then embarked in performing mass screening by collecting DNA samples from 5000 volunteering local men. Unfortunately, after almost six months of DNA examinations, no match could be were found. A DNA sample was only later collected from Pitchforth after a man named Ian Kelly was heard bragging in a Bar that he had given a sample while masquerading as the friend of Pitchfork. Later, Pitchforth's DNA was found to match the crime samples. Pitchforth confessed to committing both crimes.

Fig. 1 Example of DNA conviction and mass screening case

Evidential Value

Evidential value, or sometimes referred to as probative evidence, is the key to examining crime scenes and exhibit material effectively. The evidential value determines which exhibits to select for collection and examination. Evidence must be relevant if it is to be admitted into in court. This means it must be material (i.e., related to the specific crime under investigation) and probative (actually prove something). The exhibit material selected should contribute toward putting the pieces of the puzzle together to assist with the forensic investigation. Three key factors play a role in this selection (Manamela et al. 2015):

- What type of exhibit is it?
- Where was the exhibit found (location or the place where and/or position of the exhibit material and/or stain of the exhibit material)?
- To whom does the exhibit belong (ownership)?

A crime scene examiner, for example, may find it difficult as to what possible exhibit material to collect at a crime scene where the property was stolen. In illustrating the point, let's use the example where a factory was burgled and the perpetrator was not injured and left any bloodstains. The crime scene examiner will carefully examine the crime scene and identify entry and exit points, possible areas where the perpetrator had been or came into contact with the examiner. The crime scene examiner may have found that the perpetrator gained entrance to the factory by opening the main entrance door. To collect "skin cells" (epithelial cells/ touch DNA samples) from the handle of the main entrance door would not be of much value, since several people regularly during the day touch it to open the door. The door handle will contain a mixture of genetic material left by several contributing people, and it will be extremely difficult to interpret the forensic DNA profile. The forensic examiner will have to perform a thorough investigation and, apart from searching for other types of physical evidence, will need to identify other possible areas, which are most likely to be limited to the touch of the perpetrator. The crime scene examiner must also consider other different types of physical evidence and the different types of forensic examination types (such as trace analysis such as paint and fiber, ballistic examinations, disputed document examinations, chemical analysis) that may be performed on the physical evidence.

Evidential value of evidence means evidence which is sufficiently useful to prove something important in the crime investigation by either linking a perpetrator or providing evidence to eliminate a person from the criminal investigation of an offence or either to prove innocence or guilt of a person. To establish how the crime was committed, the crime scene examiner must collect and observe the clues and information (including information from eyewitness and tenants) at the crime scene, which will contextualize and reconstructs the events leading to the offence. Physical evidence is considered to be of little use without an interpretation of the

Scenario: *A woman is raped on the bed of her own apartment by an unknown male perpetrator. There are no signs of forced entry. The perpetrator is injured and scratched during the incident by the victim. The perpetrator left semen, hair (with roots) and his own blood (stains) on the linen of the victims' bed. She does not sustain any major injuries apart from bruise marks on the body and genital injuries.*

Analysis: *The crime occurred in the daily environment of the victim, where she continually deposits her own genetic material. The victim was not injured with a weapon and thus no bloodstain pattern marks will be observed. Consequently, the crime scene examiner will not collect genetic material deposited by the victim at the crime scene.*

Expect that the perpetrator will leave genetic material behind at the crime scene. The genetic material left will place the perpetrator on the scene where he had no earlier legal access to. Evaluate the evidential value of any object containing the perpetrator's genetic material (DNA) such as semen, saliva, hair root or blood (if he was injured), or if it links the perpetrator to the scene.

Conclusion: *The post-coital (rape sample) and nail scrapings from the victim, and the bloodstain, hair and semen stains left by the perpetrator on the linen will have evidential value and link the perpetrator to the crime scene. The exhibit material will give an excellent opportunity to get forensic DNA profiles for comparison matching.*

Fig. 2 Scenario of a rape case to illustrate evidential value

significance of its presence and/or the information that can be obtained for criminal investigation. The value of physical evidence can be measured either in generic terms, which may ultimately be subjective or in contextual terms (see Fig. 2: Scenario of a rape case to illustrate evidential value). Thus, a crime scene examiner and the forensic examiner are guided by the relevancy test about what to collect or what to the examiner.

Consider the opportunity to get a forensic DNA profile needs after determining the evidential value of which exhibit material needs to be collected for DNA analysis. Hence, the usefulness and evidential value of forensic analyses on any exhibit material are determined by the type of exhibit, ownership, and the location (the place where and/or position) of the exhibit material and/or stain.

It is important to show "ownership" when approaching forensic investigations. "Ownership" is determined by who was responsible for leaving the genetic material (DNA) at the scene or on the exhibit material. In forensic DNA analysis in the case of forensic biological evidence, it refers to the donor of the biological material such as blood, semen, "skin cells" (epithelial cells), and hair on a crime scene or object found on the crime scene. Compare the forensic DNA profile of the crime samples with the forensic DNA reference samples of the suspected perpetrator in the alleged crime to prove "ownership."

Concerning evidential value, it is critical to show whether a specific person had earlier access to a crime scene environment. Thus, establish the "location," the position where the exhibit (sample) is found in with the parties involved. The crime scene environment must be evaluated to determine whether it is a victim, perpetrator, or neutral environment. This is a crucial step to identify which donor's genetic material needs to be identified to prove the donor's presence on the crime scene during the commissioning of the alleged offence. The crime scene examiner

Example of physical evidence in a sexual assault or rape incident that may aid the investigation:

- Perpetrator's semen in the victim's orifices or deposited on an item at the crime scene
- Hair (roots) of perpetrator left at the crime scene
- A weapon with transfer evidence of some kind (for example "skin cells" from the perpetrator)
- Wound patterns on the victim
- Torn pieces of clothing
- The condom of the perpetrator left at the crime scene
- Victim's blood at the crime scene
- Fibres from ligatures used by the perpetrator to bind the victim
- Hair/Fibres of the victim in the perpetrator's vehicle
- Victims blood on the perpetrator's weapon

Example of physical evidence in property crime:

- Tool marks, fingerprints, blood (if injured), "skin cells" left at the point of entry or at strategic points where the item was handled by the perpetrator
- Broken doors/ windows (direction of broken glass)
- Missing valuables
- Footwear impressions

Fig. 3 Examples of physical evidence

must also examine the crime scene for physical evidence that may be submitted to other forensic science examination types.

Thoroughly search the crime scene for physical evidence to determine that a crime was committed. Examples of physical evidence that may have evidential value are illustrated in Fig. 3: Examples of physical evidence.

It may not be possible to scientifically find the age (time since deposition) of a body fluid stain found at the crime scene. It is, however, possible after prolong periods after deposit to detect body fluid stains.

The mere fact that a person's genetic material (DNA) is found at a crime scene does not conclude that the person is guilty or innocent of committing a crime (Hodge 2018). Forensic examiners must limit their court testimony to the value of DNA and not venture into opinions on how the genetic material (DNA) was left at the crime scenes or on the propositions or activities or on the way the DNA was transferred (Gill et al. 2020).

Forensic DNA Evidence

Deoxyribonucleic acid (DNA) is the blueprint of life. Genetic material (DNA) is the fundamental building block for a person's entire genetic makeup that contains the informational code for replicating the cell and constructing the needed proteins. It is part of virtually every cell in the human body, and a person's genetic material (DNA) is the same in every cell. Hence, the genetic material (DNA) in a person's blood is the same as the genetic material (DNA) in his or her skin cells, saliva, and other

biological material. Genetic material (DNA) is found in most cells of the body, including white blood cells, semen, and body tissue and hair roots. Traces of DNA can also be detected in body fluids, such as saliva and perspiration, due to the presence of epithelial cells therein.

The entire human genome has been sequenced, and it found that most of the human genome is the same for all people; it was demonstrated that only a small percentage of the human genome displays variation between persons. Thus, forensic examiners only need to use chemistries that focus on these regions of variation. For this reason, forensic DNA analysis is a powerful tool to link persons to crime scenes (Jakovski et al. 2017). By performing forensic DNA analysis, the identity of the donor of genetic material (DNA) is determined. A person can be linked to a crime scene by matching the forensic DNA profile derived from the exhibit material collected from the crime scene to the forensic DNA profile derived from a reference sample (buccal sample) collected from an identification subject (such as a suspected perpetrator). A forensic DNA profile is a string of alphanumeric characters, which donate identity. This sequence of alphanumeric characters is nothing more than a biometric containing information of a purely objective and irrefutable character. In a forensic DNA database, the forensic DNA profile derived from an identification subject only has value when comparing it with another profile derived from the database: if the subjects' forensic DNA profile matches that of a forensic DNA profile derived from the exhibit material collected from the crime scene.

The human genome has variable regions that make it possible to use the genetic material (DNA) information in forensic identity applications. DNA variation occurs in the form of different alleles or various possibilities at a particular locus. The two primary forms of a variation are possible at the DNA level: sequence polymorphisms and length polymorphisms (Butler 2012). Forensic DNA typing can be historically classified into fragment length polymorphism methods and polymerase chain reaction-based methods. Many countries are now using forensic DNA analysis based on STRs (Balding 2005).

More than 97% of the human genome is noncoding and repetitive. Forensic DNA analysis uses the repetitive nature and polymorphic properties of junk DNA. The number of repeats and the sequence structure of those regions vary between persons and organisms based on the entire genetic material (DNA).

Half of a person's genetic code comes from his or her mother, and half comes from his or her father. The genetic material (DNA) of close biological relatives can thus be used in paternity and kinship analysis (Saferstein 2011).

Many forensic laboratories deploy standard chemistries to generate forensic DNA profiles that contain no or very limited medical information of a person's predisposition or any physical characteristics (such as the color of the eyes or hair or how tall or short you are). There are trending chemistries becoming available in the forensic field to complement the arsenal testing chemistries available to a forensic laboratory. These chemistries can determine the additional attributes of the donor of the genetic material (DNA) on the exhibit material (Kayser 2015; Wienroth 2018). Recent innovations and developments in forensic DNA testing in the criminal field are related to the techniques of forensic DNA phenotyping, the use of ancestry-

informative markers, and familial searching. Forensic DNA phenotyping can be described as a set of techniques that aims to infer human externally visible physical features, such as the eye, hair, and skin color, and continental-based biogeographical ancestry of criminal suspects based on analysis of biological materials collected at the crime scenes (Machado and Silva 2019). The use of these additional chemistries may be limited and restricted in a laboratory by the regulatory framework of a particular country or region due to possible contravention of a human or constitutional right.

Adhere to the set of legal requirements and proper quality assurance standards when handling physical and presenting evidence in court. The regulatory framework and the prescripts of a particular jurisdiction give the guidance of how to process crime scene investigation and the seizure of exhibit material. Additionally, the international ISO standards exist and need compliance for implementing the quality management system for the processing and analyzing of crime scenes and exhibit material in laboratories. Aspects, like search warrants, the chain of custody, contamination issues, quality controls, and quality assurance, must have adhered to when physical evidence is collected, handled, and analyzed (Meintjes-Van der Walt 2010). No other forensic field has received so much scrutiny in court and through peer review as to what forensic DNA testing has been exposed to. It is for this reason that the quality assurance standards based on the ISO17025 standard for forensic DNA testing are very stringent and must meet the many guidelines required by the forensic community and by the court system.

During the forensic examination process at the crime scene or in the laboratory, the sources of genetic material (DNA) with evidential value need to be identified, documented, collected, and preserved to support the forensic DNA analysis process.

Forensic DNA Analysis

The forensic examiner must find the nature and donor of the DNA exhibit material (such as blood, semen, saliva, tissue, bones, and hair). The forensic examiner will analyze biological exhibit material with the aid of scientifically validated methods to:

- Determine the evidential value of exhibit material.
- Screening testing of exhibits using presumptive tests.
- Forensic DNA analysis.
- Bloodstain pattern analysis.
- Facial reconstruction.
- Anthropological examinations.
- The attending of crime scenes to aid the investigating officer with a reconstruction based on bloodstain pattern analysis.

Evidence recovery and presumptive testing are normally the first step in the laboratory process to analyze the exhibit material after being received. The analysis aims to:

- Document the chain of evidence of the exhibit material by documenting the condition of the packaging material, including the unique seal number and inscriptions on the packaging material.
- Document (photographing and/or sketches) the appearance and description of the exhibit material.
- Screening analysis, with the aid of presumptive test methods, is performed on the exhibits to find, whether any possible blood or semen is present on the exhibits (Exhibits which have been tested by the crime scene examiner may not be required to be retested.).
- Determine the evidential value of biological evidence identified on exhibit material.
- Removal of biological evidence from exhibit material for forensic DNA analysis.

After evidence recovery, including the type identification of the body fluid through presumptive and/or confirmation tests, is performed on exhibit material, a small stain of the exhibit material is subjected to forensic DNA analysis to give the forensic DNA profile (see Fig. 4: Steps of forensic analysis). After obtaining forensic

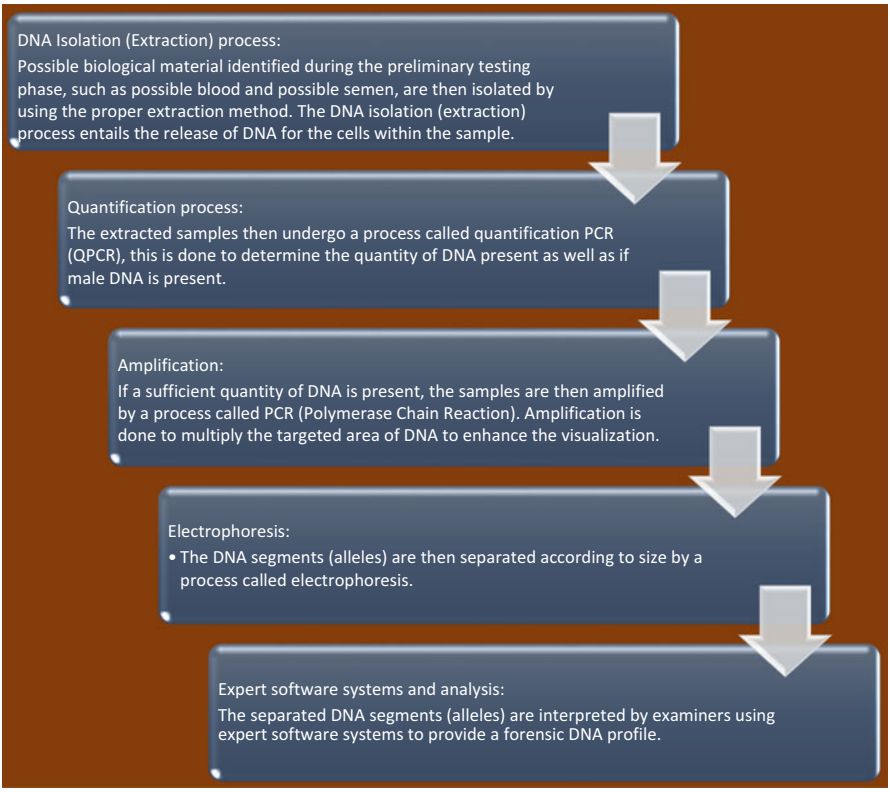


Fig. 4 Steps of forensic DNA analyses

DNA profiles from the exhibit material and reference sample from the subject or perpetrator, the forensic examiner must interpret the results and compile a forensic match report. The match (either "inclusion" or "exclusion" report) may exclude a perpetrator or provide a match between the perpetrator and the crime scene exhibit material. The inclusion match is supported with some statistical analyses to contextualize the results (statistical probabilities of a random match or likelihood ratios) that are used (Butler 2015). If required, the forensic examiner may be required to present the evidence in person to the court of law.

Forensic DNA Databases

Although forensic DNA technology has been used worldwide since 1985 to resolve crime, forensic DNA analysis was predominantly used as a prosecutorial tool for criminal and paternity casework on a case-by-case basis. The United Kingdom established the first forensic DNA database in 1995.

A forensic DNA database is a collection of forensic DNA profiles obtained from crime samples and donors, such as persons arrested and convicted offenders and used to perform comparative searches to establish an investigative lead. The value of a forensic DNA database as an investigative tool grows as the size of the database increases. The effectiveness of the forensic DNA database lies in the fact that many of crimes, such as sexual assault and property crimes, are committed by repeat offenders and that there is a likelihood that the person who committed the crime being investigated was convicted of a similar crime and their forensic DNA profile may already be loaded onto the database (see Fig. 5: Case study illustrating the effectiveness of a forensic DNA database). Access to forensic databases is mostly restricted.

By 2019, more than 69 countries have followed suit and established national forensic DNA databases to assist in resolving crime. The establishment of forensic DNA databases has raised ethical and human rights issues on the criteria of inclusion and retention framework of both samples taken from persons and forensic DNA profiles derived from these samples (Kumar et al. 2015). There are many reported instances where genetic information has been abused by governments to the detriment of its citizens. This has often included stigmatization and discrimination in areas such as insurance qualification, employment, health care, and education (National Research Council 1992; Coodly 2019).

It is evident that the ongoing public and policy debate is impacting the use and expansion of forensic DNA databases. Some safeguards are implemented at the national or regional level, but there is a lack of global standards and a need for more societal engagement and debate (Wallace et al. 2014).

It is thus important that countries should have detailed legislative framework that articulates the establishment and purpose for which a forensic national database may be used for, including aspects such as the retention periods for sample storage and profiles, access and security issues, the manner and from whom in which samples are collected, and how comparative searches is performed and reported.

CASE STUDY: EFFECTIVENESS OF A FORENSIC DNA DATABASE

The National Forensic DNA Database of South Africa (NFDD) was established in 2015. Forensic DNA profiles derived from crime scenes and certain categories of persons, such as persons arrested and charged; convicted offenders; persons under investigation and not arrested; missing persons and unidentified bodies are stored in different indices for comparative searches.

With the creation of the NFDD, comparative searches on the NFDD very quickly resulted in numerous person-to-crime forensic DNA investigative leads and several hundred crime-to-crime forensic DNA investigative leads. For example in early 2015, the forensic DNA profile of an unknown male was linked on the NFDD to 30 separate rape cases.

Sikhangele Miki, a 34-year-old male was charged with several offences committed between January 2011 and November 2015 in the areas of Delft and Khayelitsha, Cape Town, South Africa. His Modus Operandi, described by the court which ultimately convicted him, was "the signature of a person who is cruel, vicious and lacking any sense of empathy".

Miki would during the evenings, or early mornings would follow his victims, grab hold of them, threaten them at knifepoint by holding a knife to their necks and would demand money and cellular phones. He would then take them to an isolated place and rape them. Often he would either punch or hit his victims with the back of the knife handle or in some cases he stabbed them to stop them from struggling or crying for help. Nine of Miki's thirty victims were under the age of 16. One victim was even dragged to a space behind a police station where she was raped.

In 2014, Miki was arrested and convicted for an unrelated charge of assault with intent to cause grievous bodily harm for which he served 11 months in prison. During 2015, whilst serving time in prison a buccal sample was collected from him and submitted for forensic DNA analysis for uploading to the NFDD for comparative searching. The outcome of the comparative search resulted in linking Miki to the 30 unsolved rapes.

The case was brought before the High Court of the Western Cape where Deputy Judge President Patricia Goliath compared the modus operandi of Miki to a "monster lurking in the shadows, attacking, robbing and raping girls, exploiting the vulnerability of his victims." She went on to say that Miki derived pleasure in the degradation and pain inflicted his victims. She further stated, "The accused is a serial sexual predator. Even in the event of the remote possibility of rehabilitation, it is clear that he is a danger to society especially women and young girls."

Due to the overwhelming DNA evidence against the accused, Miki pleaded guilty to 84 charges which included 30 counts of rape, 27 of kidnapping, and 12 of robbery with aggravating circumstances. Miki was ultimately sentenced Miki to 15 life terms and an additional 120 years, to run concurrently:

Without the power of the NFDD and the DNA legislation which allowed DNA samples to be collected from the accused whilst he was serving time in prison as a convicted offender, it is unquestionable that Miki would have continued his reign of terror against vulnerable women in the future had he not been linked to his previous crimes by the NFDD.

Fig. 5 Case study illustrating the effectiveness of a forensic DNA database

Forensic DNA databases have now proven to be an effective investigative tool to resolve crime, which involves recidivism, and may eventually contribute to crime prevention and deterrence (Van der Beek 2015; Jakovski et al. 2017).

Studies in the United Kingdom (which launched its database in 1995) have shown that more than 60% of persons sent to prison for violent offenses and then released were rearrested for a similar offense in less than 3 years. Also, many "cold cases" before the existence of any forensic DNA databases can now potentially be linked to

an offender/perpetrator of another crime. Furthermore, when crime scene examiners focus and there is a significant increase in the collecting of genetic material (DNA) with evidential value at crime scenes such as property crime, the database becomes an effective investigative tool. This has then a profound positive impact on improving the conviction rates of the country (Asplen 2004).

In other words, not only will a forensic DNA database increase the likelihood of identifying known perpetrators, but it will also increase the possibility of linking perpetrators to multiple crime scenes.

Use of Forensic Lead Investigative Information

Several forensic databases, such as DNA, ballistic, and fingerprints, have been established to assist law enforcement to investigate and resolve the crime. After performing comparison searches between the biometric information loaded onto these databases, either a person is linked to a crime scene or different crime scenes are linked. Verified forensic investigative leads are subsequently issued to investigators to follow-up and investigate.

The mere fact that a person is linked through a database to a crime scene does not necessarily imply that the person is the perpetrator of the crime. There may exist legitimate reasons as to why the person's biometric information was found at the crime scene. The person may have had legal and regular access to premises where the crime was committed. For example, the person may be the owner or tenant or a regular visitor to the premises where the crime was committed. The fact that a person's genetic material (DNA) is found on a postcoital rape sample does not imply that the person is a rapist – the sexual intercourse may have been from consent.

Other data left on surveillance cameras or cell phone data or cell phone tracking data or geographical location data linked to images could be systematically exploited in the investigation. The use of forensic case data combined with temporal and graphical dimension may increase the likelihood of apprehending the perpetrator.

The recent technological innovation permits the data extraction of frames delineating more than one person. Software is now available that can assist to identify different perpetrators in complex DNA mixture results. A putative forensic DNA profile derived from a family member on the database can be used to link a perpetrator to the crime scene exhibit material.

In South Africa, forensic examiners routinely check for cross-leads between different types of forensic investigative leads (verified outcomes of comparison searches conducted on a forensic database such as DNA, Ballistic, and fingerprint databases (Smith 2020). A cross-lead is considered where there is common information, such as the case reference details (station and case number) and/ or the name of the person linked in the forensic investigative leads. By combining this biometric information and sharing it with the investigating officer, the number of cases linked is increased. Thus, there is a better opportunity for success in resolving the different cases by arresting a perpetrator. By increasing and simultaneous prosecuting the

number of cases, the perpetrator is more likely to be found guilty and convicted with a heavier sentence than when the cases are prosecuted in a piecemeal fashion.

Familial DNA Searches

The use of familial DNA searching can prove to be useful in a criminal investigation. Familial DNA searching assists investigators to point them toward a possible perpetrator. Persons who are close relatives tend to share more genetic (DNA) than unrelated persons. This can be used in criminal investigations by performing a familial search of exhibit material against a person's genetic information (DNA profile) on DNA databases or genetic genealogy database and to identify a possible relative of the genetic material (DNA) left by the perpetrator at the crime scene (Ram et al. 2018). Familial DNA matches are partial, or near, matches that require further investigation (Butler 2015).

Many support the use of familial searching as an investigative tool and argue that it has the potential to facilitate the identification and conviction of perpetrators, prevent crime, resolve cold cases, exonerate wrongfully convicted persons, and improve public safety; however, its use also raises important constitutional, ethical, and practical considerations that need to be taken into account by law enforcement (Niedzwiecki et al. 2017). The arguments for and justifying why DNA samples must be taken from certain categories of persons, such as convicted offenders and arrestees, are that they have a diminished expectation of privacy as a result of their conviction or arrest. This argument cannot be applied to why offenders' relatives need to be linked through a database search. Familial searches may uncover differences in genetic and social family relationships. Genetic data from distant relatives in public genetic genealogy databases have aided many cold case investigations. For example, the use of the private heritage of genetic genealogy databases was applied to identify and led to the arrest of alleged "Golden State Killer" Joseph De Angelo (Fuller 2018].

Modus Operandi and Signature of Serial Offenders

Although a comparison search on the forensic DNA database may not provide the specific identity of a known person whose forensic DNA profile is already uploaded, different crime scenes may be linked due to the forensic DNA profile matches. Offender profiling, modus operandi, and signatures left by serial perpetrators become useful during the investigation of the linked crimes (Baker et al. 2014).

The modus operandi of repeat or serial offenders is an important consideration in the investigation of crime. It is simply the way a particular perpetrator operates. Modus operandi accounts for the type of crime and property or persons attacked, the

tools or weapons used in committing the crime, and the way the perpetrator gained entrance and the victim was approached (including disguises or props used), how the perpetrator enters the crime scene, and the time and place the crime was used. The modus operandi of these perpetrators often remains stable or changes very little over the seriousness of crimes committed. In addition to the modus operandi, the signature of the serial perpetrator provides an additional information in the criminal investigation. The signature of the serial perpetrator is the imprint or expression that is used to reflect their unique identity that feeds toward the feeding of their fantasies and satisfaction. The serial murder may, for example, leave specific markings on the body or display in a bizarre position or have a particular ligature to type his victims (Baker et al. 2014).

Offender Profiling

The purpose of offender profiling is to assist investigators to narrow their search for the field of the perpetrator. This is based on the characteristics of the crime scene and initial investigative information. Information descriptors include the following: sex, race, approximate age, criminal history, residency concerning the crime scene, employment history, social adjustment, sexual abnormities or adjustment, use of narcotics or stimulants, educational level, and interpersonal skills. Although profiling may assist in sentencing mitigation, it is rather used as an aid than as evidence in court. It has proven to be vital during the investigative phase, especially in the investigation of serial murder casework (Baker et al. 2014).

Best Practices Using Forensic Investigative Leads as Effective Investigative Tool

It is important for every investigating officer, who receives a forensic investigative lead, to immediately make contact with the other investigating officers, working on the other linked cases. A forensic investigative lead will provide valuable information for a case under current investigation that has been linked to another case(s). This means that, for example, the evidence in case A matches the DNA case B and other cases (where a perpetrator may or may not have been arrested). The information indicates this is now a serial rape or serial murder investigation. A serial rapist is a perpetrator who rapes two or more victims. A serial murderer is a person who murders two or more victims because of an inner desire/urge to kill them. There may be information that indicates a serial rapist or serial murderer is active in the area because of a similar modus operandi. Similar information or investigative leads may have been provided on fingerprint evidence or IBIS (ballistic database) links that can be of assistance for solving the relevant cases as identified.

Certain best practices need to be followed to ensure a successful arrest and prosecution.

Step 1: Notification

Notify the relevant persons and stakeholders within your district or region and nationally of the fact that a serial offender is active. This will include making contact with police commanders and responsible person at behavioral investigative units, which specializes in assisting investigators in investigating serial offender linked cases.

Warn the public that a serial rapist or serial murderer is active in the area. This should be done only after approval has been obtained to involve the media. Convey this information to local newspapers, radio stations, and television.

Step 2: The Investigation

Certain investigative steps need to be taken to effectively investigate a series. The most effective way to investigate serial related cases is to form a task team. All linked cases, whether through DNA evidence, fingerprints, cellular phone data, or identikits, should be consolidated under the task team, one dedicated investigator detective. This step is recommended, even if some cases are outside the investigator's area of responsibility and investigation, for example, if a case is from another region.

Appeal to Victims to Come Forward

Since only a few cases of rape are reported to law enforcement, it means that there were other victims of the serial rapist that were not reported. These victims might have valuable information that could help solve the series.

An appeal in the media could lead to these victims coming forward and opening cases. The modus operandi and identikit can be put in the local newspapers and appear on television news with contact details, asking for victims to come forward. Even if the victim does not want to open a case, valuable information can be obtained in this way.

Search for Similar Cases

While DNA evidence (genetic material) can be a way to identify other cases, the investigator must not only rely on DNA evidence (genetic material) only to help identify other cases that are the work of the same perpetrator. The investigator should enquire from colleagues whether they have cases with similar modus operandi or to check with crime intelligence officials to check whether other cases of rape or murder have been reported in the same area.

The investigator must examine case files relating to assault, attempted rape, robbery, and even murder that occurred in the same geographical area of the linked cases. These cases might also be linked to the same perpetrator but maybe were not registered as rape cases. Serial rapists and serial murderers like to use a small geographic area in which to commit their crimes. Check for cross-forensic investigative leads.

It should be mindful that certain information collected may be sensitive or classified material requiring protection and of which unauthorized disclosure or loss could reasonably be expected to be prejudicial to lawful methods for the protection of public safety.

Revisit the Crime Scenes

It is essential to revisit the crime scenes as soon as possible. In rape cases, investigating officers often do not revisit the crime scene with the victim to locate exhibits and determine exactly where the crime had occurred. The investigator should look for evidence such as condoms, semen, and tissues the perpetrator had used and for belongings of the victim that might have the perpetrator's fingerprints on them.

Consideration may also be given to using dogs that have been trained to identify biological to assist in locating exhibits that may have the perpetrator's semen on it at crime scenes.

Revisit the crime scene as soon as possible, should a perpetrator agree to do pointing out later, after the arrest. A perpetrator might only be arrested a year after the rape and the area might look very different then.

Ideally, investigators should record the crime scene location by using GPS coordinates and mapped. Since serial criminals often continue to use the same geographical area, it will be a good idea to keep the area under observation, should the perpetrator return.

Exhibits Material

All sexual assault evidence collection kits and other exhibit material must be sent to the forensic science testing laboratory for immediate processing, and not only once a perpetrator has been identified. The investigator must maintain adequate documentation and chain of custody documentation. Ensure that the chain of evidence statements is obtained immediately, and not later.

Reinterview Victims

Investigators should remain in contact with victims and victims in other linked cases, to ensure that the victim is available for the court process. All victims need to be reinterviewed and informed that the cases are still under investigation. Often victims see the perpetrator again because serial offenders use the same areas to commit future crimes. If the victim has the investigating officer's contact number, it can assist in apprehending the perpetrator. This has happened in many serial cases because the perpetrator likes to use the same geographical area again. Sending an SMS to the victim once a week, indicating that the case is still under investigation or keep the victim informed of the progress made, will encourage the victim to remain in contact and to proceed to trial when the perpetrator is arrested.

Cellular Phones

Most serial rapists or serial murderers take the victim's phone. It is essential that detailed billing for the victim's number should be obtained from the day before the perpetrator made contact with the victim up until the date of the court order

application to obtain direct access to telephone records. Sometimes the perpetrator uses the victim's phone and SIM card to make a call, before throwing away the victim's SIM card. The number that the perpetrator called can then also be identified through the court order to see who was called. More likely, the perpetrator will replace the victim's SIM card with his own SIM card and that number should also be identified through the court order.

Besides the detailed billing of the SIM card(s), also obtain a court order for the handset profile for the stolen cellular phone for the same period as the detailed billing. It will provide information on the SIM cards that have been used in the stolen phone. The cellular phone information that is obtained through the court order should be analyzed to see how it can assist with the investigation and should not only be filed. It is important to note that the victim does not need to know their phone's IMEI number. The IMEI number can be obtained from the detailed billing of the SIM card number. Also, note that detailed billing and a handset profile can be obtained for a pay-as-you-go number.

Step 3: After the Arrest

Once a perpetrator has been identified and arrested, the perpetrator's DNA reference (buccal) sample should be immediately submitted for forensic DNA analysis.

Preparation for the Trial

All cases linked to one perpetrator must be tried together in one trial as this improves the chances of obtaining a successful conviction. If cases are from different court jurisdictions, areas, make an application to centralize the cases for trial in the same court. This can even be done for cases outside the province. It is important not to only prosecute in cases where there is DNA evidence, if the modus operandi is the same, but also consider using the modus operandi (similar fact evidence) to obtain convictions in those cases that are similar but without the physical evidence as proof. The investigators from behavior science units can be approached to also give evidence in supporting the case investigation.

Sentencing

For sentencing, the testimony must be given about the fact that the accused is a serial rapist and/or serial murderer and he or she is a danger to society. The investigators from behavior science units can give evidence in this regard.

The comments of the presiding officer, along with any sentencing reports, should accompany the convicted offender to the Department of Correctional Services and should be put on the perpetrator's file at the Department of Correctional Services. This is important because one day the offender will appear before a parole board, and unless this information is in the offender's file, the Department of Correctional Services or Parole Board cannot make a properly informed decision on the parole matter.

Reasons for Unsuccessful Investigations of Serial Casework

Investigators often become very despondent when there is a negative court outcome in serial offender casework. The following are some of the reasons why serial investigations may not have the expected outcome:

Serial Related Cases Not Consolidated

Sometimes forensic DNA-linked cases may be investigated by different units or investigators. For example, rape that occurred outside in an open field will usually be investigated by one unit; however, a rape inside a victim's home might be investigated as a house robbery and dealt with by the organized crime investigators. Rape and murder might be investigated by different independent investigators, even though all of these cases are linked to the same perpetrator based on the genetic material (DNA) evidence. The best practice would be to form a task team to deal with these cases and work together, using all the information that is available in different cases.

Serial Casework Investigated by Inexperienced Detectives

Inactivity in follow-up with investigative leads will lead to the perpetrator committing more crimes. The most experienced investigator should be in charge of the investigation. Junior investigators can still be involved to learn from the experienced investigator(s).

Cases Closed as Undetected Are Not Reopened

Even if a case had already been closed as undetected by the time the forensic DNA report stating that these cases are linked, the cases should be reopened and included in the serial investigation. Often these cases have valuable leads that were not properly followed up at the time.

Informal Identity Parades

If an investigating officer identifies a possible perpetrator and takes the perpetrator to the victim and asks the victim whether he/she indeed was the correct perpetrator, the case investigation may become compromised. Besides the fact that this is not the correct procedure, it has no evidential value. Investigators must hold formal identification parades according to the prescripts so that this identification evidence can be used in court.

Pretrial Preparation

Forensic examiners must always be objective and impartial when dealing with forensic evidence. Examination test methods used and interpretation of results must be processed following international acceptance criteria. In the interest of good justice, it is important that both defense lawyers and prosecutors adequately

prepare for a trial. Smith (2019) from the Section Forensic Database Management of the South African Police, who is responsible for managing the DNA database, and Meintjes-Van der Walt (2010), an experienced prosecutor in criminal casework, propose the following set of questions in preparing for court cases where forensic DNA evidence is relevant:

- Has the chain of custody of key evidential material been proved and supported with documentation?
- Has the possibility of contamination been considered and eliminated?
- Are all the forensic DNA test examinations supported with the required documentation?
- Was there adequate discovery?
- Was the accused able to exercise his or her right to a fair trial?
- What are the standard operating procedures of the testing laboratory? Are they available for the defense? Have they complied within the specific case?
- Who are the crime scene and forensic examiners that performed the tests on the exhibit material?
- What is the competency (including qualifications and training) of the crime scene and forensic examiners?
- What are the proficiency test results of the forensic examiners that were involved with the examination of exhibit material in the specific case?
- Has the laboratory got a quality management system based on the ISO17025 international standard? Is the laboratory accredited and been subjected to internal and external audits?
- Could the DNA exhibit material (including work and storage areas) and laboratory equipment been contaminated during the handling and testing thereof?
- What are the controlling measures, including quality controls, during the methods tests to monitor contamination? Was contamination excluded in the specific case? Was a possibility of the exhibits been contaminated by the crime scene and forensic examiners excluded?
- Are all the forensic DNA test methods validated? Are the validation studies available?
- What equipment (with serial numbers) were used in the forensic DNA examination of the exhibit material? Are the forensic DNA equipment calibrated? Are the calibration certificates available on request?
- Has potential examiner bias been addressed and excluded in the interpretation of the results in the specific case? Was the forensic DNA results and report subjected to administrative and technical review before releasing the report?
- Does the defense have access to STR electropherograms? Have the allele designations been processed correctly? Is there any ambiguity with the electropherograms?
- Were the isolated DNAs from the exhibit material quantified? What was the quantity of the DNA used for amplification?

- Is the profile of the exhibit material a single source or a mixture? What are the criteria for making a match? Consider requesting a qualitative analysis approach with the genotypes contributing to the mixture.
- What is the possibility that a close relative was involved in the crime?
- Did the accused recently have a bone marrow transplant?
- What statistical method was used to support the DNA results? Was the statistical value of the exhibit material calculated correctly and according to international acceptable practice?
- What is the possibility of an additional contributor with mixture results?
- What other physical and other evidence is there to support the DNA evidence?

Erroneous Forensic DNA Findings

Forensic DNA evidence is based on accepted scientific principles and courts have "judicially" noticed DNA as a reliable scientific test. There are, however, many instances where reported forensic DNA findings have proved to be erroneous. Instances of incompetence or deliberate fraudulent action have occurred, knowingly excluding information or removing information from an exculpatory report of the accused, knowingly giving incorrect oral, purposefully concealing that an error occurred in practice, fabricating competency and qualifications, providing evidence to tests that were never performed, concealing sample switches or contamination, falsifying laboratory reports, or providing results based on non-validated test methods, wrongfully "reading" and including an accused into a mixture forensic DNA profile, and overstating the significance of DNA results (Cooley 2010).

DNA evidence may be presented in a misleading way, where it is either exaggerated, such in the case of the prosecutor fallacy, or underestimated, such as in the defendant fallacy: It is important for all parties (expert witnesses, lawyers, and presiding officers) concerned to understand the mathematical principals supporting forensic DNA evidence. Forensic evidence is often given in terms of a mathematical probability. The forensic examiner must be cautious that they do not make the error of "the prosecutor's fallacy," in the interpretation of such probability (Thompson 2009):

DNA analysis is subject to human error based on the interpretation of the DNA results. These errors are often made when dealing with mixture samples, low copy number DNA (Gabel 2014), and degraded evidence (Hodge 2018). Low Copy Number DNA refers to trace DNA from which it is difficult to obtain a full profile such as availability of limited, damaged, or degraded DNA, oligospermic or aspermic perpetrators or from extended interval post-coital samples.

Forensic science testing laboratories must implement a quality management system that is open to peer review. The quality management system will include various quality control measures to cover the different activities and processes of the forensic DNA laboratory (Wienroth 2018).

A forensic examiner must be committed to professional ethics and seek the truth (Li 2015). His or her conduct must be impartial and without bias. The forensic examiner must embrace science and always ask if it is good science that is being practiced. Practicing good science requires an understanding of the principles of the scientific method and the limitation of the forensic method in supporting the forensic result and interpretation.

Conclusion

Although forensic DNA testing is a powerful investigative tool, it is not always the silver bullet in case examination and investigators must not neglect other forensic examination methods and database capabilities to resolve their casework.

Forensic DNA testing has proved to be very useful in resolving criminal casework to assist with the:

- Conviction of a perpetrator.
- Exoneration of an innocent person.
- Identification of a poacher or the determination that poaching has occurred.
- Proving of championship pedigree.
- Identification of the bodies and remains in mass disaster incidents.
- Determination of paternity.
- Determining who was driving a vehicle during an accident if the blood of the driver is found on the windscreen.
- Linking a perpetrator to several crime scenes by performing comparative searches on a DNA database.

The mere fact that a person's DNA is found on a crime scene or an exhibit item or rape sample does not in itself imply that he or she is the perpetrator of a crime. There may be an innocent explanation as to why the DNA person is deposited at a crime scene. The person may have had a legal reason or is the resident at the crime scene. Furthermore, in rape cases, sexual intercourse may have been a consensual sexual intercourse.

With the advent of uploading forensic DNA profiles to DNA databases, it became possible to perform comparison searches and identify serial offenders, or multiple offenders, and to distinguish copycat offenders. In most countries, on the one hand, when an offender commits two or more of the same offenses at different time intervals, the offender is known as a serial offender. This is particularly important, given the recidivist nature of sexual predators. On the other hand, an opportunist offender that commits different offenses is known as a multiple offender. Forensic DNA investigative leads and cross-leads have become important evidence in the arsenal to fight crime.

There is a growing trend for forensic DNA products to be delivered as intelligence products over and above the direct evidence presented in the courts. These

trends are opening a very exciting future of using technology to support criminal investigations and the resolution thereof.

Forensic DNA analysis must be performed efficiently and reproducibly while complying with the regulatory framework, and the forensic examiner must maintain scientific objectivity so that the evidence holds up in the courts.

References

Asplen CH (2004) The application of DNA technology in England and Wales. Smith Alling Lane. https://www.ncjrs.gov/pdffiles1/nij/grants/203971.pdf. Accessed 12 Sept 2020

Baker PK, Keppel RD, Napier MR, Sadoff RL, Schlesinger LB (2014). In: James SH, Nordby JJ (eds) Forensic science: an introduction to scientific and investigative techniques. CRC Press, Boca Raton

Balding DJ (2005) Weight-of-evidence for forensic DNA profiles. Wiley, West Sussex

Butler JM (2010) Fundamentals of forensic DNA typing. Elsevier Academic Press, Burlington

Butler JM (2012) Advanced topic in forensic DNA typing: methodology. Elsevier Academic Press, Burlington

Butler JM (2015) Advanced topic in forensic DNA typing: interpretation. Elsevier Academic, Burlington

Butler JM, Schoske R, Vallone PM, Kilne MC, Redd AJ, Hammer MF (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int 129(1):10–24. https://doi.org/10.1016/s0379-0738(02)00195-0. PMID: 12230993. Accessed 10 Sept 2020

Coodly LR (2019) DNA and its privacy. Master's thesis, Harvard extension school. http://nrs.harvard.edu/urn-3:HUL.InstRepos:42006723. Accessed 11 Oct 2020

Cooley CM (2010) Miscarriages of justice: causes and suggested reforms. In: Petherick W, Turvey B, Ferguson CE (eds) An introduction to forensic criminology. Elsevier, Burlington, pp 313–321

Fuller T (2018) How a genealogy site led to the front door of the golden state killer suspect. The New York Times, April 26. https://www.nytimes.com/2018/04/26/us/golden-state-killer.html. Accessed 23 Aug 2020

Gabel JD (2014) Realizing reliability in forensic science from the ground up. J Crim Law Criminol 104(2):2. https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=74. Accessed 10 Oct 2020

Gill P, Hicks T, Butler JM, Connolly E, Gusmão L, Kokshoorn B, Morling N, Van Oorschot RAH, Parson W, Prinz M, Schneider PM, Sijen T, Taylor D (2020) DNA commission of the international society for forensic genetics: assessing the value of forensic biological evidence – guidelines highlighting the importance of propositions. Part II: evaluation of biological traces considering activity level propositions. Forensic Sci Int Genet 44. https://doi.org/10.1016/j.fsigen.2019.102186. Accessed 12 Sept 2020

Goodwin W, Linacre A, Hadi S (2011) An introduction to forensic genetics (Essentials of forensic science), 2nd edn. Wiley, Oxford

Gould JB, Carrano J, Leo R, Young J (2013) Predicting erroneous convictions: a social science approach to miscarriages of justice. Final technical report. NCJ 241839. U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, Washington, DC. https://www.ncjrs.gov/pdffiles1/nij/grants/241389.pdf. Accessed 28 Aug 2020

Hodge S (2018) Current controversies in the use of DNA in forensic investigations. Univ Baltimore Law Rev 48(1):3. Accessed 3 Sept 2020

Jakovski Z, Janovska R, Stankov A, Poposka V, Bitoljanu N, Belakaposka V (2017) The power of forensic DNA databases in solving crime cases. Forensic Sci Int 6:e275–e276. https://doi.org/10.1016/j.fsigss.2017.09.085. Accessed 10 Sept 2020

Kayser M (2015) Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet 18:33–48. https://doi.org/10.1016/j.fsigen.2015.02.003. Accessed 11 Sept 2020

Kumar S, Verma AK, Singh P, Singh R (2015) Current scenario of forensic DNA databases in or outside India and their relative risk. Egypt J Forensic Sci 6:1–5. https://core.ac.uk/download/pdf/82220247.pdf. Boca Raton

Li R (2015) Forensic biology, 2nd edn. CRC Press, Boca Raton. pp 373, 485–496, 528

Machado H, Silva S (2019) What influences public views on forensic DNA testing in the criminal field? A scoping review of quantitative evidence. Hum Genomics 13:23. https://doi.org/10.1186/s40246-019-0207-5. Accessed 11 Sept 2020

Manamela M, Smith JH, Mokwena RJ (2015) Serological evidence and DNA analysis. In: Zinn R, Dintwe S (eds) Forensic investigation. Legislative principles and investigative practice. Juta & Co (Pty) Ltd, Cape Town, pp 96–118

Meintjes-Van der Walt L (2010) DNA in the courtroom: principles and practice. Juta & Co (Pty) Ltd., Cape Town, pp 85, 109, 113–114

National Research Council (1992) DNA technology in forensic science. The National Academies Press, Washington, DC, p 10.17226/1866

Niedzwiecki E, Sherrill SD, Field MB (2017) Understanding familial DNA searching: coming to a consensus on terminology. Document number: 251080. U.S. Department of Justice. Office of Justice Programs' National Criminal Justice Reference Service. https://www.ncjrs.gov/pdffiles1/nij/grants/251080.pdf. Accessed 3 Oct 2020

Ram N, Guerrini CJ, McGuire AL (2018) Genealogy databases and the future of criminal investigation. Science 360(6393):1078–1079. https://doi.org/10.1126/science.aau1083. Accessed 11 Sept 2020

Saferstein R (2011) Criminalistics an introduction to forensic science, 10th edn. Pearson, pp 8, 15, 69, 261, 266–269

Smith JH (2019) Ensuring confidence of supplier product from forensic testing laboratories when accepting forensic DNA profile to be loaded to the National Forensic DNA Database of South Africa. South African Police Service, Cape Town, South Africa, pp 1–3

Smith JH (2020) Follow-up and investigation of forensic investigative leads. South African Police Service, Cape Town, South Africa, pp 1–6

Thompson WC (2009) Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. Law Probab Risk:257–276. http://lpr.oxfordjournals.org/cgi/reprint/mgp013?ijkey=ehrDeU74Gi6ERv1&keytype=ref. Accessed 3 Oct 2020

Van der Beek K (2015) Measuring the effectiveness and efficiency of forensic DNA databases international symposium of human identification. 26. https://www.promega.com/-/media/files/products-and-services/genetic-identity/ishi-26-oral. Accessed 3 Oct 2020

Van Oorschot RA, Jones M (1997) DNA fingerprints from fingerprints. Nature 387:767. https://doi.org/10.1038/42838. Accessed 9 Nov 2020

Wallace HH, Jackson AR, Gruber J, Thibedeau AD (2014) Forensic DNA databases-ethical and legal standards: a global review. Egypt J Forensic Sci 4(4):57–65. https://doi.org/10.1016/j.ejfs.2014.04.002. Accessed 9 Nov 2020

Wienroth M (2018) Governing anticipatory technology practices. Forensic DNA phenotyping and the forensic genetics community in Europe. New Genet Soc 37(2):137–152. https://doi.org/10.1080/14636778.2018.1469975. Accessed 9 Nov 2020

Introduction to Forensic DNA Typing and Current Trends

2

Monika Chakravarty and Prateek Pandya

Contents

M. Chakravarty
Forensic Science Laboratory, Home Department Government of NCT of Delhi, Delhi, India

P. Pandya (✉)
Amity Institute of Forensic Sciences, Amity University, Uttar Pradesh, India
e-mail: ppandya@amity.edu

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_1

Abstract

DNA profiling in forensic context has seen tremendous growth over the last almost four decades. Forensic DNA fingerprinting started with Alec Jefferey's single and multilocus probes via autoradiography followed by fluorescence-based PCR-STRs. Lately, there has been a significant change in technology for human identification. Chip-based massively parallel sequencing is the most recent advancement. Drawing investigative leads, disaster victim identification, and unsolved cold cases created a requirement for more efficient ways of extraction, rapid DNA testing, and genetic genealogy. Pooling forensic genetic genealogy, mitochondrial DNA for challenging sample analysis and STR-SNP sequencing has opened more avenues to generate the data leading to identification. This chapter highlights some of the advanced methods adopted in forensic DNA profiling and a few appropriate cases incorporating such methods.

Keywords

InDels · Massively parallel sequencing · Forensic DNA phenotyping · Biogeographic ancestry · Genetic genealogy · GEDmatch · Rapid DNA

From Sampling to Identification

Human identification is an essential component of forensic science. In the earlier times, the emphasis was on the phenotypic aspects of human body such as anthropometric measurements, color, and related features. However, since the advent of DNA technology, the field of human identification has grown significantly in terms of establishing individuality, relationships, and cultural and ethnic correlations. In addition, DNA technology has also provided various interesting and useful avenues such as establishing the genetic basis of various diseases through marker identification, evolutionary biology, gene therapy, mapping, subspecies level identification, habitat correlation, and related areas. The journey of DNA as an identification tool in forensics for the past three decades has been comparatively simple. Generally, a crime scene sample is profiled using existing methods and compared to a suspect or uploaded to a forensic DNA database of convicted offenders. Crime investigating laboratories mostly apply polymerase chain reaction (PCR) and fluorescence-based capillary electrophoresis (CE) to detect length variations in short tandem repeats (STRs). Although current capillary electrophoresis is the gold standard for analysis of forensic samples and is the method of choice for separating STRs for most forensic laboratories, the advent of massively parallel sequencing (MPS) or next-generation sequencing (NGS) has shown new avenues for detailed genetic analysis. NGS technology is evolving rapidly over the last decade and has proved

advantageous in challenging forensic samples, including mixtures, low copy number DNA, and degraded samples. NGS technology offered comprehensive results in case of mismatches observed in disputed paternity cases (Ma et al., 2016).

Improved Software Tools

Software tools for data collection emanating from capillary electrophoresis are an essential component of DNA profiling technology workflow. Advanced data collection and analysis software could affect data analysis in capillary electrophoresis. This could facilitate data processing by reducing off-scale data, thereby increasing laboratory's output. Stochastic effects like allele dropouts, allele drop in, sister allele imbalance, and stutters occur more often in low DNA samples and may lead to stutter amplification. Upgraded 3500 Data Collection Software v4.0.1 (User Bulletin Publication Number 100075298, Thermo Fisher Scientific) provides new features to reduce spectral pull ups. Gene Mapper™ v1.6, automated genotyping software (User Bulletin Publication Number 100073905, Thermo Fisher Scientific), cuts off threshold limit from 33,000 RFU (Relative Fluorescence Unit) to 66,000 RFU and enables accurate evaluation of peak heights to determine minor or low-level contributor. It reduces off-scale data and is useful both for routine casework and database samples. Improved software provides profile comparison feature, improved marker labeling along with stutter filters so that the stutter peaks are not exported. Recently, there have been integrated CE and NGS case management workflows having data concordance (Converge software, Thermo Fisher Scientific) in which the samples can be analyzed both for CE and NGS at the same platform. Improved software tools have population statistics included in the software itself; however, the databases can be customized according to the region. Variants can be accurately detected, and samples that have been run on the same sequencer can be compared. Fast, automated analysis of complex data and multiple data export options have streamlined the analysis. Some laboratories are also in the process of developing probabilistic genotype software to estimate the number of male and female contributors (Coble and Bright 2019).

CE Data Interpretation Improvement

- Pull ups reduced.
- Signal optimization across the capillaries.
- Reduced off-scale data.
- Incorporation of 6-dye chemistry.
- Automatic spatial calibration.
- RFID tracking.
- One reagent cartridge.
- Common array length and single polymer for all applications.
- Small bench top instruments, touch display.

Massively Parallel Sequencing/Next-Generation Sequencing: Advanced Human Identification

DNA sequencing consists of identifying base sequence of certain sections or entire length of DNA molecule. Sequencing technology has evolved tremendously since first-generation technology started in the 1960s (Heather and Chain 2016). Since then, the technology has improved significantly while jumping from first-generation (Maxam and Gilbert 1977; Sanger, Nicklen, and Coulson 1977) to second- (Hyman 1988; Shendure and Ji 2008) and third-generation tools (Niedringhaus et al. 2011). Conventional capillary electrophoresis has been the chosen method for forensic laboratories to identify the perpetrators of crime and to exonerate individuals (Chakravarty et al., 2019; Shrivastava et al., 2012). Molecular biology in association with population genetics principles paved the way for excellent human identification and construction of large number of DNA databases (Dixit et al. 2019; Srivastava et al. 2020). Short tandem repeats (Autosomal STRs, Y-STRs, X-STRs, and mini-STRs,) mitochondrial DNA, and single-nucleotide polymorphisms (SNPs) are the tools used for CE fragment analysis (Fig. 1).

With recent developments in next-generation sequencing (NGS), novel methods have been devised within the forensic community to make way for the investigative proceedings when CE-STR analysis fails to produce results. Several forensic laboratories are in the process of applying the MPS technology for the analysis of conventional STR markers and mitochondrial control DNA region and the possible uses of other DNA markers not frequent in casework such as next-generation STR kits, single-nucleotide polymorphisms (SNPs), insertion/deletion (InDel) markers, and mitochondrial DNA sequence (Phillips et al., 2007). With NGS we can generate a lot of data from a single sample, viz., autosomal, Y-STRs, X-STRs, identity SNPs, as well as phenotype and biogeography ancestry (Phillips, 2015) (Fig. 2).

The NGS Technology

With the initiation of Sanger sequencing method in the 1970s (Sanger et al. 1977), DNA sequencing technology has come a long way. Several genome projects have been completed using Sanger technology. However low throughput and high cost pose a limitation to its use in more complex genome analyses (Fullwood et al., 2009). The recently introduced NGS technology has superseded these problems and

Fig. 1 Current tools for routine, degraded, kinship, and sexual assault samples

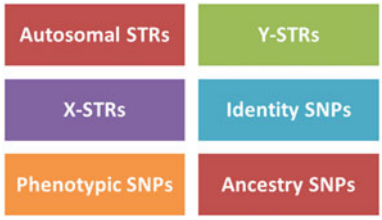


Fig. 2 NGS: Multiple markers, one amplification

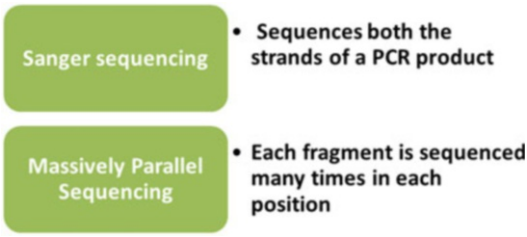


Fig. 3 Sanger sequencing versus massively parallel sequencing

is being used in forensics (Weber-Lehmann et al. 2014), diagnosing diseases (McCarthy et al., 2013), and ancient DNA analysis (Poinar et al. 2006). With Sanger sequencing, both the strands of a PCR product are sequenced, i.e., we read the sequence in one direction followed by reading in other direction, eventually sequencing each base twice. In MPS/NGS each base is sequenced many times yielding much more information about a sequence (Yang 2014). Billions of molecules can be sequenced in parallel, hence the name massively parallel sequencing (Fig. 3). Sequencing multiple reads at the same time leads to reduction in time as well as cost. Roche introduced the world's first high-throughput sequencing system utilizing pyrosequencing based on sequencing by synthesis in 2005 (Margulies et al. 2005) followed by the technologies offered by Thermo Fisher Scientific, Illumina, Ion Torrent Inc., and Pacific Biosciences (PacBio) to name a few. Since then, the NGS technology has been offering new horizons for forensic genetics.

Comparing CE and NGS

Extracted DNA in CE technology is PCR amplified using commercially available kits with fluorescent dye labeled primers to multiplex STRs with overlapping size ranges. PCR products are then separated through CE according to their molecular weight. The result is in the form of electropherogram comprising peaks with their sizes expressed as base pairs and height expressed in relative fluorescent units (Fig. 4) (Riman et al., 2020). NGS workflows (Fig. 5) also have PCR amplification to enrich STR markers. The target-specific primers contrarily are not fluorescently labeled but are taken for library construction. PCR products have adapters attached

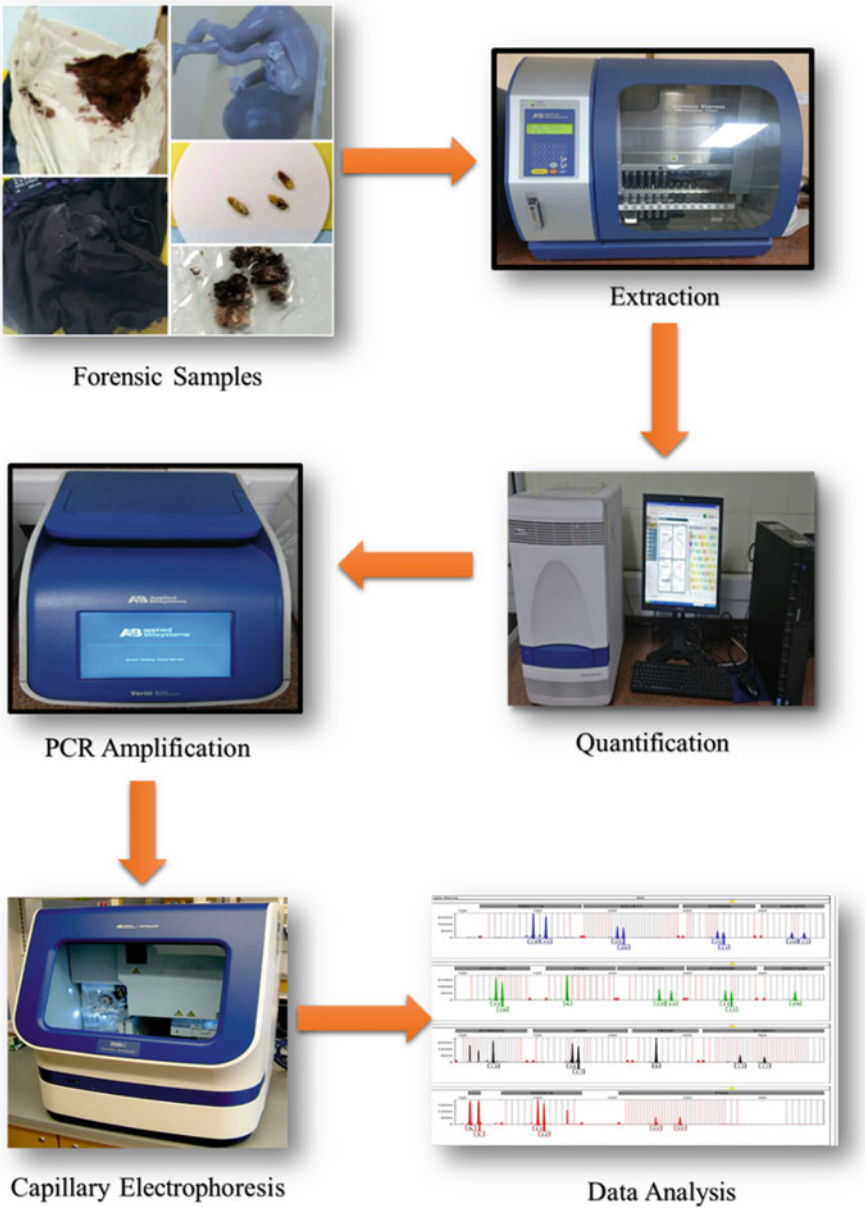


Fig. 4 Workflow of a traditional DNA Analysis

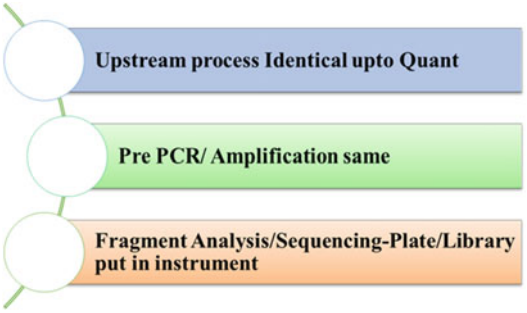
at both the ends producing DNA libraries which can be sequenced (Müller et al. 2018). Raw digital data of sequencing reads is obtained as a result of sequenced DNA libraries (Mardis 2017). Multiple libraries can be pooled into one reaction as

Fig. 5 NGS Workflow

Table 1 CE versus NGS

Parameters	Capillary electrophoresis	NGS/MPS
Variation	Length based	Sequence based
Primers	Fluorescent	Non-fluorescent
Random matching probability	Length-based discrimination	Sequence -based discrimination
Population studies	Uses allele frequencies	Uses frequencies for sequence-based alleles

Fig. 6 Similarities between CE and NGS/MPS methodology



the samples are barcoded. Available algorithms or software can determine the sequence and length-based polymorphisms (Woerner et al., 2017). Table 1 shows differences between CE and NGS technologies, whereas Figs. 6 and 7 depict the similarities between methodologies adopted in the processes.

Forensic Applications of NGS Technology

The application of massively parallel sequencing is increasing in forensic DNA analysis as the crime investigation laboratories are looking for methodologies to obtain maximum information from a trace or degraded forensic sample (Fig. 8) (Montano et al. 2018). Presently it is possible to map whole genomes with constantly increasing speed and decreasing costs (Børsting and Morling 2015). The true diversity in core forensic loci has been explored, thereby adding statistical weightage to the evidence.

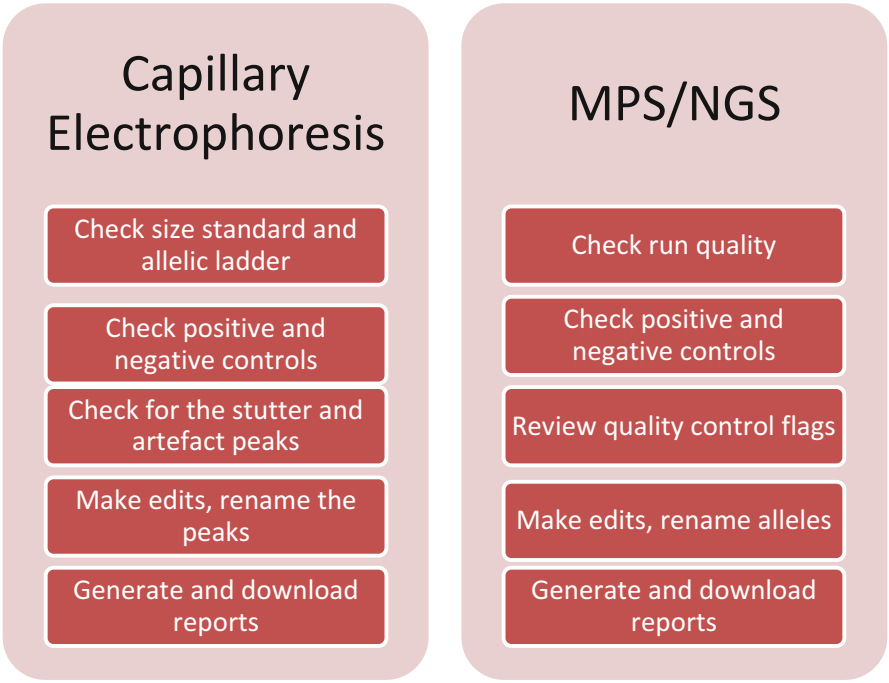


Fig. 7 Data analysis similarities between CE and NGS

STR and SNP Sequencing

CE explores length based, whereas NGS is sequence-based variation (Gettings et al. 2015). Multi-application sequencing has helped in achieving new heights. With NGS, it is possible to derive full sequence information (SNPs and InDels) within the STR loci to derive investigative leads and to estimate mutational events in kinship testing (Dalsgaard et al., 2013; Gettings et al., 2015). Insertion/deletion polymorphisms (InDels) which comprise the characteristics of both STRs and SNPs (Kidd et al. 2012) have now been used for forensic case work examination, databasing, and anthropological studies (Liu et al., 2020). SNPs and mitochondrial DNA (mtDNA) provide an effective accompaniment to traditional CE-STR analysis by increasing the amount and kind of genetic information that a single sample may yield. By putting together this information, a powerful investigative profile can be generated for use with missing persons and mass disaster victim identification (DVI) and to determine the number of contributors in a mixture (Petrovick et al. 2020) and cold cases. It is helpful particularly in those cases where traditional CE-based methods do not provide a full profile. Isometric alleles having identical size, but different sequence, could be identified. MPS offers high resolution as well as high power of discrimination, and degraded sample can still be used. Lots of sequence variation can be analyzed within an

Fig. 8 NGS facilitated forensic tools

allele. NGS technology can target just a few genes to all the nucleotides in a whole genome (Guo et al., 2017). A large number of panels have been developed for major as well as admixed populations. Successfully reliable and responsible sequencing of autosomal as well as Y-STR loci leads to conviction of a rape suspect by a court in Amsterdam, Netherlands. This case worldwide was the first case where conviction was based on MPS.

Mitochondrial DNA Sequencing in Forensics Using NGS

Mitochondrial genome has higher copy numbers as compared to nuclear DNA and is less prone to degradation. Degraded samples could yield good results using mt DNA. Challenging samples are difficult to analyze with nuclear DNA; forensically fruitful information can be obtained by means of mitochondrial (mt) DNA (Holland & Parsons, 1999). It is the only source of information from samples like hair shafts

where nuclear DNA is generally depleted (Higuchi et al., 1988). Mt. DNA analysis is an indispensable tool in forensic DNA examination (Holland et al., 2019); Mt. DNA is important for lineage, ancient DNA applications, and maternal inheritance for bio-ancestry. Capillary electrophoresis-based Sanger sequencing has been the gold standard for the last few decades (Berglund et al., 2011). But the technique is time-consuming and laborious and can focus only a portion of mitochondrial genome, i.e., HVI and HVII. Detection of heteroplasmy and occurrence of two or more mitochondrial genotypes in a cell can increase the discrimination power of analysis to a great extent. Analyzing heteroplasmy with Sanger sequencing is exceedingly difficult. Introduction of NGS technologies has revolutionized the arena of genomics (Kircher et al., 2012;). Last few decades have witnessed the utility of varied NGS technologies in assessing mtDNA, viz., Roche's 454 (Payne et al. 2013; Illumina's GAII (Li et al. 2010), Illumina's HiSeq 2000 (Tang et al. 2013), and Ion Torrent's Personal Genome Machine (PGM) (Parson et al. 2013). MPS addresses the limitations of traditional sequencing and sequences the entire mtDNA in one reaction without consuming much sample. MPS has large multiplex panels which makes it feasible to sequence the entire mitochondrial genome. With MPS complete, mtDNA sequence can be deduced from hair samples that come across in forensic case work (Parson et al. 2015). For aged and ancient genetic material, quantification and estimation of the level of degradation is the correct approach. In the highly degraded sample where nuclear DNA profile is not possible, NGS/SNP and mitochondrial DNA is the best alternate method. Tsunami victims case where high temperature and moisture resulted in degradation, the Russian Czar family missing children, and missing Mexican students of 2014 where apparently the individuals were burnt and killed and the remains were not amenable to DNA typing were some of the cases solved using mitochondrial analysis.

DNA Intelligence

DNA intelligence or forensic DNA phenotyping (FDP) facilitates the prediction of bio-geographic ancestry and external visible characteristics of the donors of forensic samples (Fig. 9). This sort of intelligence administers priceless investigative leads in cases where database matches do not yield results from STR analysis (Phillips et al., 2019).

Forensic DNA Phenotyping (FDP): Visible Phenotype Estimation

STR analysis fails to identify a person whose profile is unknown to the analysts. The drawbacks of comparative STR analysis lead to the birth of a new field in forensic genetics, i.e., forensic DNA phenotyping (FDP) (Kayser and De Knijff; 2011). Forensic DNA Phenotyping implies to the prediction of human appearances from traits such as hair color, eye color, and skin color from unknown crime scene samples

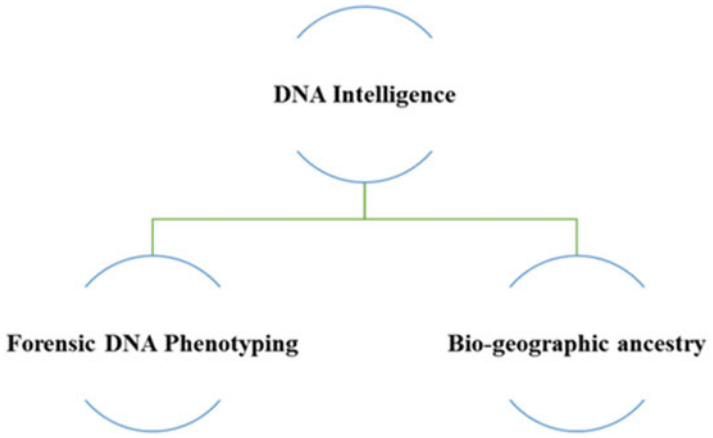


Fig. 9 DNA intelligence

(Kayser 2015). Investigative leads can be obtained from unknown sample donors, unidentified with the current CE technology. The technology can be of great help in disaster victim identification and missing person identification cases.

Ancestry Informative Markers

Bio-geographic ancestry (BGA) can be predicted using ancestry informative markers (AIM) (Xavier et al., 2020). During the course of the past few years, a large number of panels of the ancestry informative markers have been recommended for analyzing population genetic structure (Jiang et al. 2018). Also, ancestry can be established from the components of mixtures.

Degraded Samples

Conventional forensic analysis has a limitation of allele size with respect to degraded samples. To overcome this, shorter markers (MiniSTRs) were adopted (Martín et al., 2006). Currently SNPs have been incorporated in sequencing panels (Gettings et al., 2015). In CE, as a ski slope is obtained for a degraded sample, but NGS is not size based, so interpretation of results is easier, and also it also works with less amount of sample.

Identifying Monozygotic Twins

Monozygotic twins, having the same genetic structure, cannot be differentiated by conventional techniques like STR, SNP, and mitochondrial DNA analysis.

Identification of extremely rare mutations by NGS can differentiate between monozygotic twins (Weber-Lehmann et al., 2014).

Emerging Applications of NGS

Species origin of a sample, age range of contributors, metagenomics or human microbiome analysis, and methylation analysis are some of the emerging applications of NGS. Microbiome gives unique identity to an individual and can be used to determine the site of origin of the sample (Tozzo et al., 2020). NGS has a wide variety of application in body fluid and tissue identification via mRNA and methylation studies (Ingold et al. 2018). Studies have reported the use of miRNAs (microRNAs), a group of small noncoding RNAs having applicability in age prediction for body fluids or crime scene stains using MPS (Fang et al. 2020). NGS is also useful in wildlife forensics.

Utility of NGS/MPS over CE

The analysis of conventional STR markers using MPS provide a number of benefits over standard CE analysis, namely, particularly increased number of loci, higher discrimination power, and shorter amplicon length for a conclusive analysis of degraded and trace DNA evidence:

- Low DNA samples.
- Male minor contributor.
- Degraded samples.
- Unknown tissue origin.
- Cold cases.
- Suspect untraceable cases.
- Failed CE cases.
- No STR profile match in DNA database.
- Cold homicide cases.
- Mixture interpretation.
- Complex kinship cases or familial search.

The technique has certain drawbacks also, such as being expensive, lacking standardization, requiring huge resources for bioinformatics, and data storage besides accumulating sensitive personal data which is difficult to protect as entire genome including the coding regions can be sequenced. Although CE has a limitation of panels being limited by size ranges and fluorescent labels, still CE-based STR typing will remain the standard casework and database application as it is cheap, fast, and reliable and can be undertaken on regular basis, whereas MPS-based STR typing represents a specialized tool in forensics and can be used for specific cases. It involves time taking, tedious steps in library and DNA template preparation,

effective bioinformatics tools for sequence alignment, efficient data storage servers for the storage of sequence data files, and lack of standardization. The said shortcomings are being worked upon and do not seem to be a major long-term issue (Alonso et al. 2018). The technology guarantees endless advancement and may soon become the standard benchmark of quality. Today NGS platforms are being used not only in forensics (Fordyce et al. 2011) but also in microbial (Caporaso et al. 2012) and cancer research (Kirsch and Klein 2012).

Forensic Genetic Genealogy: The New Means of Genetic Identification

DNA analysis is being widely used in human identification, missing person/disaster victim identification and kinship testing (Fig. 10). A CE-STR based analysis involves uploading a crime scene profile to a database to obtain a hit or match (Fig. 11), whereas forensic genealogy is based on SNP testing and uploading the results to genealogy database to measure the genetic relatedness (Fig. 12).

Genetic genealogy has helped solve dozens of cold cases that would not have been cracked otherwise (Greytak et al., 2019). Matching is done by searching for DNA segments which are shared among the relatives. Genome-wide SNP data is used to measure the genetic relatedness (Fig. 13) (Kling and Tillmar 2019). Genetic genealogy though is not linked to CODIS but cases end the same way. In its evolutionary stage, genetic genealogy started with Y STR analysis. For the last 2 years, over 100 cases have been solved that have not been solved over decades. It has proved to be extremely accurate. Family trees have been constructed to find birth

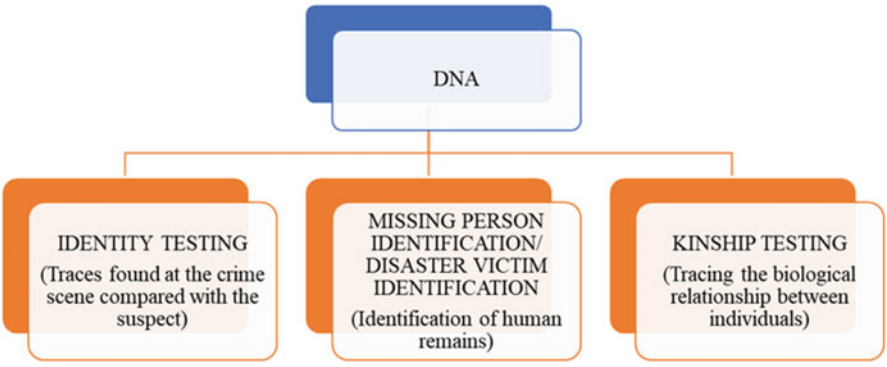


Fig. 10 Applications of DNA in legal field

Fig. 11 Flow chart for CE-STR-based analysis

Fig. 12 Flow chart for forensic genetic genealogy

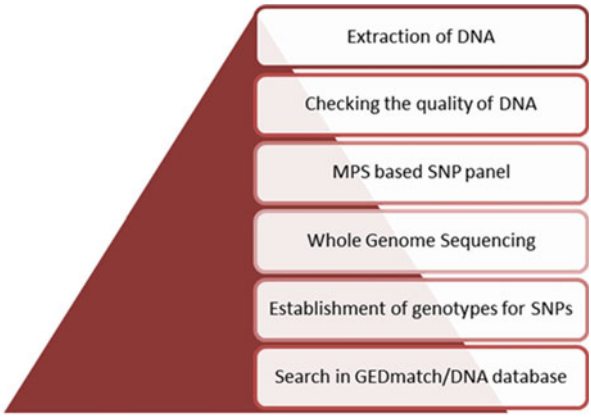


Fig. 13 From remains to database search: protocol for FGG

families by genealogists. Reference samples are compared with evidence material to confirm identity.

Cold cases, missing person identification, unsolved heinous crimes, and exoneration of the innocent can be solved by forensic genetic genealogy. Voluntarily submitted DNA data for forensic purposes is being used by "Genealogy Data Matching" (GEDmatch, Verogen, Inc.) software to generate investigative leads. GEDmatch allows the user to upload and compare SNP data in database. Autosomal profiles are uploaded to GEDmatch, to find out a potential match, and most recent common ancestors of a suspect were found on family tree, although there still exist issues in genetic genealogy relating to privacy notwithstanding the advancements (Court S. Denise, 2018; Wickenheiser, 2019). Moreover, the technique is expensive, time intensive, and resource intensive as is outsourced by forensic agencies and requires specialized knowledge. Low-quality uploads result in false genealogy connections. A cold case, a 16-year-old double homicide case, was solved by FGG in Sweden. SNP profiles were generated from the crime scene samples using whole genome sequencing and advanced bioinformatics tools. Relatives were searched in GEDmatch, and family tree DNA databases were constructed ultimately giving leads. With the arrival of IGG, alternate route for DNA analysis based on SNPs has been described. The following are the technical requirements for SNP testing:

- At least 1 ng of DNA.
- Single source DNA samples preferred.
- Degradation of samples could be a problem.

- DNA quality.
- Partial SNP profile.
- Success is genealogy database dependent.

Two most famous cases solved by this path breaking technique have been discussed below:

The Ekeby Man Case

The technique puts forward a great challenge for low copy number degraded DNA as in the case of forensic samples. Progression in DNA sequencing methodology has enabled the probability to process low copy number degraded biological samples (Prüfer et al., 2014). One such example is to identify an unknown male remains ("the Ekeby man case") found murdered in Sweden in 2003, where whole genome sequencing was done on a bone sample along with bioinformatics tools which generated around 1.4 million SNPs (Tillmar et al. 2020). The SNP genotypes were searched for relatives on DNA database GED match. A list of relatives was prepared to identify the unknown remains, and investigative leads were obtained.

The Golden State Killer

The man known as the Golden State killer (De Angelo serial killer, 74 years now) was sentenced to multiple life sentences for dozens of crimes he committed. He was convicted for the rapes and killings which he committed from 1975 to 1986 covering a wide geographical area that was thought initially because of multiple people. He began with home burglaries before committing many rapes and killings across southern California. He often broke into people's homes at midnight and carry out rapes and killings. The crimes mysteriously ended in 1986. The heinous crimes committed might have gone unsolved without the innovative genetic technique that was developed to reunite adoptees with their biological parents. DNA recovered from one of the crime scenes was put into the DNA database to find the relatives of the killer. A common ancestor among them was found, and family trees down to the present day were created (Phillips, 2018). De Angelo appeared as a possible suspect. Eventually an item containing De Angelo's DNA was found by the investigators and compared with the DNA recovered from the crime scenes, and a match was obtained. He was arrested in April 2018. Dozens of victims testified before the court. He was sentenced to life imprisonment without parole on August 21, 2020.

The Rapid DNA Instrument

Generating DNA profile from capillary electrophoresis involves a large number of tedious steps. The setup required for performing STR analysis includes centrifuge machines, thermal cyclers, and capillary electrophoresis instrumentation in a

centralized laboratory consuming at least 10 hours. Advancements in instrumentation have led to automation during the last few years (Hopwood et al. 2010). Robotic platforms (Frégeau et al., 2010) have reduced hands-on time. The rapid DNA instruments combine the steps of isolation, faster amplification, denaturation, sizing, and genotyping (Fig. 14). RapidHIT® 200 and the RapidHIT ID Systems have been developed by IntegenX for generating STR profiles from reference samples in nearly 90 minutes. This also reduces the chance of contamination between the samples (Pleasanton, CA) (Shackleton et al., 2019). One could employ such kind of a unified instrument for reference samples with the least amount of time required by the analyst (Dash et al., 2020). This kind of technique not only reduces the risk of contamination but has the added advantage of swab reusability for conventional DNA methods. Such kind of instruments can be used by the law enforcement personnel in the police stations, in the crime investigation labs for increased lab productivity, and in the field as well according to the need.

The RapidHIT™ ID system (Thermo Fisher Scientific) is a fully automated system for human identification. Run information is transferred to Rapidlink™ software to process. Reagents used for each instrument can be traced at a central location. All the data is fed into computer, and any DNA profile can be matched. Quality flags/colored flags can tell about the run quality. One person at a central position can see all the data coming from all the instruments and can address the problem. Data quality is monitored at one workstation by the software. The system is cartridge based with two kinds of cartridges, viz., primary cartridge and sample cartridge:

1. Primary cartridge contains polymers, capillary, and buffers.
2. Sample cartridge are of two types the Ace GlobalFiler Express sample cartridge for single source reference samples and the Rapid INTEL™ sample cartridge for forensic samples such as bones (Buscaino et al. 2018).

The developmental validation studies for this system were performed with mock casework samples according to the Scientific Working Group on DNA Analysis Methods (SWGDAM) guidelines (Scientific Working Group on DNA Analysis

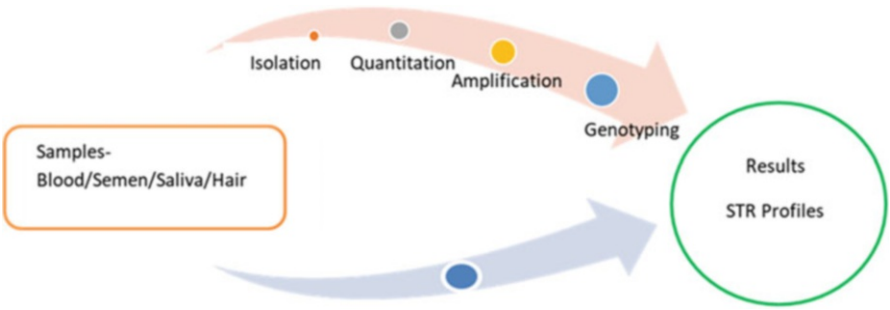


Fig. 14 Rapid hit system

Methods Validation Guidelines for DNA Analysis Methods. December 2016) (User Bulletin, "RapidINTEL™ Sample Cartridge for blood and saliva samples, Thermo Fisher Scientific). The samples are analyzed on the GeneMarker®HID STR Human Identity software on the instrument (Holland & Parson, 2011). Samples that require fast results may benefit from the rapid platform. Numerous Rapid Hit systems linked with RapidLink software can prove to be useful in criminal investigations.

Conclusion

Rapid DNA examination will soon facilitate new applications (Butler 2015). Massively parallel sequencing (MPS) provides the potential to multiplex diversified forensically relevant markers and multiple samples together in a single run in comparison with traditional capillary electrophoresis method (Churchill et al., 2016). Sequence based allele frequency data with the establishment of within STR allele sequence variants, will aid forensic community to enhance the power of discrimination for human identification and mixture deconvolution by increasing the effective allele number (Gaag et al. 2016) and kinship analysis. Forensic DNA phenotyping (FDP) and bio-geographic ancestry (BGA) from an unknown crime scene sample are gaining interest of the forensic community (Xavier et al., 2020). Whole genome sequencing has been successfully used to constitute genealogy DNA databases for generating investigative leads in cold cases (Tillmar et al. 2020). DNA intelligence marks a considerable peculiar application of genetic evidence unlike the one presented in the courtroom (Kayser 2015). NGS is an up-and-coming technology and is increasingly being implemented in forensic case work examination.

References

Alonso A, Barrio PA, Müller P, Köcher S, Berger B, Martin P, Bodner M, Willuweit S, Parson W, Roewer L, Budowle B (2018) Current state-of-art of STR sequencing in forensic genetics. Electrophoresis 39;(21):2655–2668

Berglund EC, Kiialainen A, Syvänen AC (2011) Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet 2(1):23

Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18:78–89

Buscaino J, Barican A, Farrales L, Goldman B, Klevenberg J, Kuhn M, Lin F, Nguyen P, Salceda S, Schueren R, Smith C, Troup C, Tsou D, Vangbo M, King D (2018) Evaluation of a rapid DNA process with the RapidHIT® ID system using a specialized cartridge for extracted and quantified human DNA. Forensic Sci Int Genet 34:116–127

Butler JM (2015) The future of forensic DNA analysis. Philos Trans R Soc B Biol Sci 370 (1674):20140252

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. Int Soc Microb Ecol 6(8):1621–1624

Chakravarty M, Pandya P, Raina A, Priyanka (2019) Resourcefulness of Str markers in forensic casework analysis. Int J Med Toxicol Legal Med 22(3–4):110–113

Churchill JD, Schmedes SE, King JL, Budowle B (2016) Evaluation of the Illumina® Beta version ForenSeq™ DNA signature prep kit for use in genetic profiling. Forensic Sci Int Genet 20:20–29
Coble MD, Bright JA (2019) Probabilistic genotyping software: an overview. Forensic Sci Int Genet 38:219–224
Court S. Denise (2018) Forensic genealogy: some serious concerns. Forensic Sci Int Genet 36:203–204
Dalsgaard S, Rockenbauer E, Gelardi C, Børsting C, Fordyce SL, Morling N (2013) Characterization of mutations and sequence variations in complex STR loci by second generation sequencing. Genet Suppl Serie Forensic Sci Int 4(1):e218–e219
Dash HR, Pankaj S, Surajit D (2020) RAPID DNA technology: a boon to forensic DNA. Typing 313–316
Dixit S, Pankaj S, Kumawat RK, Kaitholia K, Dash HR, Sharma H, Choubey G (2019) Forensic genetic analysis of population of Madhya Pradesh with PowerPlex fusion 6C ™ multiplex system. Int J Legal Med133(3):803–805
Fang C, Xu L, Jing Z, Bingbing X, Jialin Q, Wenli l, Baoming L, Xiaoli Z, Huijuan W, Jiangwei Y (2020) Age estimation using bloodstain MiRNAs based on massive parallel sequencing and machine learning: a pilot study. Forensic Sci Int Genet 47:102300
Fordyce SL, Ávila-Arcos MC, Rockenbauer E, Børsting C, Frank-Hansen R, Petersen FT, Willerslev E, Hansen AJ, Morling N, Gilbert TP (2011) High-throughput sequencing of Core STR loci for forensic genetic investigations using the Roche genome sequencer FLX platform. BioTechnique 51(2):127–133
Frégeau CJ, Marc Lett C, Fourney RM (2010) Validation of a DNA IQ™-based extraction method for TECAN robotic liquid handling workstations for processing casework. Forensic Sci Int Genet 4(5):292–304
Fullwood MJ, Wei CL, Liu ET, Ruan Y (2009) Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. Genome Res 19(4):521–532
Gaag VD, Kristiaan J, De Leeuw RH, Hoogenboom J, Patel J, Storts DR, Laros JFJ, De Knijff P (2016) Massively parallel sequencing of short tandem repeats – population data and mixture analysis results for the PowerSeq™ system. Forensic Sci Int Genet 24:86–96
Gettings KB, Aponte RA, Kiesler KM, Vallone PM (2015a) The next dimension in STR sequencing: polymorphisms in flanking regions and their allelic associations. Genet Suppl Serie Forensic Sci Int 5:e121–e123
Gettings KB, Kiesler KM, Vallone PM (2015b) Performance of a next generation sequencing SNP assay on degraded DNA. Forensic Sci Int Genet 19:1–9
Greytak EM, Moore CC, Armentrout SL (2019) Genetic genealogy for cold case and active investigations. Forensic Sci Int 299:103–113
Guo F, Yu J, Lu Z, Li J (2017) Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA signature prep kit on the MiSeq FGx™ forensic genomics system. Forensic Sci Int Genet 31:135–141
Heather JM, Chain B (2016) The sequence of sequencers: the history of sequencing DNA. Genomics 107(1):1–8
Higuchi R, Von Beroldingen CH, Sensabaugh GF, Erlich HA (1988) DNA typing from single hairs. Nature 332(6164):543–546
Holland MM, Parsons TJ (1999) Mitochondrial DNA sequence analysis – validation and use for forensic casework. Forensic Sci Rev 11(1):21–50
Holland MM, Bonds RM, Holland CA, McElhoe JA (2019) Recovery of MtDNA from unfired metallic ammunition components with an assessment of sequence profile quality and DNA damage through MPS analysis. Forensic Sci Int Genet 39:86–96
Holland MM, Parson W (2011) GeneMarker® HID: a reliable software tool for the analysis of forensic STR data. J Forensic Sci 56(1):29–35
Hopwood AJ, Hurth C, Yang J, Cai Z, Moran N, Lee-Edghill JG, Nordquist A, Lenigk R, Estes MD, Haley JP, McAlister CR, Chen X, Brooks C, Smith S, Elliott K, Koumi P, Zenhausern F, Tully G (2010) Integrated microfluidic system for rapid forensic DNA analysis: sample collection to DNA profile. Anal Chem 82(16):6991–6999
Hyman ED (1988) A new method of sequencing DNA. Anal Biochem 174(2):423–436

Ingold S, Dørum G, Hanson E, Berti A, Branicki W, Brito P, Elsmore P, Gettings KB, Giangasparo F, Gross TE, Hansen S, Hanssen EN, Kampmann ML, Kayser M, Laurent FX, Morling N, Mosquera-Miguel A, Parson W, Phillips C, Porto MJ, Pośpiech E, Roeder AD, Schneider PM, Schulze Johann K, Steffen CR, Syndercombe-Court D, Trautmann M, van den Berge M, van der Gaag KJ, Vannier J, Verdoliva V, Vidaki A, Xavier C, Ballantyne J, Haas C (2018) Body fluid identification using a targeted MRNA massively parallel sequencing approach – results of a EUROFORGEN/EDNAP collaborative exercise. Forensic Sci Int Genet 34:105–115

Jiang L, Wei YL, Zhao L, Li N, Liu T, Liu HB, Ren LJ, Li JL, Hao HF, Li Q, Li CX (2018) Global analysis of population stratification using a smart panel of 27 continental ancestry-informative SNPs. Forensic Sci Int Genet 35:e10–e12

Kayser M (2015) Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet 18:33–48

Kayser M, De Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nature Rev Genet 12(3):179–192

Kidd KK, Kidd JR, Speed WC, Fang R, Furtado MR, Hyland FCL, Pakstis AJ (2012) Expanding data and resources for forensic use of SNPs in individual identification. Forensic Sci Int Genet 6 (5):646–652

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res 40(1):e3–e3

Kirsch S, Klein CA (2012) Sequence error storms and the landscape of mutations in Cancer. Proc Nat Acad Sci U S A 109(36):14289–14290

Kling D, Tillmar A (2019) Forensic genealogy—a comparison of methods to infer distant relationships based on dense SNP data. Forensic Sci Int Genet 42:113–124

Li M, Schönberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M (2010) Detecting Heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. Am J Hum Genet 87(2):237–249

Liu J, Weian D, Wang M, Liu C, Wang S, He G, Zheng W (2020) Forensic features, genetic diversity and structure analysis of three Chinese populations using 47 autosomal InDels. Forensic Sci Int Genet 45:102227

Ma Y, Kuang JZ, Nie TG, Zhu W, Yang Z (2016) Next generation sequencing: improved resolution for paternal/maternal duos analysis. Forensic Sci Int Genet 24:83–85

Mardis ER (2017) DNA sequencing technologies: 2006–2016. Nat Protoc 12(2):213–218

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Lei D, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density Picolitre reactors. Nature 437(7057):376–380

Martín P, García O, Albarrán C, García P, Alonso A (2006) Application of mini-STR loci to severely degraded casework samples. Int Cong Series 1288:522–525

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. Proc Nat Acad Sci 74(2):560–564

McCarthy JJ, McLeod HL, Ginsburg GS (2013) Genomic medicine: a decade of successes, challenges, and opportunities. Sci Trans Med 5(189):189sr4–189sr4

Montano EA, Bush JM, Garver AM, Larijani MM, Wiechman SM, Baker CH, Wilson MR, Guerrieri RA, Benzinger EA, Gehres DN, Dickens ML (2018) Optimization of the Promega PowerSeq™ auto/Y system for efficient integration within a forensic DNA laboratory. Forensic Sci Int Genet 32:26–32

Müller P, Alonso A, Barrio PA, Berger B, Bodner M, Martin P, Parson W (2018) Systematic evaluation of the early access applied Biosystems precision ID Globalfiler mixture ID and Globalfiler NGS STR panels for the ion S5 system. Forensic Sci Int Genet 36:95–103

Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE (2011) Landscape of next-generation sequencing technologies. Anal Chem 83(12):4327–4341

Parson W, Huber G, Moreno L, Madel MB, Brandhagen MD, Nagl S, Xavier C, Eduardoff M, Callaghan TC, Irwin JA (2015) Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples. Forensic Sci Int Genet 15:8–15

Parson W, Strobl C, Huber G, Zimmermann B, Gomes SM, Souto L, Fendt L, Delport R, Langit R, Wootton S, Lagacé R, Irwin J (2013) Evaluation of next generation MtGenome sequencing using the ion torrent personal genome machine (PGM). Forensic Sci Int Genet 7(5):543–549

Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels DC, Santibanez-Koref M, Chinnery PF (2013) Universal Heteroplasmy of human mitochondrial DNA. Hum Mol Genet 22(2):384–390

Petrovick MS, Boettcher T, Fremont-Smith P, Peragallo C, Ricke DO, Watkins J, Schwoebel E (2020) Analysis of complex DNA mixtures using massively parallel sequencing of SNPs with low minor allele frequencies. Forensic Sci Int Genet 46:102234

Phillips C, McNevin D, Kidd KK, Lagacé R, Wootton S, de la Puente M, Freire-Aradas A, Mosquera-Miguel A, Eduardoff M, Gross T, Dagostino L, Power D, Olson S, Hashiyada M, Oz C, Parson W, Schneider PM, Lareu MV, Daniel R (2019) MAPlex – A massively parallel sequencing ancestry analysis multiplex for Asia-Pacific populations. Forensic Sci Int Genet 42:213–226

Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, Calaza M, Casares de Cal M, Ballard D, Lareu MV, Carracedo Á (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 1(3–4):273–280

Phillips C (2015) Forensic genetic analysis of bio-geographical ancestry. Forensic Sci Int Genet 18:49–65

Phillips C (2018) The Golden state killer investigation and the nascent field of forensic genealogy. Forensic Sci Int Genet 36:186–188

Poinar HN, Schwarz C, Qi J, Shapiro B, MacPhee RDE, Buigues B, Tikhonov A, Huson DM, Tomsho LP, Auch A, Rampp M, Miller M, Schuster SC (2006) Metagenomics to Paleogenomics: large-scale sequencing of mammoth DNA. Science 311(5759):392–394

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, De Filippo C, Li H, Mallick S, Dannemann M, Qiaomei F, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PLF, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Pääbo S (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505(7481):43–49

Riman S, Iyer H, Borsuk LA, Vallone PM (2020) Understanding the characteristics of sequence-based single-source DNA profiles. Forensic Sci Int Genet 44:102192

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with Chain-terminating inhibitors. Proc Nat Acad Sci U S A 74(12):5463–5467

Shackleton D, Pagram J, Andrews N, Malsom S, Ives L, Vanhinsbergh D (2019) Development of enhanced sensitivity protocols on the RapidHIT™ 200 with a view to processing casework material. Sci Justice 59(4):411–417

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26(10):1135–1145

Shrivastava P, Trivedi VB, Singh AK, Mishra N (2012) Application of DNA fingerprinting technology in Forensic Investigation. Int J Scient Res Publ 2(1):2250–3153

Srivastava V, Prabhakar Rao K, Surekha Rani H, Kumawat RK, Mishra A, Shrivastava P (2020) Genomic diversity in the Goud population of Telangana, India inferred using twenty-three autosomal marker PowerPlex® fusion 6C system. Meta Gene 25:100718

Tang S, Wang J, Zhang VW, Li FY, Landsverk M, Cui H, Truong CK, Wang G, Chen LC, Graham B, Scaglia F, Schmitt ES, Craigen WJ, Wong LJC (2013) Transition to next generation analysis of the whole mitochondrial genome: a summary of molecular defects. Hum Mutat 34(6):882–893

Tillmar A, Sjölund P, Lundqvist B, Klippmark T, Älgenäs C, Green H (2020) Whole-genome sequencing of human remains to enable genealogy DNA database searches – a case report. Forensic Sci Int Genet 46:102233

Tozzo P, D'angiolella G, Brun P, Castagliuolo I, Gino S, Caenazzo L (2020) Skin microbiome analysis for forensic human identification: what do we know so far? Microorganisms 8(6):873

Weber-Lehmann J, Schilling E, Gradl G, Richter DC, Wiehler J, Rolf B (2014) Finding the needle in the haystack: differentiating 'identical' twins in paternity testing and forensics by ultra-deep next generation sequencing. Forensic Sci Int Genet 9:42–46

Wickenheiser RA (2019) Forensic genealogy, bioethics and the Golden state killer case. Forensic Sci Int Synergy 1:114–125

Woerner AE, King JL, Budowle B (2017) Fast STR allele identification with STRait razor 3.0. Forensic Sci Int Genet 30:18–23

Xavier C, de la Puente M, Phillips C, Eduardoff M, Heidegger A, Mosquera-Miguel A, Freire-Aradas A, Lagace R, Wootton S, Power D, Parson W, Lareu MV, Daniel R (2020b) Forensic evaluation of the Asia Pacific ancestry-informative MAPlex assay. Forensic Sci Int Genet 48:102344

Xavier C, de la Puente M, Mosquera-Miguel A, Freire-Aradas A, Kalamara V, Vidaki A, Gross TE, Revoir A, Pośpiech E, Kartasińska E, Spólnicka M, Branicki W, Ames CE, Schneider PM, Hohoff C, Kayser M, Phillips C, Parson W (2020a) Development and validation of the VISAGE AmpliSeq basic tool to predict appearance and ancestry from DNA. Forensic Sci Int Genet 48:102336

Yang Y, Xie B, Yan J (2014) Application of next-generation sequencing Technology in Forensic Science. Genom Proteom Bioinform 12(5):190–197

Biological Sources of DNA: The Target Materials for Forensic DNA Typing

3

Pankaj Shrivastava, R. K. Kumawat, Pushpesh Kushwaha, and Manisha Rana

Contents

Abstract

Among physical evidences encountered at the crime scene, biological evidences, viz., blood, semen, vaginal secretion, saliva, urine, and sweat, are the most ubiquitous in nature, and their presence aids in linking perpetrator to the victim as well as crime scene. Recognition of biological fluids as substantive evidences

P. Shrivastava (✉)
DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

R. K. Kumawat
DNA Division, State Forensic Science Laboratory, Jaipur, India

P. Kushwaha
Dr. APJ Abdul Kalam Institute of Forensic Science and Criminology, Bundelkhand University, Jhansi, India

DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

M. Rana
Department of Forensic Science, Guru Ghasidas University, Bilaspur, India

DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_2

is endorsed by presence of DNA in them. Advancing forensic DNA typing techniques have a great potential in characterization and individualization of biological evidences encountered during criminal investigation, but their applicability on each biological fluid for human identification varies in a great deal due to variation in the amount of nucleic acid available within the fluid. DNA concentration is relatively high in fluids such as blood that contains large number of cells while fluids such as urine and sweat possess low amount of DNA. Besides reference biological samples, forensic DNA typing can be implemented for detecting presence of traces of biological fluids on physical surfaces. The concept of "Touch DNA" or "Transfer DNA" involves analysis of low amount of DNA deposited on the surfaces that have come in human contact and can constructively help in associating evidences with perpetrator and/or victim. However, identification of sources of such materials, their collection methods, and preservation conditions can influence the quality of result. Further to this, degradation, impurity, contamination, and presence of inhibitors in such evidences demand purification and isolation of high-quality DNA. This chapter deals with various biological sources of DNA commonly encountered at the scene of crime and their evidential value, along with various factors and conditions affecting forensic DNA typing of such samples. This chapter deals with various biological samples used for forensic DNA typing along with various factors and conditions that affect forensic DNA typing.

Keywords

Biological fluids · Forensic DNA typing · Touch DNA · Crime Scene · Perpetrator

Introduction

Among physical evidences encountered at the crime scene, biological evidences, viz., blood, semen, vaginal secretion, saliva, urine, and sweat, are the most ubiquitous in nature, and their presence aids in linking perpetrator to the victim as well as crime scene. Recognition of biological fluids as substantive evidences is endorsed by presence of DNA in them. Advancing forensic DNA typing techniques have a great potential in characterization and individualization of biological evidences encountered during criminal investigation, but their applicability on each biological fluid for human identification varies in a great deal due to variation in the amount of nucleic acid available within the fluid. DNA concentration is relatively high in fluids such as blood that contains large number of cells while fluids such as urine and sweat possess low amount of DNA. Besides reference biological samples, forensic DNA typing can be implemented for detecting presence of traces of biological fluids on physical surfaces. The concept of "Touch DNA" or "Transfer DNA" involves analysis of low amount of DNA deposited on the surfaces that have come in human contact and can constructively help in associating evidences with perpetrator and/or victim.

However, identification of sources of such materials, their collection methods, and preservation conditions can influence the quality of result. Further to this, degradation, impurity, contamination, and presence of inhibitors in such evidences demand purification and isolation of high-quality DNA. This chapter deals with various biological sources of DNA commonly encountered at the scene of crime and their evidential value, along with various factors and conditions affecting forensic DNA typing of such samples.

Blood

Blood is one of the preeminent evidences that are alighted on at various crime scenes including cases of sexual assault, homicide, suicide, accidents, and burglary in disparate forms of blood-pool and blood stains adhered to the surfaces such as floor, walls, clothes, and the weapons involved in the crime. Whole blood is a composite of various blood components that are classified into liquid element, plasma, and cellular (or formed) elements, erythrocytes (or red blood cells), leukocytes (or white blood cells), and thrombocytes (or platelets).

Discovery of ABO blood typing in 1900 instigated the scope of human blood identification that further strode toward individualization in the 1980s with the advancements of molecular techniques. Existence of individual-specific minisatellites in the human genome that can aid in human identification cases was affirmed by analyzing blood samples of 20 unrelated individuals using southern blot hybridization technique (Jeffreys et al. 1985). Thereafter several protocols and their modifications with varied incubating reagents, time frame, and techniques were designed and reported for extraction and purification of human genomic DNA (Table 1).

A number of considerable pre-analytical factors influence the quantity and quality of genomic DNA extracted from the whole blood or clotted blood sample. Sample collection with zero or minimal contamination is an important consideration that affects the stability of blood sample. Use of sterilized syringe is preferred for collection of blood pools. Wet blood soaked objects are air-dried and collected; if object is inflexible, stains can be collected on sterilized cotton swabs and air-dried. Scraping or tape-lifting is employed for collection of dried blood stains. Stability of blood is influenced by other factors such as use of stabilizing agents such as heparin or anticoagulant agents such as EDTA, time difference between collection, and storage of sample (Vaught 2006). Storage period, storage conditions such as temperature of blood, as well as isolated DNA are shown in Table 2. Exposure to ultraviolet radiations, heat, light, humidity, and soil contaminations (McNally et al. 1989) have adverse effect on extraction of good quality of DNA.

Blood evidence in various conditions, viz., frozen blood, clotted blood, or dried blood spots, can also be found during forensic investigation. Such samples require additional pre-analytical treatment or modified protocols for extraction of DNA. Modification in conventional proteinase K/phenol chloroform isoamyl alcohol (PCIA) protocol with pre and post-trypsination of frozen blood samples with the

Table 1 Protocols and their modifications for DNA extraction from blood

S. no	Protocol	Material used	Technique used	Result	Reference
1	Rapid method of DNA isolation from human leukocyte	SDS for lysis, potassium acetate for precipitation	Diethylaminoethyl (DEAE) cellulose chromatography	50–70 µg of DNA in 10 ml of blood sample	Potter et al. (1985)
2	Rapid method for the purification of DNA from blood	Guanidine hydrochloride, ammonium acetate, sodium sarkosyl, and proteinase K	Southern blotting	20 µg of DNA in 1 mL of blood	Jeanpierre (1987)
3	Salting out procedure for extracting DNA from human nucleated cells	1 ml of saturated NaCl after digestion	Centrifugation at 2500 rpm for 15 min	DNA quantity comparable with that of phenol-chloroform protocol	Miller et al. (1988)
4	Non-organic procedure for the isolation of genomic DNA from blood	Ice cold CLB (0.32 M sucrose, 10 mM Tris-HCl pH 7.6, 5 mM $MgCl_2$, 1% Triton X-100)	Southern blotting	Isolation period less than 4 h	Grimberg et al. (1989)
5	Direct PCR from whole blood, without DNA extraction	Blood introduced directly to the PCR reaction of 50 mM KCl, 10 mM Tris-HCl pH 8.0, 1.5 Mm $MgCl_2$, 0.1 mg/ml gelatin, 200 uM each dNTP	3 PCR cycle of 3 min at 94 °C then cooled for 3 min at 55 °C	Convenient alternative to the tedious DNA extraction process	Mercier et al. (1990)
6	Isolation of fetal DNA from nucleated erythrocytes in maternal blood	Phosphate buffered saline (PBS: 0.137 M NaCl/0.002 M KCI/0.008 M Na_2HPO_4/ 0.0015 M KH_2PO_4, pH 7.4), 2% fetal calf serum, and 0.1% sodium azide on ice	TfR Analysis	0.1–1 ng of fetal DNA present in maternal blood at 15–16 weeks of gestation	Bianchi et al. (1990)

(continued)

Table 1 (continued)

S. no	Protocol	Material used	Technique used	Result	Reference
7	Alkaline extraction of Human Genomic DNA	5 µL of sample incubated with 20 µL 0.2 M NaOH at room temperature for 5 min in case of blood and at 75 °C in case of stain	Centrifugation at 12000xg for 5 min	30 ng of nuclear DNA per µL of blood	Dissing et al. (1996)
8	Modified salting out method using laundry detergent	Additional treatment with laundry powder solution, glass beads, and NaCl prior to DNA precipitation	Centrifugation at 15000 rpm for 5 min	30 mg/ml of powder yielded sufficient DNA 56.3 µg/mL	Nasiri et al. (2005)

Table 2 Storage conditions for blood samples

Storage condition	Maximum time	Reference
45 °C	6–7 weeks	Madisen et al. (1987)
23 °C	1 week	
−30 °C	12 years	Chen et al. (2018)
Dried blood spots on FTA cards	16 years	Rahikainen et al. (2016)
Blood stains at room temperature	15 years	Barbaro and Cormaci (2006)

inference that trypsination before cell lysis yielded DNA with 88.17% purity that declined to 63.23% in case of untrypsinized frozen blood (Ahmad et al. 1995). Another rapid protocol with a modified composition of cell lysis buffer and extraction buffer substituting the toxic reagents yielded DNA from frozen blood that was comparable with fresh blood (Guha et al. 2018). In case of clotted blood, mechanical shearing of clot by homogenization, scraping, or slicing (Xu et al. 2010) raised the quality of DNA. Use of nylon mesh and serum separator in various studies reports a good yield of DNA from clotted blood.

The texture, quality, and absorptivity of the surface acting as blood-stain carriers also influence the yield of DNA. Prinz and Berghaus (1990) successfully isolated sufficient amount of DNA from two days old dried blood stains on eleven different stain-carrier surfaces including variety of fabrics stores at room temperature. DNA isolated from surfaces such as wool, denim, suede, and carpet was chemically contaminated possibly due to surface-specific challenges in extracting leukocyte from the carrier.

A commonly encountered forensic situation is the deliberate removal of visible blood stains with cleaning agents that does not possibly obstruct the visualization of the stains but generate contamination of stain and degradation of genomic DNA (Tas 1990) (Thabet et al. 2018). Bleach has the most detrimental impact on the yield of

DNA as compared to other chlorinated and non-chlorinated detergents (Harris et al. 2006).

Bloodstains treated with fingerprint-enhancing and/or blood-enhancing reagents are another set of challenges for recovery of DNA. Fingerprint enhancement reagents such as Cyanoacrylate Fuming (Newall et al. 1996; Mutter et al. 2018), silver nitrate (Lee et al. 1989), and other bloody fingerprint enhancement chemicals such as luminol (3-aminophthalhyrazide) (Manna et al. 2000), benzidine, leucomalachite green (LMG), phenolphthalein KM, Amido Black (methanol based), Crowle's Double Stain, and Hungarian Red (fuchsin acid) (De Almeida et al. 2011; Frégeau et al. 2000; Everson et al. 1993; Tobe et al. 2007) have deteriorating impact on the quality and quantity of DNA. Yield of DNA is also reduced by exposure to short ultraviolet rays that are used for enhancing fingerprints found in the blood (Andersen and Bramble 1997).

DNA fingerprinting technique is influenced by diverse range of factors including ecological factors, improper collection techniques or mishandling during the chain of custody, as well as analytical procedure. The inference drawn from the well-known O. J. Simpson trial (1985) (Thompson 1996) elicited the necessity of improving the diligence of evidence collection and analysis by the forensic DNA laboratories as well as upgrading the presentation of DNA evidence in the court-room. The defense alleged on the collection and preservation of evidence more willingly than raising questions on the validity of DNA evidence by presenting evidence of negligence by Los Angeles Police Department (LAPD) in form of cross-contamination, switching of dried swatches, and premeditated planting of blood onto the evidences (Butler 2005).

Semen

Sexual assault cases contribute nearly 50% of the total cases received in the Forensic DNA Laboratories. The importance of DNA in sexual assault cases was known to be long ago since the "Pitchfork Case" of Leicestershire (1988), the first trial involving DNA analysis of semen for conviction of rape and murder. Semen, a viscous, slightly yellowish or grayish fluid mainly comprises of seminal fluid (made up of water, proteins, sugars, minerals, and vitamins) and sperm cells (spermatozoa). A typical ejaculation releases 2–5 mL of semen of which spermatozoa (50 μM in length) make up approximately 5% of the total volume of semen. Swabs, clothing, vaginal slides, and bedding items are generally collected for DNA analysis. For a successful STR (Short Tandem Repeats) analysis, nucleated cells with a sufficient amount of DNA are required, and once seminal fluid is detected on samples by applying different preliminary and confirmatory tests, the next step is DNA extraction.

DNA fingerprinting has two basic utilities in sexual assault cases. Firstly, the individualization of semen (when spermatozoa is less in number) and secondly is to differentiate mixed stains when the number of perpetrator is more than one. The situation becomes complex when contaminated samples are there or semen is mixed with other body fluids like blood. And here, the hemoglobin acts as an inhibitor in

the PCR (Polymerase Chain Reaction) process. Generally, semen is found to be mixed with vaginal secretion which cartons the presence of spermatozoa as the amount of vaginal secretion is much higher. Male DNA can be separated from the DNA of other cells using differential extraction method which was firstly detailed in the publication by Gill et al. (1985) and later by Wiegand et al. (1992) and Yoshida et al. (1995) (Table 3). Another difficulty which DNA experts face is when multiple males are contributors and two or more alleles may be present for one marker and in this situation chances of allele dropout can't be ruled out. If spermatozoa is absent in

Table 3 Differential extraction protocols and their modifications

Protocol for differential extraction	Reference
Sperm nuclei lysis with SDS, proteinase K, and DTT mixture	Gill et al. (1985)
Sample treated with Lysis buffer I (10 mM Tris-HCl, 0.4 M NaCl, 2 mM EDTA) with proteinase K and SDS and incubated at 37 °C for 40 min followed by centrifugation at 13000 rpm for 5 min. Supernatant lysed with lysis buffer II (proteinase K, SDS, and DTT), incubated at 56 °C for 1 h, centrifuged at 13000 rpm for 5 min, and DNA isolated using organic extraction protocol	Wiegand et al. (1992)
First incubation with TNE buffer, 1% SDS and proteinase K at 70 °C for 3 h, followed by centrifugation at 15000 rpm for 5 min. Second incubation with TNE buffer with 1% SDS, 100 ml proteinase K, and 0.04 M dithiothreitol (DTT)) for more than 8 h at 56 °C in a shaking water bath	Yoshida et al. (1995)
Tris(2-carboxyethyl)phosphine (TCEP) in Triton X-100 as lysis agent to lyse sperm cells and collect DNA on-chip, incubated for 15 min followed by addition of 40 µL of proteinase K solution (1 µg mL − 1) and incubation for 4 h at 55 °C. 100 µL of Buffer AL and 100 µL of ethanol were added to the samples, mixed by vortexing and run through gDNA extraction using a Qiagen spin column protocol	Inci et al. (2018)
Sample treated with extraction buffer (10 mM Tris-HCl (pH 8.0), 100 mM sodium chloride, 1 mM EDTA (pH 8.0), TNE, 1% SDS, and 0.2 mg/ml proteinase K) and incubated for 2 h at 37 °C followed by centrifugation of samples in spin baskets at 18000 g for 5 min, separation of supernatant, and multiple washing of pallet. Pallet lysed with sperm extraction buffer (10 mM Tris-HCl (pH 8.0), 100 mM sodium chloride, 1 mM EDTA (pH 8.0), 2.5% sarkosyl, 0.39 M dithiothreitol, and 0.5 mg/ml proteinase K) and incubated at 37 °C for 2 h. DNA purified from epithelial and sperm portion using Qiagen EZ1 Advanced XL system	Alderson et al. (2018)
Sample lysis with stain extraction buffer (1 mol/L Tris–HCl, ddH2O, 5 mol/L NaCl, 0.5 mol/L EDTA, 10% SDS, pH ¼8.0) and 15 mL of proteinase K (20 mg/mL) (Fisher Scientific, Pittsburgh, PA, USA) followed by an overnight 56 C incubation. Isolation of lysate (non-sperm fraction) using DNA IQTM spin baskets (Promega Corporation, Madison, WI, USA) at 7500 g spin for 5 min. The sperm pellet was subsequently resuspended in 200 mL of phosphate buffered saline solution (Fisher Scientific), 20 mL of Qiagen proteinase K stock solution, and 20 mL of 1 mol/L DTT (Fisher Scientific), vortexed, 200 mL of Buffer AL (Qiagen) was added, and samples were incubated at 56 °C isolation with QIAamp DNA Investigator kit. DNA was eluted in final volumes of 100 mL (nonsperm fractions) or 60 mL (sperm fractions) of Buffer ATE (Qiagen)	Goldstein et al. (2019)

semen (oligospermic, azoospermic, or normospermic), Positive Semelogenin (Sg) samples may be suitable. Y-chromosome identification and Sg biomarker should be thoroughly examined in the laboratories (Martínez et al. 2015).

The condition of the exhibit before examination remains to be a censorious aspect for successfully detecting and analyzing semen, and for this appropriate handling parameters during drafting, collecting, packaging, storing, and transportation of samples are the fundamental strides.

Vaginal Secretion

Another important evidence found in cases of sexual assault is vaginal secretion. In these cases, the identification of vaginal secretion is crucial as it can support in verifying the allegations of sexual assault. For instance, a stain is observed during investigation of a sexual assault case, and forensic DNA analysis affirms that the stain is originated from the victim, thus creating a link among the suspect and the victim. But, the litigant may contradict any criminal act by claiming that the stain is originated from sweat due to spontaneous contact. So, the evidence would have a significant value if vaginal secretion stain was found. A mixture of vaginal secretion and semen stain is generally found, and the presence of vaginal secretion confirms the incidence of sexual assault.

Human vagina is composed of squamous mucosa (comprises of stratified squamous epithelial tissue), submucosa, and muscularis. The vaginal secretion basically consists of epithelial debris, tissue fluid, leukocytes, electrolytes, lactic acid, and proteins which is generally derived from the glands of the uterus, cervix, transudation of the vaginal epithelium, and Bartholin's glands. Forensic laboratories use various methods for identification of vaginal secretion, and once it is identified, then DNA analysis is done. The main problem experts face is the ominous proportion of male to female DNA, with a surplus of the victim's material. Differential lysis is applied in this situation to isolate male DNA from epithelial cells (Gill et al. 1985).

Some other factors which should be kept in mind in order to get better results are the type of material used for collection and storage must be selected correctly. The contamination of genetic material from other sources (e.g., from the examiner and other biological evidence) should also be avoided (Butler 2005). Contamination may occur during the sexual contact (e.g., if there is more than one executioner), during collection and packaging, during transportation, during the medical examination, and in the laboratory. Proper care must be taken to restrict cross-contamination between sexual assault evidences.

Oral Fluids

Oral fluids or whole saliva is a mixture of secretion produced from major specific salivary glands, numerous minor salivary glands along with secretion from non-salivary sources such as nasal secretion, gingival crevicular fluid, bronchial

mucus, buccal cells, bacterial products, and food remains. Non-invasiveness, easy collection, and less contamination are some of the beneficial features of saliva over blood for DNA typing. Saliva are generally recovered from the scene of crime in conjunction with bite mark evidence on the skin as in cases of violent crimes, on eatables, clothing, cigarette butts, chewing gums, chewed betel quid stains, documents, postage stamps, and other objects (Anzai-Kanto et al. 2005; de Oliveira Musse et al. 2019). Despite frequent occurrence, quick drying of saliva stains makes them indiscernible, hampering their recognition and collection.

The earliest DNA isolation from the saliva on cigarette butts involved PCR amplification at HLA-DQ alpha and D1S80 markers and analysis by reverse dot-blot technique and polyacrylamide gel electrophoresis (Hochmeister et al. 1991). Similar study of DNA isolation from saliva and saliva-stained samples such as buccal swabs, gags, envelopes, and cigarettes stored at different conditions demonstrated identical DNA banding patterns as obtained from blood, hair, semen, or mixed saliva (Khare et al. 2014; Walsh et al. 1992) (Table 4). Watanabe et al. (2003) reported inhibitory impact of certain dyes present in the cigarette butts on PCR amplification. Sweet et al. (1996) were constantly involved in the studies related to DNA extraction from saliva in various conditions. In 1996, they proposed modified Chelex method involving pre-analytical use of proteinase K, incubation at 56 °C for 60 min, and 100 °C for 8 min and subsequent microconcentration of solution (Sweet et al. 1996). Saliva deposited on the skin is present in limited amount. Double swab technique ensures maximum collection of saliva stains with

Table 4 Salivary DNA extracted from different surfaces

Substrates	Number/type of samples	Conclusion	Reference
Betel quid	50	92% success rate for DNA isolation from 4 years old forensic BQ samples	Chiou et al. (2001)
Food	20 cheese pieces	Collection of saliva from the center instead of peripheral surface yielded better results	de Oliveira Musse et al. (2019)
	2 surfaces of cheese	Variation in DNA concentration recovered from upper and lower surface	Sweet and Hildebrand (1999)
Skin	5	Lower DNA recovered from skin probably due to degradation during saliva deposition, collection, and extraction	Anzai-Kanto et al. (2005)
	15	Double swab technique yielded better DNA quantity than filter paper and single swab technique	Sweet et al. (1997)
	A body submerged in water	Mixed DNA profile with minor component correlating with DNA profile of suspect	Sweet and Shutler (1999)
Cigarette butts	200	Deterioration in concentration of DNA with storage time, but all DNA were PCR amplifiable	Hochmeister et al. (1991)
	100	Inhibitory effects of dyes present in cigarette	Watanabe et al. (2003)

minimal contamination (Cherian et al. 2015). It involves swabbing of the skin surface with first swab immersed in sterile water in circular motion followed by second swabbing with dry swab using same pressure and motion (Sweet et al. 1997). In a drowning case, the bite mark present on the victim's body submerged for 5.5 h in water served as the source of salivary DNA, and DNA profile of the suspect was identified and distinguished at HUMTH01 and HUMvWA loci (Sweet and Shutler 1999).

Storage of saliva at −70 °C yields fair quality of DNA up to 1 month. Storage at 4 °C led to bacterial growth but yields sufficient PCR product (Ng et al. 2004). Quantity and quality of DNA from saliva stored at 4 °C and −20 °C for up to 3 months was comparable to that of fresh saliva samples. However, gradual deterioration was observed when the storage period was extended to 5 months (Kim and Kim 2006).

Trace quantity of saliva can also be transferred to the surfaces during speaking, coughing, and flipping pages of documents. Double swab technique for collection of such samples is preferred. Deposition of trace samples can also be the result of contamination at the crime scene by adventitious transfer or during investigation and analysis. Use of face mask during investigation, collection, and analysis of samples is recommended to minimize such contamination.

Sweat

Sweat became unfailing evidence that are deposited unconsciously at various touched surfaces or handled objects and due to their transparent and evaporative nature, rarely engages one's attention during deliberate cleaning of other evidences. Biologically, sweat is a watery fluid secreted from eccrine and apocrine sweat glands present throughout the body along with dissolved mineral, metabolites, and epithelial cells. Around 650 sweat glands are present in average square inch of skin resulting in primary transfer, i.e., deposition of trace amount of sweat on surfaces that come in contact with the skin. The concept of "Touch DNA" by primary transfer and "Transfer DNA" by secondary transfer of DNA relating to Locard's Principle of exchange has gained much attention over the last few years (Kisilevsky et al. 1999; Ladd et al. 1999). Touch DNA or trace DNA are described as low levels of DNA deposited on handled, touched, or worn object without presence of detectable body fluid. Minute traces of epidermal cells along with sweat generally result in deposition of touch DNA. Transfer DNA, on the other hand, are resulted from secondary transfer and include foreign DNA present on individual's hand from previous contact that are subsequently deposited on other surfaces (Wickenheiser 2002). This may include skin-skin-object mode of transfer or skin-object-skin mode of transfer (Burrill et al. 2019).

Quality and quantity of touch DNA recovered from any surface is affected by a wide variety of factors. Shedding status is one of the factors that influences the yield of touch DNA and is described as the tendency of an individual to lose skin cells. It is reported to be higher in women compared to men due to the presence of

thicker stratum corneum in men, making it more stable (Faleeva et al. 2018). Shedding rate also depends upon the individual's age due to high proliferation rate and less degraded DNA in children as compared to elderly people (Poetsch et al. 2013). The yield of trace DNA is affected by the number of deposited DNA-bearing cells, nature of the surface carrying the deposited DNA, lapse of time between deposition and recovery coupled with exposure to environmental conditions, method of sample collection and DNA extraction employed and is independent of timeframe for which skin remains in contact with the surface (Alketbi 2018).

Worn clothing, footwear, beddings, wallets, and bags and door handles are some of the common substrate bearing sweat stains. A study on potential transfer of touch DNA revealed that samples of sweat collected from the beddings after one night of sleep provide good DNA profile. DNA profiles of former individual in contact with the bedding can also be generated. Similar study on sweats of foot and soles of footwear inferred higher DNA amount from the top of foot than the soles. Microbial impacts, cell compressions, and presence of certain PCR inhibitors at the underside of foot and sole justify the loss of DNA on the sole. Synthetic sport shoes yielded better amount of DNA than the leather shoes. The areas as well as techniques of sample collection were also found to affect DNA recovery (Bright and Petricevic 2004). Adhesive tape lifting, dry swab, and cutting out are some of the commonly employed techniques of sample collection. Double swabbing technique for collection of sweat stains, with first wet swab and second dry swab, yields greater DNA recovery from dry swabs than the wet swabs.

Minimal amount of available touch DNA necessitates maximum sample collection and extraction of DNA in shortest possible time period. Zhou et al. (2016) described use of 96-well centrifugal filtration plate and automated DNA extraction on liquid workstation from swabs from door handles, gloves, beverage bottles, cigarette butts, tools, etc. resulting in 54.43% successful profile rate. Recovery of touch DNA from metal surfaces such as ammunition, door handles, and furniture is affected by interactive nature of DNA with metal cations as well enzymatic actions of certain metals on degradation of DNA. Combination of collection method and buffer specific to the metal surface carrying DNA resulted in improved recovery of DNA (Tucker 2015).

Latent fingerprints resulting from deposition of sweat, skin cells, and particulate matters, frequently occurring in any crime scene, are also an efficient source of samples for DNA profiling. Recovery of DNA from fingerprints depends on the substrate, standard pressure, frequency of hand washing, and exposure to fingerprint-enhancing chemicals (Table 5).

Urine

The importance of bodily fluids as sources of DNA for identification purpose has been known since a longer period of time (Hilhorst et al. 2013). Urine, being a useful tool as a source of genetic material, has not been densely applied as a

Table 5 Effect of fingerprint-enhancement chemicals on DNA profiling

Fingerprint-enhancement techniques	Surface	Effect of DNA profiling	Reference
White powders	Glass and wood	Better DNA yield from wood surface than glass. BVDA white, Faurot white, magnetic black, special black fingerprint powders had no impact on DNA and yielded a good profile	Van Hoofstat et al. (1999)
Black powders			
Metal powders			
Black magnetic powders			
UV light, DFO, ninhydrin	Paper	Increased inhibitory effects observed	Raymond et al. (2008)
White powders, black powder, magnetic powder	Glass	No effect on DNA yield	
Cyanoacrylate fuming, Cyanoacrylate + rhodamine 6G, Cyanoacrylate + VMD	Plastic	Rhodamine stained latent fingerprint resulted in better DNA profile	
Ninhydrin	Paper	Less amount of DNA but sufficient for DNA profiling without any inhibitory effects	Schulz et al. (2004)
Vacuum metal deposition (VMD)	Plastic	No effect on DNA quality and quantity	Bhoelai et al. (2011)
Cyanoacrylate fuming	Plastic	No effect on DNA quality and quantity	
Ninhydrin	Paper	Degradation and contamination of DNA	
1,8-diaza-9-fluorenone (DFO)	Paper	Degradation and contamination of DNA	
Black fingerprint powder	Glass	Acceptable decline of DNA quality and quantity	Alem et al. (2017)
Magnetic latent print powder	Glass	Less efficient genetic analysis	
Luminescent cyanoacrylate fuming	Plastic	Observable extent of DNA degradation	Khuu et al. (2018)

potential source of DNA for identification purposes in Forensic Sciences (Junge et al. 2002). Urine may be submitted as forensic evidence in violent crimes, hanging, and illicit drug screening tests (Ng et al. 2018).

The DNA is contained in epithelial cells of human urine, such as renal tubular, squamous cells, transitional urothelial, leukocytes epithelial cells (renal tubular, squamous cells and transitional urothelial), and malignant cells. There are a lot of factors on which the quantity of extractable DNA depends such as the extraction procedure, storage condition, gender, bacterial contamination, and release of nucleases from cells. As the concentration of DNA in urine is low and the instability of DNA in urine preservation, personal identification using urine samples becomes difficult. Healthy individuals, especially males, contain very less nucleated cells and also shelter bacteria that can act as inhibitor during amplification of DNA. The major urine component, i.e., urea, also acts as an inhibitor and can affect the yield of DNA (Aoki et al. 2017).

The yield of DNA is dependent on many factors like the gender of the urine sample, temperature and storage conditions, quality, quantity, and age of the sample.

Generally, yield is greater in females than males and in fresh urine sample rather than stored urine sample (Aoki et al. 2017; Ng et al. 2018). In cases of urine stains, sample should always be collected from the largest stains that are available. The collection and storage of samples should be properly done in order to avoid degradation so that maximum DNA could be extracted (Yokota et al.1998).

Fecal Matter

Fecal matter identification is helpful in providing relevant clue during a criminal investigation, and the individual peculiarities of a fecal sample can be adequately determined by DNA analysis. Fecal as a forensic evidence can be found in many cases where crimes with cruelty are done like sexual assault, sodomy, vandalism, and burglary during which the executioner defecated at the scene of crime. In cases related to animals/wildlife species, genetic evidence from animals, plants, bacteria, and viruses has been used in criminal investigations as forensic tool for identification and individualization purpose (Forgacs et al. 2019).

Feces are a type of waste matter formed in the intestines during the last phase of digestion as a direct result of food. Its composition includes a complex mixture of undigested foodstuffs, intestinal bacteria, intestinal epithelial cells, electrolytes, bile pigments, soluble and insoluble gastrointestinal tract products, mucus, and water. Due to low number of cells and co-extraction of various unwanted substances (due to presence of several PCR inhibitors from digestive system, soil, and foodstuffs), DNA analysis from fecal matter is a challenging aspect. Although the quality and quantity of DNA from fecal matter is comparatively lower than traditional sources of DNA (such as blood, semen, saliva, etc.), few studies have suggested fecal samples as a crucial and valid source of genetic material by comparing the results from the same individuals to high-quality DNA samples (Forgacs et al. 2019). So, DNA analysis from fecal matter can be highly important in criminal investigation.

After three and half decades, due to technologocal advancements, forensic DNA technology has become a potential tool in the court of law, resulted into increasing the rate of conviction. Nowadays, reliable and sensitive techniques for DNA isolation from a variety of biological samples have been developed which attained through gradual technological advancement which shown in Fig. 1.

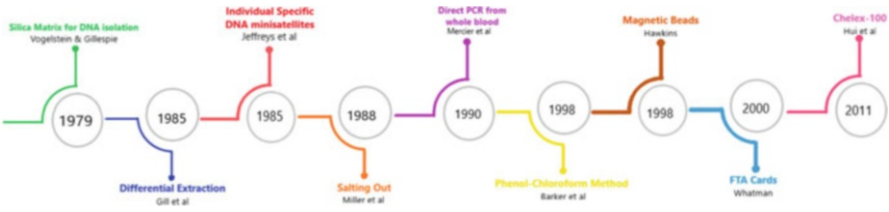


Fig. 1 Timeline of DNA extraction protocols

References

Ahmad NN, Cu-Unjieng AB, Donoso LA (1995) Modification of standard proteinase K/phenol method for DNA isolation to improve yield and purity from frozen blood. J Med Genet 32:129–130

Alderson G, Gurevitch H, Casimiro T et al (2018) Inferring the presence of spermatozoa in forensic samples based on male DNA fractionation following differential extraction. Forensic Sci Int Genet 36:225–232

Alem L, Valentin ESB, Cunha MA et al (2017) Efficiency of DNA recovery from fingerprints enhanced with black and magnetic powders. Forensic Sci Int Genet Suppl Ser 6:e490–e491

Alketbi SK (2018) The affecting factors of touch DNA. J Forensic Res 9:424

Andersen J, Bramble S (1997) The effects of fingermark enhancement light sources on subsequent PCR-STR DNA analysis of fresh bloodstains. J Forensic Sci 42:303–306

Anzai-Kanto E, Hirata MH, Hirata RDC et al (2005) DNA extraction from human saliva deposited on skin and its use in forensic identification procedures. Braz Oral Res 19:216–222

Aoki K, Tanaka H, Ueki M (2017) DNA typing for personal identification of urine after long-term preservation for testing in doping control. Drug Test Anal 9:1116–1123

Barbaro A, Cormaci P (2006) DNA typing from 15-year-old bloodstains. In: International Congress series. Elsevier 1288:550–552

Bhoelai B, de Jong BJ, de Puit M, Sijen T (2011) Effect of common fingerprint detection techniques on subsequent STR profiling. Forensic Sci Int Genet Suppl Ser 3:e429–e430

Bianchi DW, Flint AF, Pizzimenti MF et al (1990) Isolation of fetal DNA from nucleated erythrocytes in maternal blood. Proc Natl Acad Sci 87:3279–3283

Bright J-A, Petricevic SF (2004) Recovery of trace DNA and its application to DNA profiling of shoe insoles. Forensic Sci Int 145:7–12

Burrill J, Daniel B, Frascione N (2019) A review of trace "Touch DNA" deposits: variability factors and an exploration of cellular composition. Forensic Sci Int Genet 39:8–18

Butler JM (2005) Forensic DNA typing: biology, technology, and genetics of STR markers. Elsevier, London

Chen WC, Kerr R, May A et al (2018) The integrity and yield of genomic DNA isolated from whole blood following long-term storage at− 30° C. Biopreserv Biobank 16:106–113

Cherian E, Rajkumar K, Cherian B et al (2015) Utility of saliva and stains of saliva in deoxyribonucleic acid fingerprinting: application in forensic odontology. J Int Oral Heal 7:66–71

Chiou FS, Pai CY, Hsu YPP et al (2001) Extraction of human DNA for PCR from chewed residues of betel quid using a novel "PVP/CTAB" method. J Forensic Sci 46:1174–1179

De Almeida JP, Glesse N, Bonorino C (2011) Effect of presumptive tests reagents on human blood confirmatory tests and DNA analysis using real time polymerase chain reaction. Forensic Sci Int 206:58–61

de Oliveira Musse J, Marques JAM, Remualdo V et al (2019) Deoxyribonucleic acid extraction and quantification from human saliva deposited on foods with bitemarks. J Contemp Dent Pract 20:549

Dissing J, Rudbeck L, Marcher H (1996) A five minute procedure for extraction of genomic DNA from whole blood, semen and forensic stains for PCR. In: 16th Congress of the International Society for Forensic Haemogenetics (Internationale Gesellschaft für forensische Hämogenetik eV), Santiago de Compostela, 12–16 September 1995. Springer, pp 269–271

Everson RB, Mass MJ, Gallagher JE et al (1993) Extraction of DNA from cryopreserved clotted human blood. BioTechniques 15:18–20

Faleeva TG, Ivanov IN, Mishin ES et al (2018) Possibilities of DNA identification of foreign sweat and grease substance on human skin. Russ J Genet 54:746–752

Forgacs D, Wallen RL, Boedeker AL, Derr JN (2019) Evaluation of fecal samples as a valid source of DNA by comparing paired blood and fecal samples from American bison (*Bison bison*). BMC Genet 20:1–8

Frégeau CJ, Germain O, Fourney RM (2000) Fingerprint enhancement revisited and the effects of blood enhancement chemicals on subsequent Profiler Plus™ fluorescent short tandem repeat DNA analysis of fresh and aged bloody fingerprints. J Forensic Sci 45:354–380

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints'. Nature 318: 577–579
Goldstein MC, Cox JO, Seman LB, Cruz TD (2019) Improved resolution of mixed STR profiles using a fully automated differential cell lysis/DNA extraction method. Forensic Sci Res 5:1–7
Grimberg J, Nawoschik S, Belluscio L et al (1989) A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. Nucleic Acids Res 17:8390
Guha P, Das A, Dutta S, Chaudhuri TK (2018) A rapid and efficient DNA extraction protocol from fresh and frozen human blood samples. J Clin Lab Anal 32:e22181
Harris KA, Thacker CR, Ballard D, Court DS (2006) The effect of cleaning agents on the DNA analysis of bloodstains deposited on different substrates. In: International Congress series. Elsevier 1288:589–591
Hilhorst M, Theunissen R, van Rie H et al (2013) DNA extraction from long-term stored urine. BMC Nephrol 14:238
Hochmeister MN, Dirnhofer R, Borer UV et al (1991) PCR-based typing of DNA extracted from cigarette butts. Int J Legal Med 104:229–233
Inci F, Ozen MO, Saylan Y et al (2018) A novel on-chip method for differential extraction of sperm in forensic cases. Adv Sci 5:1800121
Jeanpierre M (1987) A rapid method for the purification of DNA from blood. Nucleic Acids Res 15: 9611
Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA. Nature 316:76–79
Junge A, Steevens M, Madea B (2002) Successful DNA typing of a urine sample in a doping control case using human mitochondrial DNA analysis. J Forensic Sci 47:1–3
Khare P, Raj V, Chandra S, Agarwal S (2014) Quantitative and qualitative assessment of DNA extracted from saliva for its use in forensic identification. J Forensic Dent Sci 6:81
Khuu A, Chadwick S, Moret S et al (2018) Impact of one-step luminescent cyanoacrylate treatment on subsequent DNA analysis. Forensic Sci Int 286:1–7
Kim Y-W, Kim Y-K (2006) The effects of storage of human saliva on DNA isolation and stability. J Oral Med Pain 31:1–16
Kisilevsky AE (1999) DNA PCR profiling of skin cells transferred through handling. In: Proceeding of the annual meeting of the Canadian Society of Forensic Science; 1999
Ladd C, Adamowicz MS, Bourke MT et al (1999) A systematic analysis of secondary DNA transfer. J Forensic Sci 44:1270–1272
Lee HC, Gaensslen RE, Pagliaro EM et al (1989) The effect of presumptive test, latent fingerprint and some other reagents and materials on subsequent serological identification, genetic marker and DNA testing in bloodstains. J Forensic Ident 39:339–358
Madisen L, Hoar DI, Holroyd CD et al (1987) The effects of storage of blood and isolated DNA on the integrity of DNA. Am J Med Genet 27:379–390
Manna AD, Montpetit S (2000) A novel approach to obtaining reliable PCR results from luminol treated bloodstains. J Forensic Sci 45:886–890
Martínez P, Santiago B, Alcalá B, Atienza I (2015) Semen searching when sperm is absent. Sci Justice 55:118–123
McNally L, Shaler RC, Baird M et al (1989) Evaluation of deoxyribonucleic acid (DNA) isolated from human bloodstains exposed to ultraviolet light, heat, humidity, and soil contamination. J Forensic Sci 34:1059–1069
Mercier B, Gaucher C, Feugeas O, Mazurier C (1990) Direct PCR from whole blood, without DNA extraction. Nucleic Acids Res 18:5908
Miller SA, Dykes DD, Polesky H (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215
Mutter N, Deacon P, Farrugia KJ (2018) The effect of cyanoacrylate fuming on subsequent protein stain enhancement of fingermarks in blood. J Forensic Ident 68:545–556
Nasiri H, Forouzandeh M, Rasaee MJ, Rahbarizadeh F (2005) Modified salting-out method: high-yield, high-quality genomic DNA extraction from whole blood using laundry detergent. J Clin Lab Anal 19:229–232

Newall PJ, Richard ML, Kafarowski E et al (1996) Homicide case report: successful amplification and STR typing of bloodstains subjected to fingerprint treatment by cyanoacrylate fuming. Can Soc Forensic Sci J 29:1–5
Ng DPK, Koh D, Choo SGL et al (2004) Effect of storage conditions on the extraction of PCR-quality genomic DNA from saliva. Clin Chim Acta 343:191–194
Ng HH, Ang HC, Hoe SY et al (2018) Simple DNA extraction of urine samples: effects of storage temperature and storage time. Forensic Sci Int 287:36–39
Poetsch M, Bajanowski T, Kamphausen T (2013) Influence of an individual's age on the amount and interpretability of DNA left on touched items. Int J Legal Med 127:1093–1096
Potter AA, Hanham AF, Nestmann ER (1985) A rapid method for the extraction and purification of DNA from human leukocytes. Cancer Lett 26:335–341
Prinz M, Berghaus G (1990) The effect of various stain carriers on the quality and quantity of DNA extracted from dried bloodstains. Int J Legal Med 103:191–197
Rahikainen A-L, Palo JU, de Leeuw W et al (2016) DNA quality and quantity from up to 16 years old post-mortem blood stored on FTA cards. Forensic Sci Int 261:148–153
Raymond JJ, van Oorschot RAH, Walsh SJ, Roux C (2008) Trace DNA analysis: do you know what your neighbour is doing?: a multi-jurisdictional survey. Forensic Sci Int Genet 2:19–28
Schulz MM, Wehner H-D, Reichert W, Graw M (2004) Ninhydrin-dyed latent fingerprints as a DNA source in a murder case. J Clin Forensic Med 11:202–204
Sweet D, Hildebrand D (1999) Saliva from cheese bite yields DNA profile of burglar: a case report. Int J Legal Med 112:201–203
Sweet D, Shutler GG (1999) Analysis of salivary DNA evidence from a bite mark on a body submerged in water. J Forensic Sci 44:1069–1072
Sweet D, Lorente M, Valenzuela A et al (1996) Increasing DNA extraction yield from saliva stains with a modified Chelex method. Forensic Sci Int 83:167–177
Sweet D, Lorente M, Lorente JA et al (1997) An improved method to recover saliva from human skin: the double swab technique. J Forensic Sci 42:320–322
Tas S (1990) Purification of DNA from clotted blood. Clin Chem (Baltimore) 36:1851
Thabet HZ, Ghandour NM, Salama RH (2018) Effect of some cleaning products on blood DNA retrieval from cloth. Egypt J Forensic Sci Appli Toxicol 18:53–66
Thompson WC (1996) DNA evidence in the OJ Simpson trial. Univ Colo Law Rev 67:827
Tobe SS, Watson N, Daeid NN (2007) Evaluation of six presumptive tests for blood, their specificity, sensitivity, and effect on high molecular-weight DNA. J Forensic Sci 52: 102–109
Tucker KL (2015) Comparative study of touch DNA recovery from metals, a thesis submitted to fulfillment of the requirements for the Degree of Master of Science by Tucker, Kambra. Oklahoma State University, ProQuest Dissertations Publishing, 2015, 10140171
Van Hoofstat DEO, Deforce DLD, Hubert De Pauw IP, Van den Eeckhout EG (1999) DNA typing of fingerprints using capillary electrophoresis: effect of dactyloscopic powders. Electrophor An Int J 20:2870–2876
Vaught JB (2006) Blood collection, shipment, processing, and storage. Cancer Epidemiol Prev Biomarkers 15:1582–1584
Walsh DJ, Corey AC, Cotton RW et al (1992) Isolation of deoxyribonucleic acid (DNA) from saliva and forensic science samples containing saliva. J Forensic Sci 37:387–395
Watanabe Y, Takayama T, Hirata K et al (2003) DNA typing from cigarette butts. Legal Med 5: S177–S179
Wickenheiser RA (2002) Trace DNA: a review, discussion of theory, and application of the transfer of trace quantities of DNA through skin contact. J Forensic Sci 47:442–450
Wiegand P, Schürenkamp M, Schütte U (1992) DNA extraction from mixtures of body fluid using mild preferential lysis. Int J Legal Med 104:359–360

Xu R, Ye P, Luo L et al (2010) A simple and efficient method for DNA purification from samples of highly clotted blood. Mol Biotechnol 46:258–264

Yokota M, Tatsumi N, Tsuda I et al (1998) DNA extraction from human urinary sediment. J Clin Lab Anal 12:88–91

Yoshida K, Sekiguchi K, Mizuno N et al (1995) The modified method of two-step differential extraction of sperm and vaginal epithelial cell DNA from vaginal fluid mixed with semen. Forensic Sci Int 72:25–33

Zhou R, Meng Q, Li P et al (2016) Improved high throughput processing of touch DNA samples. Aust J Forensic Sci 48:2–9

Collection, Preservation, and Transportation of Biological Evidences

4

Hirak Ranjan Dash and Kamayani Vajpayee

Contents

Abstract

DNA fingerprinting technique is considered to be one of the most irrefutable evidence in the criminal justice system. To obtain a proper result to be produced in the court of law, proper collection, preservation, and transportation of

H. R. Dash (✉)
School of Forensic Science, National Forensic Sciences University, Delhi Campus, New Delhi, India

K. Vajpayee
DNA Fingerprinting Unit, Forensic Science Laboratory, Bhopal, Madhya Pradesh, India

School of Arts and Sciences, Ahmedabad University, Ahmedabad, India

Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, Gujarat, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_3

biological evidences play an imperative role. To conduct a DNA test, the biological samples can be divided into two types, the questioned exhibits and the reference exhibits. The results obtained from both the exhibits are analyzed and compared to draw a conclusion. Questioned sample may be obtained from varied sources such as bone, teeth, hair, vaginal swab, clothing, knife, or any other article suspected to contain any biological material. However, the reference samples are always the peripheral blood or the buccal swab of the suspected individual or the survivor. In this regard, the present chapter describes the various nature of the biological exhibits along with their proper collection, preservation, and transportation to the laboratory to obtain a proper DNA profiling result.

Keywords

Biological exhibits · DNA profiling · Collection · Preservation · Transportation

Introduction

In the past few decades, DNA profiling technique has gained a huge attention among the investigators in solving criminal and civil cases of varied nature such as paternity dispute, identification of mutilated bodies, sexual assault, and murder. Courts have appraised physical witnesses such as DNA technology to be one of the most irrefutable evidences in the criminal justice system. DNA profiling-based physical evidences take advantage of trace DNA technologies and link between suspect/victim to a crime *independently* and *objectively* (Lee and Ladd 2001). Over the years, the technology has undergone a huge transformation in terms of its sensitivity and discrimination power. After the invention of STR-based PCR amplification technique, biological evidences that were not useful in earlier days has also become of huge importance. Further advancement of this technique also allows the amplification and detection of DNA fragments in previously unsuitable evidences such as samples containing PCR inhibitors and samples with degraded DNA. Over the time, the nature of collection, preservation, and transportation of biological evidences has been modified as per the technological advancements.

Generating a quality DNA profile is highly dependent on the proper collection, preservation, and transportation of biological evidences. DNA fingerprinting technique relies on the analysis of DNA found inside every biological cell. Such biological samples are perishable in nature and need appropriate storage conditions with the use of suitable preservatives. Every forensic technology is different and we should not assume that the same biological samples should be stored in the same preservative for different analyses. For example, blood samples are examined for the presence of alcohol or other intoxicants by toxicological analyses. In such instances, 1–5% sodium or potassium fluoride is used as a preservative (Dinis-Oliveira et al. 2016). Whereas, for DNA analysis, liquid blood sample should be collected in EDTA.

Crime scenes contain evidences which are ad rem to any investigation. Investigators recon mostly on the samples that have been collected from the crime scene to use them as a proof or evidence in a trial. These evidences play an important role either in exonerating a person or to prove him guilty (incrimination) in front of jury. Evidences, on the basis of their nature, can be classified broadly into various categories, viz., physical evidence (glass, soil, paint, fiber, etc.) and biological evidence (blood, semen, saliva, etc.). Among all, biological evidence has gained much of the attention in the recent years. The advent of new and sensitive technologies has aided in the processing and analysis of these trace evidences.

Earlier, the biological evidences especially blood was used for blood typing to discriminate the suspect and to establish the link between the crime, victim, and criminal. Later with the invention of the polymerase chain reaction (PCR) technology, the use of short tandem repeat (STR) markers, the genetic analysis of the biological evidences became possible and has revolutionized the field of forensic genetics. Therefore, the technology initially was applied to the cases where abundant and a good quality of biological evidences could be recovered such as in the case of sexual assault, murder, and homicide. However, the field of genetics developed further and allowed the invention of more sensitive and reliable technologies which could handle the analysis of trace evidences (sample) having negotiated quality such as cold, degraded, and contaminated samples.

The basic principle behind forensic genetics relies on a fact that every biological substance (cell) contains genetic material and that it can be isolated to generate a profile identical to an individual (individualization). Thus, this very concept allowed investigators to collect all the possible biological exhibits from the scene of crime. These include a wide range of biological samples like blood, semen, saliva, bone, teeth, or even the samples which may have the possibility of containing touch DNA like clothing (Hess and Haas 2017). The technological advancements in the field of forensic genetics or the DNA profiling methodology has opened the field for the array of crime scenes for which the DNA profiling can now be performed such as theft, burglaries, motor vehicle crimes, terrorism, etc. Investigators are now collecting and submitting wide array of samples for their DNA analysis.

While dealing with trace evidences and sensitive instruments, it has now become more important to formulate and carefully follow the protocols regarding collection, preservation, and storage of the exhibits. The cautious use of collection and preservation methods will protect the samples from further degradation occurring due to environmental, physical, and chemical factors, cross contamination. There are specific protocols and procedures laid down by several accrediting agencies that are being recommended to professionals and crime scene investigators. A thorough training of officers on importance of handling different types of evidences, proper collection, and preservation procedures should be mandatory. Attention should also be made on identifying the different types of evidences, their evidentiary value, and proper packaging and transportation procedures. Thus, realizing the importance of proper collection, preservation, and transportation of the exhibits encountered at crime scene, this chapter aims at laying out the standardized guidelines expected to be used by investigators at the scene of crime.

Biological Evidence

Forensic science deals with the application of well-established scientific techniques on the biological exhibits to draw a conclusion which is accepted in the criminal justice system. As it is a science of comparison, the biological exhibits can be divided into two types, i.e., the questioned exhibits and the reference exhibits. Firstly, the examination is conducted on the questioned sample and the results obtained from the questioned sample are compared with the reference sample. Based on the observation and matching the results between the questioned and the reference samples, conclusion is drawn in a case.

In a simple case of paternity dispute, the mother is considered to be known, child is questioned, and the father is the accused/alleged. Similarly in a case of identification, the unknown/questioned sample may be of either bone, teeth, fetus, hair, or any other biological object. Here, the reference samples can be considered of the biological samples obtained from the putative father, mother, child, wife, or any other patrilineal/matrilineal relatives. The questioned sample in case of a murder may involve the clothing, blood stained soil, any murder weapon of any other sample that presumably contains biological fluids of either deceased or the accused which has the potential to link the deceased, accused, crime scene, and the murder weapon. Additionally, the questioned samples in a case of sexual assault involve any article collected from the sexual assault survivor's body or its belongings. Here the Locard's exchange principle comes into play, where the body fluid of the culprit and the survivor gets exchanged during the course of sexual assault (Sammons 2014).

Reference sample for forensic DNA analysis can be defined as a sample with high quantity and quality DNA and is collected in front of witnesses to be accepted in the court of law ("Glossary for Crime Scene Investigation: Guides for Law Enforcement" 2021). Hence, irrespective of all type of cases, the widely used reference samples include the peripheral blood collected either in a K_3EDTA vial or in a FTA® card. Alternatively, buccal swab can also be collected as an alternative to the liquid blood sample as a useful noninvasive sample (Ghatak et al. 2013). However, there exists certain limitations; at many instances, the sample donor is a habitual tobacco chewer. In such cases, buccal swab samples may not yield an optimum DNA result as tobacco acts as a potential PCR inhibitor (Adamowicz et al. 2014); hence, peripheral blood samples should be treated as a reference sample.

DNA Content in Biological Samples

Various types of biological samples are being collected from the crime scene and are sent to forensic laboratory for their analysis. The DNA content in these samples varies significantly (Andréasson et al. 2006). There are samples like liquid blood where a sufficient amount of DNA can be extracted, and on the other hand, there are

Table 1 Summary of DNA content in common biological evidences

Type of biological material	DNA content	References
Blood	30–40 μg/mL	("How much DNA and RNA can be expected from human blood cells? – QIAGEN" 2021)
Bone	100 ng/gram	(Iwamura et al. 2004)
Hair	0.4 ng per strand	(Heywood et al. 2003)
Buccal swab	18 ng/μl – 433 ng/μl	(Livy et al. 2011)
Teeth	247.807 ± 23.8 ng/μl	(Rubio et al. 2018)

samples like hair from which a very small amount of DNA can be recovered. Table 1 summarizes the amount of DNA that can be recovered from various biological samples.

While analyzing the crime scene, an investigator must have a prior knowledge about the genetic content in the biological materials so that materials with higher DNA content can carefully be collected for better interpretation and conclusion. Further this would also help in preserving the only samples collected from the crime scene with little DNA content.

- **Blood**: Blood is an excellent source of DNA. Blood contains erythrocytes, leucocytes, platelets, and plasma. Among all, leucocytes are the sources of genetic material since mature erythrocytes lack nuclei. Approximately 30–40 μg of DNA is found in per milli liters of blood ("How much DNA and RNA can be expected from human blood cells? – QIAGEN" 2021).
- **Bone**: Bone is the choice as the DNA source only in case of decomposed body. Bones can also be taken for consideration in case of ancient DNA analysis. DNA can be isolated from the demineralized long bones since calcium is a potential PCR inhibitor. Usually bones may contain 100 ng of DNA per gram of bone (Iwamura et al. 2004).
- **Teeth**: Similar to bones, teeth also act as a DNA source in case of decomposed, burnt, and charred body. The tooth pulp is known to be the richest source of DNA. Studies have shown that more DNA is retrieved from multi-rooted teeth than single-rooted teeth. Usually teeth may contain 247.807 ± 23.8 ng of DNA per micro liters of demineralized tooth (Rubio et al. 2018).
- **Hair**: Shed hairs are mostly found in crime scenes although plucked hair samples are also collected sometimes as reference material and/or in the sexual assault cases. In hair, the sufficient quantity of nuclear DNA is only found in the root sheath whereas mitochondrial DNA is abundantly found in hair shaft. Approximately 0.4 ng of nuclear DNA can be isolated from a single hair. It is difficult to standardize the amount of DNA that can be recovered from the hair since every hair from a single source differ in their DNA content (Heywood et al. 2003).

- **Buccal Swab**: Buccal swabs are alternate to invasive sample collection method. When blood could not be taken as reference sample, buccal swabs are considered as second option. Buccal swabs are prone to contamination, and thus, it is recommended to use mouth wash before sample collection. Buccal swabs contain epithelial cells and saliva as the source of DNA. Researchers have estimated that a buccal swab may contain 18 ng/μl – 433 ng/μl of DNA (Livy et al. 2011).

Collection of Biological Samples

Once the investigating officer reaches the scene of crime, he must first secure the crime scene so as to maintain its integrity (Pepper 2010). He must observe the scene of crime with utmost care and look for the best method of examining it. Moreover, a brief observation is required to locate all the possible evidences. The expert must try to collect fragile evidences in first place so that degradation due to environmental factors could be avoided. It is the primary duty of the crime scene investigator to ensure that all the materials required for collection, preservation of the evidences from the crime scene is available with him. Table 2 enlists the materials that are required for the proper handling of the evidences at the crime scene. Although there might be the cases where some specialized materials, instruments or reagents would be required apart from those listed in the table.

Depending upon the type of crime committed (homicide, sexual assault, murder, theft, motor vehicle accidents, etc.), a range of biological evidences can be found in a crime scene. These may include blood, semen, saliva, sweat, hair, bone, urine, etc. Before collection, it is necessary to carefully observe and recognize all the evidences present at the scene of crime. There are samples like blood, bone, tissue which are visible to the naked eye and can thus be detected easily. Detection of the samples like semen, saliva, fingerprints, etc. can be difficult owing to their color and quantity. Such samples are made visible by using either alternative light sources or chemicals. For example, Polilight is used to detect seminal and blood stains whereas Luminol can be used to detect trace amounts of blood. The areas with a doubt of being touched can be analyzed for the presence of latent fingerprints by using enhancement powders (Miranda et al. 2014; Sterzik et al. 2016; van Oorschot et al. 2010).

Apart from collection, investigating officer must take some precautionary measures to ensure proper preservation of the exhibits. The preservation methods are purely dependent upon the type of sample to be preserved. Table 3 outlines some important measures that are to be taken care of while collecting, preserving, and transporting the biological evidences.

As regards to the storage conditions, every sample has its own optimum temperature under which the sample integrity is maintained. Usually short-term and long-term storage conditions vary for the samples. Tables 4 and 5 defines the short-term and the long-term storage conditions for the biological evidences, respectively.

Table 2 List of materials required at the crime scene (Modified from Mozayani and Parish-Fisher 2018)

Items
Protective personal equipment (PPE)
Gloves (latex, nitrile, cotton)
Facemasks
Hair cap
Disposable overshoes
Safety goggles
Sterile sealed swabs
Sterile tweezers
Plastic tubes
Sterile disposable plastic pipettes
Disposable sterile scalpels or razor blades
Biological hazard bags and sharp bins
Polythene bags
Brown paper sacks
Cardboard boxes (flat pack)
Scissors
Adhesive tape
Stapler and pins
Evidence tape and tags
Scene of crime barrier tape
Identification labels
Pens and markers
Ruler and tape measure
Plain paper (A4 and A3)
Thermometer
Magnifying glass
Torch
Fingerprint brushes and fingerprint powders
Lifting tape with acetate sheet
Alternative light source (ALS)
Digital and video camera
Clipboard
Crime scene investigation forms and necessary laboratory documents
Reagents
Deionized/ultrapure water
Ethanol
Bleach
Reagents for presumptive testing of biological stains
Fingerprint enhancement reagents

Methodology for the Collection and Preservation of Biological Evidence

An investigator encounters a wide range of biological evidences at the crime scene. For their better recovery, it is essential to follow standard procedures of collection. There are several methods like double swabbing method, cutting method, etc. that allow the collection of biological evidences while preserving their integrity. These methods can be used at the scene of crime as well as in the laboratory.

Table 3 Measures to be taken while collecting, preserving, and transporting the biological evidences

While attending the scene of crime, investigator must ensure that he/she is carrying all the necessary items required for collection and preservation of the exhibits.
Investigating officers along with the other officers dealing with the crime scene must wear proper safety garments which include gloves, shoe cover, hair cap, mask, and safety suit.
To avoid any contamination and to maintain the integrity of the evidences, touching the anything bare hands must be avoided.
A proper documentation of the crime scene along with the exhibits before collection should be prepared that may include notes, sketching, photography, and videography.
Attention must be paid to avoid any kind of cross contamination among the biological evidences. Further individual packaging of the exhibits should be practiced.
Liquid samples like blood must be dried in air before packaging. Similarly, swabs and slides should be allowed to dry before packaging.
In the case of collection of the stain, a controlled sample must also be collected and sent to laboratory. This controlled sample has to be collected from unstained portion.
Paper bags and card board boxes should be used to pack biological stains. Plastic bags have the tendency to retain moisture and therefore must be avoided.
After collection and packaging, the package needs to be sealed, labelled with brief description about the type of evidence, date of collection, location, name, etc. the package may also include appropriate storage conditions.
Investigator must prepare a checklist of all the items collected from the scene of crime.
After collection, all the evidences must be sent promptly to the laboratory for analysis.
Care must be taken while transporting the exhibits to the laboratory. Samples liable to temperature and light must be kept in ice box so as to avoid further degradation or damage.
Chain of custody is important; hence, a proper documentation regarding chain of custody must be prepared before transporting the exhibits to the laboratory.

Scraping Method

The scrapping method is used to recover dry materials on porous surface. The method is followed under controlled environment so as to avoid any kind of contamination. Following steps are to be taken while working with this method:

- Locate the stain of interest.
- Check if the stain is dry before performing further steps.
- Carefully scrape the stain from the object on to a clean, sterile paper piece.
- Avoid using bare hands; always wear hand gloves before recovering the stain.
- Secure the scrapings and place the paper in an envelope.
- Seal and label the envelope.
- Store as per physiological conditions (Mozayani and Parish-Fisher 2018).

Double Swabbing Method

Dry samples on nonporous surface (glass, metal, etc.) are collected using double swabbing method (Pang and Cheung 2007). This method can also be used to collect evidences for touch DNA (Pang and Cheung 2007). The steps are as follows:

Table 4 Recommended physiological conditions for short-term storage of biological evidences. (Modified from Ballou et al. 2013)

Type of evidence	Frozen	Refrigerated	Temperature Controlled	Room Temperature
Liquid blood	Never	Best	Less than 24 hrs	
Urine	Best	Less than 24 hrs		
Dry Stains			Best	Acceptable
Wet Stains	Best	Acceptable	Less than 24 hrs	
Bone	Acceptable		Acceptable	Acceptable
Hair			Best	Acceptable
Swab		Best (wet)	Best (dried)	
Vaginal Slide/Smear			Best	
Buccal Swab			Best	Less than 24 hrs

- Locate the stain on a surface.
- Moisten the swab with normal saline water.
- Wet the dry stain using the moisten swab which will soften and rehydrate the cells.
- Using another moist swab recover the stain from the surface.
- Allow the swab to dry in air and place it in an envelope.
- Seal, label, and store the envelope (Mozayani and Parish-Fisher 2018).

Cutting Method

When stains are found on large objects like sofa, carpet, cutting method is used to recover them. However, if stains are found on small objects like clothing, cigarette butts, experts use this method for sample collection. The steps taken are as follows:

Table 5 Recommended physiological conditions for long-term storage of biological evidences. (Modified from Ballou et al. 2013)

Type of evidence	Frozen	Refrigerated	Temperature Controlled	Room Temperature
Liquid blood	**Never**	Best		
Urine	**Best**			
Dry Stains			**Best**	
DNA Extracts	**Best (Liquid)**	**Acceptable**	**Acceptable**	
Bone			**Best**	
Hair			**Best**	**Acceptable**
Swab			**Best (Dry)**	
Vaginal Slide/Smear			**Best**	
Buccal Swab			**Best**	

- Locate the stain on a surface.
- Allow the stain to dry if wet.
- Cut around the stain with care.
- Pick the cut pieces and place them in an envelope.
- Seal, label, and store the envelope (Mozayani and Parish-Fisher 2018).

Tape Lifting Method

Dry stains on nonabsorbent materials can be recovered using this method. Tape lifting method is majorly used to lift fingerprints and the areas suspected to be touched for the collection of touch DNA. The steps taken are as follows:

- Locate the stain on a surface.
- Place the adhesive tape over stain and press gently.
- Carefully lift the tape off the object and secure the adhesive side with acetate paper.
- Place the tape in an envelope.
- Seal, label, and store the envelope (Mozayani and Parish-Fisher 2018).

Picking Method

Picking method is used to collect solid items found at scene of crime like hair, bone, etc. The steps followed are as follows:

- Locate the evidence.
- With the help of sterile forceps/tweezers, pick the solid biological materials.
- Transfer the material into an envelope or box.
- Seal, label, and store the envelope/box (Mozayani and Parish-Fisher 2018).

Collection, Preservation, Transportation, and Storage Procedures for Commonly Encountered Biological Evidences at the Crime Scene

Liquid Blood

For DNA typing technology, liquid blood is the most suitable evidence. Liquid blood is usually collected as reference sample from an individual may be victim or suspect. It is also at times encountered by the investigating officer at the crime scene (Dash et al. 2020). Liquid blood is collected by an experienced medical practitioner only after taking a written consent from the person. All the medical concerns have to be documented along with the sample such as organ or blood transfusion, etc.

Approximately 2 ml of blood sample in duplicates is collected in EDTA vial. Blood sample can also be taken in a vacutainers containing 5 mM sodium citrate, heparin as anticoagulant. Blood from the deceased body can also be collected in a similar way. Moreover, it must be collected from areas like heart or major internal blood vessels and within 24 h of death. After collection, the vial should be labelled well including the name of the person, name of the collector, exhibit number, date, location, and time. The package must be sealed properly and should be stored at low temperatures before transporting it to the laboratory.

The liquid blood can also be collected on Whatman FTA® cards by dropping approximately 125 µl of blood per inch inside the marked area. The card should be labelled well and dried in air before packaging in individual envelopes. These cards can be transported to laboratories at room temperature. The blood samples are stored at low temperatures but are not freezed.

Liquid Blood and Wet Bloodstains from Crime Scene

Liquid blood can be found in a crime scene along with the blood clots. Blood from the blood pool can be collected through a syringe and transferred either to EDTA vial or vacutainers containing 5 mM sodium citrate, heparin as anticoagulant. In a similar way, the blood clot can be collected using sterile tweezers or spatula.

Blood serum and the blood around the clot can be collected by soaking a sterile cotton swab.

The exhibits can then be labelled with name of the collector, exhibit number, date, location, and time, sealed, and packaged. The samples are refrigerated (not freezed) and transported to the laboratory at the earliest. Sometimes liquid blood stains can be found on the surface of the objects. In such cases, the stains can be cleaned using a sterile cotton swab. The swab is allowed to dry in air followed by packaging in an envelope with proper markings.

Dried Blood Stains

Dried blood stains can be found either on large immovable objects or on removable objects in a crime scene. If the blood stain is on immovable object, the blood stain pattern must be documented before collection. Methods of collection like tape lifting and scrapping can be used to collect the stain. Moreover, the blood stain can also be collected using double swabbing method where a moist cotton swab is used to wipe out the stain. Prior to packaging, the swabs must be allowed to dry in air so as to avoid any fungal growth and degradation. In both the cases, the sample must be packed separately to avoid cross contamination. Also the source could be different for the different stains; thus, packaging of stains in separate envelopes becomes important.

On the other hand, if the blood stain is found on removable objects like cloths, knife, etc., the whole object can be collected, packed, labelled, and sent to the laboratory. In addition, if the blood stain is found on the object which can be cut, then cutting method of evidence collection can be utilized to recover the stain. These objects can be transported to the laboratory at normal room temperature (Dash et al. 2020).

Semen and Seminal Stains and Evidences from Sexual Assault Cases

In sexual assault cases, semen, seminal stains on objects like bedsheet, pillows, clothing, etc. can easily be found. Before recovering the seminal stains, it is always practiced to document the stain either by sketching, photography, or videography. If the encountered stain is dry, it can be collected using scrapping method, cutting method, or double swabbing method. Alternatively, if the stain is wet, it must be allowed to dry in air before collection so as to avoid any contamination. The swabs are to be packed in an envelope and stored in lower temperatures.

Liquid semen sample can be transferred to a tube either by using sterile disposable pipette or syringe. In both the cases, there must be a proper labelling and sealing. In sexual assault cases, medical examination of the victim remains priority and must be carried out at the earliest. Since samples like seminal fluid starts

degrading faster in female body due to the presence of enzymes and acidic pH, other samples like buccal swab, anal swab, vaginal swab, etc. must be collected based upon the case. The slides, swabs prepared must be dried before packaging. The package must be properly labelled with name of the victim, date of collection, name of the medical expert, exhibit number, etc.

Similarly, clothing of victim and the suspect should also be recovered, dried, and properly packed with all the relevant details. All these samples collected are sent to the laboratory in normal temperature. Samples like vaginal slides, seminal slides are to be stored in lower temperatures but are not to be freezed (Dash et al. 2020).

Soft Tissues and Organs

Soft organs or tissues must be collected with the help sterile forceps. They should be packed in a clean, dry, sterile, and airtight glass jar. The tissues/organs should be preserved in normal saline solution. Additionally, antibiotics could be added to the solution to prevent the growth of microbes. The tissues should be packed separately and labelled well. The jars should be sealed and sent to laboratory in ice boxes. In case where the organs are collected from a deceased body having advanced decomposition, a deep muscle along with few other organs should be collected in said manner and sent to laboratory at the earliest.

It is always recommended to preserve heart or brain for DNA analysis. Generally, liver and kidneys are not recovered in such cases. It is important to note that these tissue/organ samples are never preserved in formalin or paraffin wax (Dash et al. 2020).

Teeth and Bone

Teeth and bones are often collected from the deceased skeletonized body. It is the duty of the investigator to the source of the sample from any skilled medical practitioner prior to packaging and transporting to the laboratory. If possible, intact and rooted molars are recommended for collection. In case of bone, long bones especially femur should be collected for DNA analysis. The samples are picked using sterile tweezers and are wrapped cautiously with paper. The package must contain all the information like name of the collector, exhibit number, date, location, and time. The package is stored at room temperature and is sent to the laboratory for further analysis (Dash et al. 2020).

Hair

Nuclear as well as mitochondrial DNA can be isolated from the hair samples. For nuclear DNA analysis, hair needs to have root sheath, whereas mitochondrial DNA

can be isolated from hair shaft samples. If hairs are found in the crime scene, it should be picked up using sterile tweezers and must be packed in a ziplock polybag or envelope. The package needs to be transported to laboratory and stored in room temperature. In case of reference sample collection, approximately 20–25 hair samples have to be submitted since it may require for additional microscopic examination (Dash et al. 2020).

Touch DNA Samples

With the advent of DNA profiling technique from RFLP method to PCR-based STR analysis, the amount of DNA required for the examination is getting decreased day by day. Most of the currently available commercial STR kits require 05–1.0 ng of genomic DNA to generate a complete DNA profile (Cisana et al. 2017; Ludeman et al. 2018). Hence, the analysis of transfer of trace DNA from an object or a person due to mere physical contact has been explored widely recently. Touch DNA samples do not attribute to a particular body fluid, rather, the source of touch DNA may include the shed corneocytes, endogenous/transferred nucleated epithelial cells, fragmented cells and nuclei, and cell-free DNA (Tang et al. 2020). Various factors such as physical activity of the donor, sex, age, substrate, temperature, and humidity contribute to the amount of touch DNA contributed by a donor.

In most of the crime scene evidences, wet/dry swabbing technique is commonly applied. However, a study revealed the usefulness of scraping or tape lift method during collection of touch DNA samples (Williamson 2012). Before collection of a trace DNA evidence from the crime scene, the primary aim should be to locate the possible target surfaces harboring the trace DNA evidence. In suspected homicide cases, in addition to the deceased's body fluid, touch DNA can be explored from surface of murder weapons such as revolver, knife, or pistol. In sexual assault cases, touch DNA sample of the perpetrator can be targeted from the skin or clothing of the survivor (Sowmyya 2016). Looking at the crime scene, the investigator and the forensic expert can ensure the presence of any trace DNA evidence to be collected for a DNA testing.

Chain of Custody of Forensic Samples

Chain of custody is the most important step of evidence documentation to assure the court of law regarding the authenticity of the evidence. Proper documentation of the continuity of possession of evidence, and its movement and location from the point of discovery, recovery, or collection till laboratory examination and admission in the court collectively accounts for the chain of custody of an evidence. In this regard, a record of chain of custody must be maintained and established in the court before assigning it as an exhibit. The lack of which may make the evidence inadmissible in the court and its legitimacy, integrity, and

related examinations become questionable. The chain of custody document ensures that none other than the authorized person has access to the exhibits. The document should be prepared in a comprehensive manner citing information regarding the evidence collection, people handling the evidence, period of possession of the evidence, safekeeping conditions, and other information. Thus, maintaining chain of custody of a DNA evidence becomes utmost important in addition to the other evidences. As the biological samples require proper preservatives and storage conditions, hence, the appropriate storage conditions need to be mentioned in the chain of custody form for any biological exhibit.

References

Adamowicz M, Stasulli D, Sobestanovich E, Bille T (2014) Evaluation of methods to improve the extraction and recovery of DNA from cotton swabs for forensic analysis. PLoS One 9(12): e116351. https://doi.org/10.1371/journal.pone.0116351

Andréasson H, Nilsson M, Budowle B, Lundberg H, Allen M (2006) Nuclear and mitochondrial DNA quantification of various forensic materials. Forensic Sci Int 164(1):56–64. https://doi.org/10.1016/j.forsciint.2005.11.024

Ballou S, Stolorow M, Taylor M, Bamberger P, Brown L, Brown R et al (2013) The biological evidence preservation handbook: best practices for evidence handlers; technical working group on biological evidence preservation. https://doi.org/10.6028/nist.ir.7928

Cisana S, Cerri N, Bosetti A, Verzeletti A, Cortellini V (2017) PowerPlex® fusion 6C system: evaluation study for analysis of casework and database samples. Croat Med J 58:26–33. https://doi.org/10.3325/cmj.2017.58.26

Dash HR, Shrivastava P, Das S (2020) Collection, transportation, and preservation of biological evidences for DNA analysis. In: Principles and practices of DNA analysis: a laboratory manual for forensic DNA typing. Springer Protocols Handbooks, Humana/New York. https://doi.org/10.1007/978-1-0716-0274-4_3

Dinis-Oliveira RJ, Vieira DN, Magalhaes T (2016) Guidelines for collection of biological samples for clinical and forensic toxicological analysis. Forensic Sci Res 1:4251

Ghatak S, Muthukumaran R, Nachimuthu S (2013) A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. J Biomol Tech jbt.13-2404-001. https://doi.org/10.7171/jbt.13-2404-001

Glossary for Crime Scene Investigation: guides for law enforcement. National Institute of Justice. (2021). Retrieved 13 April 2021, from https://nij.ojp.gov/topics/articles/glossary-crime-scene-investigation-guides-law-enforcement

Hess S, Haas C (2017) Recovery of trace DNA on clothing: a comparison of Mini-tape lifting and three other forensic evidence collection techniques. J Forensic Sci 62(1):187–191. https://doi.org/10.1111/1556-4029.13246

Heywood D, Skinner R, Cornwell P (2003) Analysis of DNA in hair fibers. J Cosmet Sci 54:21–27

How much DNA and RNA can be expected from human blood cells? – QIAGEN. Qiagen.com. (2021). Retrieved 13 April 2021, from https://www.qiagen.com/be/resources/faq?id=01070e31-3a4c-42d7-870c-e8005285889f&lang=en&Print=1#:~:text=Blood%20of%20a%20healthy%20individual,10x%20more%20DNA%20than%20RNA

Iwamura E, Soares-Vieira J, Muñoz D (2004) Human identification and analysis of DNA in bones. Revista Do Hospital Das Clínicas 59(6):383–388. https://doi.org/10.1590/s0041-87812004000600012

Lee HC, Ladd C (2001) Preservation and collection of biological evidence. Croat Med J 42: 225–228

Livy A, Lye S, Jagdish C, Hanis N, Sharmila V, Ler L, Pramod B (2011) Evaluation of quality of DNA extracted from buccal swabs for microarray based genotyping. Indian J Clin Biochem 27 (1):28–33. https://doi.org/10.1007/s12291-011-0154-y

Ludeman MJ, Zhong C, Mulero JJ, Lagacé RE, Hennessy LK, Short ML, Wang DY (2018) Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. Int J Legal Med 132:1555–1573. https://doi.org/10.1007/s00414-018-1817-5

Miranda G, Prado F, Delwing F, Daruge E (2014) Analysis of the fluorescence of body fluids on different surfaces and times. Sci Justice 54(6):427–431. https://doi.org/10.1016/j.scijus.2014.10.002

Mozayani A, Parish-Fisher C (2018) Forensic evidence management, 1st edn. CRC Press, pp 29–41

Pang B, Cheung B (2007) Double swab technique for collecting touched evidence. Legal Med 9(4): 181–184. https://doi.org/10.1016/j.legalmed.2006.12.003

Pepper IK (2010) Crime scene investigation: methods and procedures, 2nd edn. Open University Press

Rubio L, Sioli J, Gaitán M, Martin-de-las-Heras S (2018) Dental color measurement to predict DNA concentration in incinerated teeth for human identification. PLoS One 13(4):e0196305. https://doi.org/10.1371/journal.pone.0196305

Sammons J (2014) Digital forensics. Introd Inf Sec:275–302. https://doi.org/10.1016/b978-1-59749-969-9.00013-4

Sowmyya T (2016) Touch DNA: an investigative tool in forensic science. Int J Curr Res 8: 26093–26097

Sterzik V, Panzer S, Apfelbacher M, Bohnert M (2016) Searching for biological traces on different materials using a forensic light source and infrared photography. Int J Legal Med 130(3):599–605. https://doi.org/10.1007/s00414-015-1283-2

Tang J, Ostrander J, Wickenheiser R, Hall A (2020) Touch DNA in forensic science: the use of laboratory-created eccrine fingerprints to quantify DNA loss. Forensic Sci Int 2:1–16. https://doi.org/10.1016/j.fsisyn.2019.10.004

van Oorschot R, Ballantyne K, Mitchell R (2010) Forensic trace DNA: a review. Investig Genet 1 (1):14. https://doi.org/10.1186/2041-2223-1-14

Williamson AL (2012) Touch DNA: forensic collection and application to investigations. J Assoc Crime Scene Reconstr 18(1):1–5

5 Forensic DNA: From New Approaches for the Bio-stain Identification to the Evaluation of the Genetics Evidence in Courtroom

E. D'Orio, P. Montagna, M. Mangione, and G. Francione

Contents

Abstract

The correct application of DNA analysis in forensic field is the final objective of every forensic biologist who carries out their work to support justice by providing technical-scientific information. In recent years, the advancement of technology significantly contributed to achieve higher levels of accuracy and precision in analysis. This chapter will examine the various aspects with regard to new areas of development of forensic biology, covering both new approaches to improve the pre-analytical investigations related to the search and identification of traces and new procedure for genotyping. Lastly, the topic of the correct methodology for presenting genetic evidence to the judge will be investigated. The aim of this chapter is to allow scientific data to be appreciated in a balanced and accurate way

E. D'Orio (✉) · P. Montagna · M. Mangione · G. Francione
Bio Forensics Research Center, Angri, Italy
e-mail: eugenio.dorio@bioforensics.it

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_58

by judges, preventing any potential form of error caused by excessive or limiting evaluations relating to genetic data offered by forensic scientists.

Keywords

Forensic DNA · New approaches · Biological stain identification · Forensic biotechnology · DNA in Court

Introduction About the Bio-Stain Identification

In the field of Forensic Science, the identification of biological traces is a critical point. Forensic genetics technologies allow scientists to determine the identity of the person who released a trace on a place or object of investigative interest. Although forensic genetics techniques have extremely high levels of accuracy, techniques for the detection of latent biological traces do not have such high margins of accuracy. It should be noted that genetic analysis is considerably conditioned by pre-analytics, i.e., those operations that take place before the typical procedures of genetic investigations. Especially in a context where biological traces (Gardner and Krouskup 2018) are hardly visible, the methods of exaltation of the traces are very important; this makes it possible to identify the different traces present on an object (or on a place) with precision and further allows the operators to carry out the sampling with accuracy. If the biological traces are not visible, then there is a strong risk that operators will sample inaccurately – known as random sampling – which can generate the following conditions:

1. A failure to acquire biological trace evidence
2. The acquisition of several distinct biological traces in the same sample (generating a mixed genetic "false")

Figure 1 shows evidence coming from a crime scene with potential biological-forensic interest.

Figure 2 shows a latent biological macro-trace.

Figure 3 shows the latent biological trace enhanced with the use of special forensic lights.

There are many techniques for the detection of latent biological traces evidence. Among the tools available are as follows:

– Forensic lights
– Wood's lamp
– Luminol

Forensic lights are light sources that emit variable wavelengths of light. They generally emit light in the visible spectrum, between 350 nm and 480 nm. They enhance biological traces that are not visible to the naked eye. However, observation

Fig. 1 Shows a find coming from a crime scene with potential biological-forensic interest

Fig. 2 Shows a latent biological macro-trace

by the technician can only occur correctly if he or she uses specific visual filters for light (goggles). Several scientific industries produce a variety of forensic lights: some with specific emission, while others with variable emission from the same instrument. Thanks to the particular light emission, these instruments enhance the latent biological traces due to the fluorescence of some biological tissues (such as semen, saliva and urine), as well as to the absorbency, which is most typical of blood tissue.

Fig. 3 Shows the latent biological trace enhanced with the use of special forensic lights

Wood's lamp (Al Aboud and Gossman 2020)is a light source that emits electromagnetic radiation mainly in the ultraviolet range, to a negligible extent, in the field of visible light. The use of Wood's lamp, even for long periods of time, is not toxic for humans; in fact, this use involves the emission of ultraviolet radiation called UV-A, which does not generate phenomena of DNA alteration (unlike UV-B and UV-C). Wood's lamp produces light that is not directly visible to the human eye; however, it can be used to illuminate materials on which ultraviolet radiation induces effects of florescence. In the biological-forensic field, Wood's lamp is used to search for organic traces that are not visible to the naked eye. Many other applications of Wood's lamp are possible: from the fight against banknote forgery (documentary forgery) to medical, microbiological, commercial, and paleographic applications.

Luminol (Barni et al. 2007) (also known as IUPAC 5-amino-2,3-dihydro-1,4-phthalazindione) is a chemical compound often used by the Judicial Police to detect latent blood. Luminol is a liquid compound, which is sprayed onto the surface of interest directly on-site. It is a very versatile substance which, when mixed with an appropriate oxidizing agent, develops blue luminescence.

Luminol is particularly useful to operators when looking for traces of "washed" blood. In fact, once it is sprayed on the surface of investigative interest in dark conditions, the luminol generates luminescence. This reaction occurs due to the chemical reaction and, more specifically, to the presence of hydrogen peroxide. In the presence of iron, the reaction is activated and luminescence is developed (for about 40–60 s). Iron is one of the typical components of hemoglobin, a protein present in the blood that functions as an oxygen carrier. However, the luminol reacts in a non-specific way with everything that contains iron, which is why it can generate

multiple "false negatives." The most classic of these is the reaction of luminol with bleach.

It should be noted that all technologies for the search of latent biological traces are presumptive, never confirmatory.

It should also be noted that, unlike what happens in DNA analysis for forensic purposes, the analysis aimed at identifying latent biological traces is still without fixed protocols. Only generic guidelines are presently available.

Obviously, this generates a condition of subjectivity in the search for latent biological trace evidence. It is therefore appropriate that a highly specific research activity is carried out, dedicated to the validation of protocols for the use of these technologies and in order to ensure reproducibility and homogeneity of the analysis results.

New Approaches for the Bio-Stain Identification

The potential of screening technologies for latent biological trace evidence is strong, but it still requires further validation.

The definition of an operational protocol would, in fact, ensure the controlled and controllable performance of the acts of biological trace evidence research at every stage.

Currently, partly due to the presence of the guidelines, and partly due to the presence on the market of multiple types of technological tools useful for the research of biological traces, it is not possible to conduct this preliminary phase of genetic analysis in an extremely accurate and standard way.

It should also be preliminarily considered that the subject of "biological trace evidence research" has taken a back seat to DNA studies for forensic purposes. Yet, a successful outcome of the trace research phase directly affects the success of the genetic analysis that will result, and as such, it is therefore deserving of greater scientific attention.

In addition, there are many publications (Baranowska 2016) and case reports that demonstrate the validity of the application of technologies for the screening of biological traces, as well as how this research, done in a targeted way, can lead to excellent investigative results.

Considering the potential application of the technologies for the biological evidence detection in the forensic field, certainly not inferior to those of DNA analysis, there are still active research projects aimed at the validation of standard operating protocols for the development of the "bio-stain identification" phase.

A first study was carried out at the University of Copenhagen's Faculty of Forensic Medicine, Forensic Genetics Section, Crime Unit group; this study has brought to light some very important data upon which increasingly accurate and refined research projects in the field can begin to develop (D'Orio et al. 2018). Firstly, these studies have generated a "reference sample," thanks to which the ability of the human eye to see biological traces of the size of 50 μl was calculated. To carry out this evaluation, different types of substrates were used, and the biological trace

was adhered to. The data analysis showed that the human eye is able to reveal only about 15% of all existing biological traces under the standard conditions considered.

Figure 4 shows the detection increase (in %) using "the human eye," an ALS protocol, and a second ALS protocol.

This data fully confirms that, for a high-quality forensic investigation, it is essential that operators use specific tools for the identification of biological traces.

The study also led to the definition of new parameters that were previously unconsidered, but which, however, must necessarily be taken into account:

1. The color of the substrate – find – on which the forensic operator conducts the search for biological trace evidence
2. Minimum Quantity Detectable (MQD), i.e., the minimum amount of biological trace that can still be detected

The study concluded that there is a direct correlation between substrate color and MQD. This is a necessity that forensic practitioners must take into account in two stages:

1. At the moment of research concerning the biological traces on the findings
2. At the moment of presentation of the data in the technical report/in court. In fact, the forensic scientist must describe not only the number of latent biological traces found but also the maximum degree of accuracy that can be reached starting from the examination of that specific type of evidence/result (you choose). All this is necessary to ensure the correctness in the communication of scientific data to judges and lawyers and to avoid that the weight of the scientific data that arises may be overestimated or underestimated.

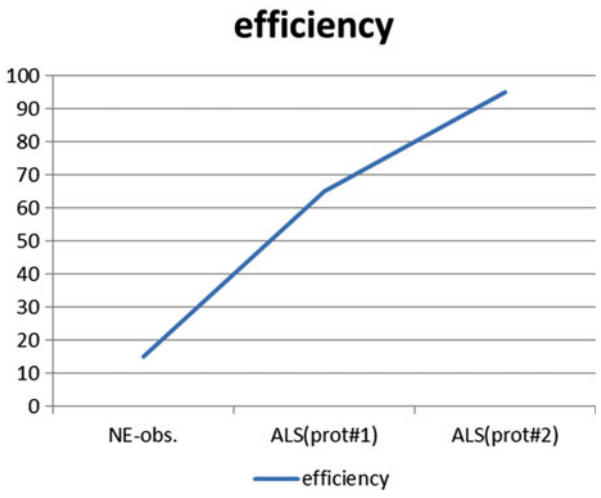


Fig. 4 Shows the detection increase (in %) using "the human eye," an ALS protocol, and a second ALS protocol

In detail, the study also showed that there is a need for at least two different protocols for the use of detection technologies that, when used in synergy, allow forensic biologists/investigators (you choose) to achieve an operational efficiency of more than 95%.

Figure 5 shows that the use of a single protocol, in fact, did not allow the efficiency to exceed 65%.

A detailed study of the data identifies the cause of this limitation. In particular, textile substrates with dark and medium-dark colors, examined with this particular type of screening protocol, and did not provide positive results.

Figure 6 shows the particular textile fibers, used as substrates, from which it was impossible to make a successful identification of biological traces.

For these reasons, a different protocol of use of the same instrument used for the search of biological traces has been standardized, and the samples found to be negative at the first examination have been reanalyzed.

Figure 7 shows the data related to this synergistic screening.

In terms of absolute efficiency, with the synergistic combination of the two experimental protocols, latent biological traces are detected in 19 samples out of 20 (all in triplicate, to confirm the reproducibility of the data), with a relative efficiency of 95%.

Data were also collected in mixed latent biological trace contexts. In these cases, the data showed that the relative efficiency is even more accurate.

Figure 8 shows the efficiency of technological tools for the detection of latent biological traces composed of two different biological tissues in equal concentrations.

Figure 9 shows the efficiency of technological tools for the detection of latent biological traces composed of two different biological tissues in different concentrations.

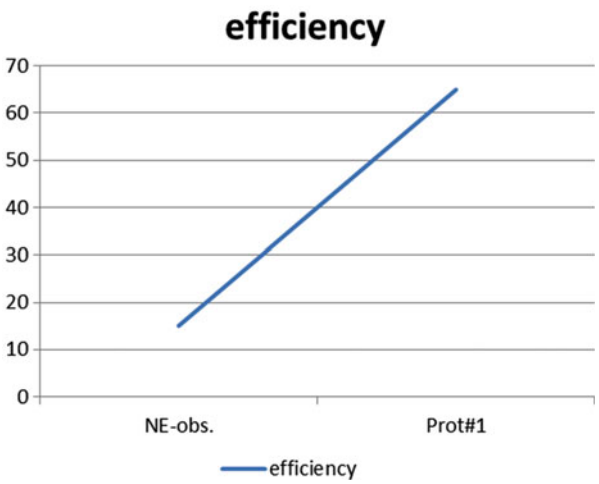


Fig. 5 Shows that the use of a single protocol, in fact, did not allow the efficiency to exceed 65%

ALS PROT. #1	Black t-shirt	Grey shirt	jeans	Blue leggins	White t-shirt
Blood	-	+	+	-	+
Semen	-	+	+	+	+
Saliva	-	+	+	-	+
Urine	-	+	+	-	+

Fig. 6 Shows the particular textile fibers, used as substrates, from which it was impossible to make a successful identification of biological traces

Protocol1	Black t-shirt	Grey shirt	jeans	Leggins blue	White t-shirt
BLOOD	-	+	+	-	+
SEMEN	-	+	+	+	+
SALIVA	-	+	+	-	+
URINE	-	+	+	-	+

protocol2	Black t-shirt	Grey shirt	jeans	Leggins blue	White t-shirt
BLOOD	+	+	+	+	+
SEMEN	+	+	+	+	+
SALIVA	+	+	+	+	+
URINE	-	+	+	+	+

Fig. 7 Shows the data related to this synergistic screening

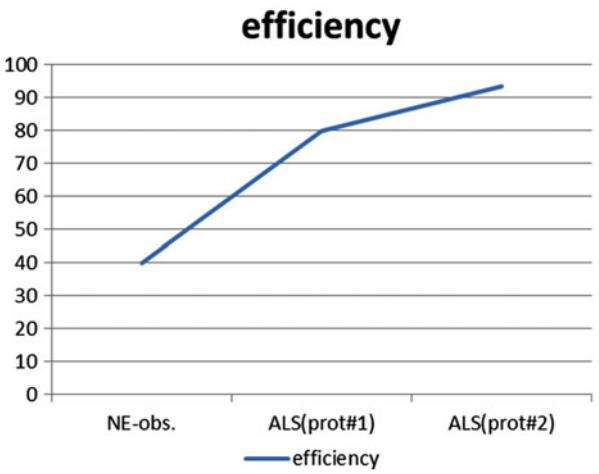


Fig. 8 Shows the efficiency of technological tools for the detection of latent biological traces composed of two different biological tissues in equal concentrations

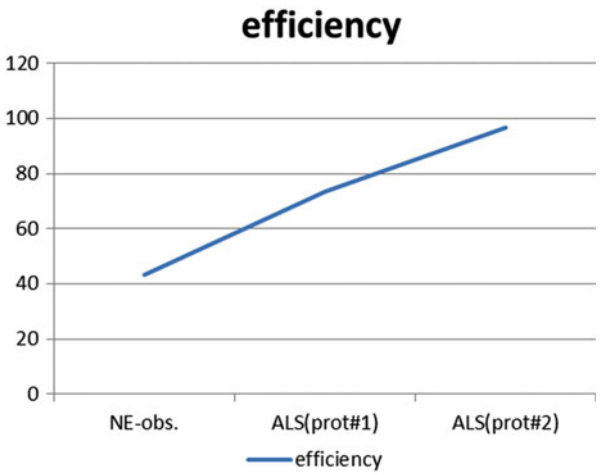


Fig. 9 Shows the efficiency of technological tools for the detection of latent biological traces composed of two different biological tissues in different concentrations

The experimental study also showed that, in terms of absolute latent trace, under certain conditions, with protocol 1, the forensic operator is able to detect and correctly document biological traces consisting of 0.78 μl.

Nevertheless, it should be pointed out that although this technology has presently shown most interesting data in terms of identification of the biological trace evidence, it is not yet able to lead to information on the histological tissue from which the biological trace under examination derives.

Further research activities are currently in progress; these activities are precisely aimed at the definition of an operational protocol that will allow the following:

1. To identify latent biological traces
2. To facilitate homogeneity of procedures and data
3. To identify the histological tissue from which the biological trace evidence under examination is generated (e.g., blood, semen, saliva, etc.)

Other recent experimental activities have been focused on the analysis of biological traces exposed in critical contexts, such as fire conditions – fire scene (D'Orio et al. 2020).

Analysis Techniques in Forensic Genetics and New Technologies

Since the publication of the article "The forensic use of the genetic fingerprint" by Alec Jeffreys, Peter Gill, and David Werrett in 1985, the application of genetics in the field of forensic science has undergone an exponential development due to the rapid evolution of methods of analysis, such as the discovery of the polymerase

chain reaction (PCR). This is a widely used method to produce thousands of miles of copies of a specific DNA sample, allowing scientists to take a very small sample of DNA and amplify it to a suitable amount to study it in detail. PCR was developed in 1984 by the American biochemist Kary Mullis (1987). At that time, a large sample of biological material was needed in order to generate a full DNA profile for forensic analysis.

In fact, before the introduction of this technique, human identification was carried out through the characterization of human protein polymorphisms. In particular, the ABO, Rh, and HLA systems were analyzed using immunological and electrophoretic techniques (Spinella et al. 1997). However, this method gave poor results, relating to the type and size of the evidence. The analysis of protein polymorphisms has been progressively replaced by new molecular techniques, which analyze DNA polymorphisms.

The sequencing of the human genome showed that only 1.5% of DNA is composed from coding regions, while the remaining 98.5% is represented by non-coding sequences. These sequences can be present both in single and in multiple copies. In total, the repeated DNA make up more than 50% of the whole genome. Unlike coding sequences, which are highly conserved, non-coding sequences are subjected to mutation events that lead to an increase in inter-individual differences that often do not affect the phenotype (Fowler et al. 1988). However, it is useful to point out that most of the human genetic material (more than 99.5%) does not vary between individuals. Therefore, only a small fraction of our genome, less than 0.5%, is subject to variability. This minimal dose of variability makes each individual unique, and through its analysis it is possible to use this information for personal identification.

DNA variability can be explained by the concept of polymorphism: a DNA sequence is defined as polymorphic when at least two allelic forms are present in the reference population. The less frequent of these allelic forms is present with a frequency greater than or equal to 1%. In the case that the frequency of an allele in the population is less than 1%, this form is defined as a rare variant.

Forensic genetics uses DNA polymorphisms for personal identification. The power of discrimination of the polymorphisms used for this analysis is based on their ability to distinguish two individuals in a population. The ability to discriminate between genetically unrelated individuals depends on the number of alleles characterizing the locus and their distribution in the reference population. Polymorphic regions are presumed on both autosomal and sex chromosomes.

There are two categories of DNA polymorphisms, based on the molecular mechanism that gives rise to such variability: sequence polymorphism and length polymorphism.

The first polymorphism, also known as SNP (single nucleotide polymorphism), arises from a single nucleotide substitution determining differences between two homologous DNA sequences. The SNPs basically produce biallelic polymorphisms for this they can provide limited information for personal identification. In order to obtain sufficient discriminatory power for personal identification, it would be necessary to analyze a set composed of fifty Snps (Sanchez et al. 2006).

Because SNPs are smaller and more abundant in each cell than STRs, SNP analysis can be useful when DNA is highly degraded.

The VNTRs (variable number of tandem repeats) are length polymorphisms consisting of a variable number of DNA sequences repeated in tandem. Each characterized by multiallelic forms, each defined by the variable length of the sequence and the number of repetitions. VNTRs are divided into two categories, based on the size of the repeating sequence: *minisatellites* which are characterized by repetitions in an interval ranging from 16 to 70 bp and *microsatellites*, or STRs (Short Tandem Repeat) with a repetition interval that ranges from 2 to 6 bp (Weber and May 1989; Beckman and Weber 1992; Hearne et al. 1992).

STRs are the markers used in forensic genetics. For the high degree of polymorphism and the high rate of heterozygosity, they can be used as genetic markers to generate a DNA profile that is extremely rare in a population of unrelated individuals (Biondo et al. 2001).

Typically, markers are examined at a minimum of 16 loci plus a sex marker in an individual's DNA. These are the European Standard Set (ESS) (set out from the European Network of Forensic Science Institute (ENFSI) and the European DNA Profiling Group (EDNAP) (Welch et al. 2012). These are visualized as a series of peaks on a graph and the position of which corresponds to the length of the STR and is recorded as a number. Each location has two STRs (one from our mother and one from our father), which means that an individual's genetic profile can be represented as a series of numbers. Each pair of digits always corresponds to a specific location on a chromosome.

Commercial kits that amplify more than 20 STR loci have now been adopted by many laboratories worldwide. These new kits enable more international sharing of DNA data with increased compatibility between STR data, and as more loci are examined, the probability of two individuals having an exact match decreases rapidly.

In fact, when a DNA profile has been generated from a crime scene sample, it can then be compared to other profiles, such as DNA from other crime scenes, the suspect's DNA, the victim's DNA, or DNA profiles within a National DNA Database. If two profiles are identical, this is called a *full match*; if only parts of the profiles match, this is a *partial match*. Once a match has been declared, the strength of evidence supporting the identification of an individual can be calculated. It is important to calculate the probability that the DNA on the evidence collected will match that of someone else. It depends on how many loci in the DNA we look at.

Although the chances of two full DNA profiles from two different unrelated individuals matching are extremely rare, DNA profiles from crime scenes are rarely perfect, and they may not contain information about every genetic marker analyzed resulting in a partial DNA profile or in a mixture of DNA from two or more people. The smaller the number of genetic markers the DNA profile is composed of, the greater the risk of a false match occurring.

Once a DNA sample has been analyzed, two kinds of statistics may be reported. The simplest is the *match probability*, which addresses the question of how rare a DNA profile is in a population of random unrelated individuals. This must not be

confused with how likely the person is to be innocent of the crime. These match probabilities are fine if we are dealing with full DNA profiles from a single individual, or we have a partial DNA profile from one individual as well. However, this statistical approach cannot be used in more complex cases, such as mixtures of two or more individuals. Under these contexts, a second kind of analysis is applied: a *likelihood ratio*. This weighs the evidence in favor of competing hypotheses, one from the perspective of the prosecution and one that of the defense. It compares how probable the observed evidence is under each argument. Even if statistical experts agree that likelihood ratios are the best approach for complex DNA profiles, their adoption by many labs has proven to be slow. This is partially done to fear that courts may misunderstand the results. In fact, due to the complexity of the calculations, particularly once several people's DNA gets mixed together, forensic genetics now uses specially designed computer programs. Several such programs are employed around the labs, but because different programs are prepared using different mathematical approaches and assumptions about the data, they could give different values for the likelihood ratio. There have been cases where the prosecution has used one program and the defense another, meaning that the court had been presented with two different answers about the strength of the DNA evidence. Usually the difference is slight, but cases where one program favors the prosecution and another favors the defense are obviously important and need further investigation. To assist in the appropriate use of one of the commonly used mixture interpretation approaches, some rules for the combined probability of inclusion (CPI) were spelled out (Bieber et al. 2016). The most popular probabilistic genotyping software (PGS) to assist DNA mixture interpretation is generally systems use a "semi-continuous" model that uses the presence or absence of peaks along with probabilities of allele drop-out or drop-in or a "continuous" model that takes peak heights into account as well as the presence or absence of peaks along with probabilities of allele drop-out or drop-in.

Several validation studies were published for the PGS system STRmix including the challenges of estimating the number of contributors with low levels of DNA were explored (Norsworthy et al. 2018). The variation of results with four different continuous PGS models was studied (Swaminathan et al. 2018), and responses to court admissibility challenges with STRmix were provided (Buckleton et al. 2019).

However, as forensic techniques have improved, their ability to detect smaller and smaller amounts of DNA has increased. This means that tiny, invisible traces of DNA can now be recovered and analyzed.

The analysis of such limited quantities of DNA is called low copy number (LCN) (Gill et al. 2000) or LT DNA typing. Under these conditions, stochastic substitutes in different amplification reactions (of the same sample) that produce different replication result are defined unstable genetic profiles (Whitaker et al. 2001). In fact, these apparently conflicting results, resulting from polymerization artifacts, the univocal determination of the genetic profile of a biological sample is not obtained (Gill et al. 2007).

With LCN sample, the electropherogram does not reflect the real DNA profile composition. This is due to the formation of stochastic variations (drop in, drop out, allelic imbalance and stutters) during sample amplification. These stochastic effects

introduce a high degree of uncertainty and difficulty in the interpretation of LCN genetic profiles. A method to check the accuracy of genetic profiles obtained from complex forensic samples consists in the preparation of replicate assays, aimed at investigating the repeatability of the result in distinct PCR reactions. To do this, at least one duplicate of independent amplifications is required (Graham 2008; Buckleton 2009). In this way, multiple amplifications of the same extracted DNA can be compared in order to obtain a "consensus profile." As such, alleles that occur more than once are considered "reliable" as they are reproduced by separate assays (Cowen et al. 2011).

In addition, the Y-chromosome DNA testing is important for forensic evidence. It can be crucial in sexual assault cases, making the examination of a male perpetrator's profile possible even in a mixed sample with high levels of female DNA concentration. It makes use of genetic information on the Y chromosome (which only males have). The Y chromosome is inherited from father to son only, which means that all male relatives on the paternal side of the family will normally share the same Y chromosome. Consequently, this type of analysis is not identifying, because the male individuals of the same family have the same Y DNA profile. The new kit containing a large battery of Y-STR loci is expected to be very helpful to increase the power of discrimination and help differentiate male individuals within the same family in a higher percentage of cases (Ottaviani et al. 2014).

Mitochondrial DNA analysis uses DNA from mitochondria. Mitochondrial DNA (mtDNA) is more abundant than other types of DNA and can be useful in cases where biological material is limited or damaged by environmental exposure, such as heat, light, or water, which breaks up the DNA strand. MtDNA is inherited by a child from its mother, so all relatives on the maternal line of the family will share the same mtDNA. Like the Y-DNA analysis, it is not identifying but can be useful for exclusion.

Next-generation sequencing (NGS) opens potential new applications including biogeographical ancestry, phenotyping of externally visible characteristics, and finer details on STR alleles to possibly improve mixture component resolution. In fact, NGS describes a suite of emerging DNA sequencing technologies, where sensitive tests can be done simultaneously and STR profiling, biogeographic ancestry testing, and phenotyping tests at the same time. Its use in forensic science is still starting out, but a special issue of the journal *Electrophoresis* on novel applications of MPS in forensic DNA analysis was published in "November 2018" (McCord and Lee 2018).

The current limits for implementation of NGS were identified as lack of consistent nomenclature and reporting standards, lack of compatibility with an existing national DNA database, lack of population data to support statistical calculations, and lack of an adequate legislative framework. In May 2019, the NDIS Board of the FBI Laboratory began accepting data from approved NGS kits for upload to the US national DNA database.

Biogeographic ancestry testing is a technique that enables an individual's broad geographic origins (Africa, Asia, Europe) to be estimated based on genetic differences in their DNA. This method uses DNA markers that are common in different parts of the world and can help narrow down a pool of suspects when no match in a national DNA database has been found.

A new application is the development of forensic DNA tests that can predict aspects of someone's physical appearance. This approach is called forensic DNA phenotyping (Kayser 2015). This technique uses DNA to make predictions about someone's appearance (hair and eye color). It is another way of narrowing down a pool of suspects using DNA markers found in the genes that determine aspects of human appearance. As it is a new technique, it has been used in a very small number of forensic cases. In fact, not all externally visible characteristics are equally predictable from DNA information at present. Predicting eye color is difficult, due to the fact that it is influenced by many genes, of which six are currently used in forensic DNA phenotyping tests. Predicting other externally visible traits such as height is equally as difficult and currently not yet possible because they are determined by large numbers of genes, many of which remain unknown. They are also influenced by environmental factors which cannot be predicted from DNA. Phenotyping can be used as an investigative tool, to reduce the number of potential suspects when the suspect pool is very large and to help prioritize who to focus on first or next. But it cannot be used as final evidence in court.

It is currently possible to predict eye and hair color from a DNA sample, although none of these tests are 100% accurate. Some of these tests have been forensically validated (Chaitanya et al. 2018). Skin color is likely to be the next appearance trait that forensic scientists will be able to predict from DNA with tests that are currently being developed and validated.

Forensic DNA phenotyping raises some ethical issues too. Whereas standard forensic DNA profiling involves genetic markers found in parts of the human genome that are non-coding regions, the markers used in forensic DNA phenotyping are located within or close to genes involved in coding regions. If forensic DNA phenotyping techniques were extended to also include non-visible characteristics, such as genetic risk of disease, they could reveal sensitive and private information. This can be avoided through national regulation.

With the goal of generating faster DNA results, Rapid DNA instruments have been developed that can produce a DNA profile in less than 2 h. These instruments offer a full automation of the STR typing process consisting of DNA extraction, amplification, separation, detection, and allele calling from reference oral swabs. Two rapid DNA instruments are currently available: the ANDE 6C (6-color) Rapid DNA System from ANDE (Longmont, CO) and the Rapid HIT ID instruments by Thermo Fisher Scientific. These instruments can be operated at police booking stations and border crossings and in traditional forensic laboratories.

Currently, these instruments would appear to be used successfully to process single-source reference samples and not crime scene evidence containing mixtures (SWGDAM; NDAA).

Genetic Traces in a Fire Scene

In an investigative-forensic context, genetic traces have a central importance in reconstructing the dynamics of the events that occurred in the fire scene to identify the persons responsible for the crime.

However, often with the express purpose of erasing the traces, many criminals use fire as a tool to weaken the investigative process, making the biological traces potentially present on a given crime scene or on finds less usable and available.

One of the greatest merits of *Salvatore Ottolenghi*, father of the Italian school of forensic sciences, was to understand how the police investigation should be traced back to the more general area of natural sciences.

The biological traces present on the scene of the crime are, in fact, "natural signs."

It is no coincidence that English criminologists speak of physical evidence to mean circumstantial evidence.

Therefore, the places and things that relate to a crime, if properly examined, have much to reveal, identity of the victim, on that of the aggressor, on their relationships, and more generally on the most fleeting or deep interrelations between human action and the environment.

Genetic Traces in a Fire Scenario

An investigative context of a fire scene with the presence of genetic/biological traces is certainly complex, as it requires the experience and skills of different professionals: on the one hand, forensic biologists and, on the other, forensic engineers (Augenti et al. 2011; Mangione et al. 2015).

In fact, for a suitable investigative analysis in these cases, only the synergistic work between these two professional figures can, in fact, trigger a forensic investigation that has excellent objective chances of success.

Specifically, in these investigations, there is a synergy between the forensic biologist, who is a specialist in biological trace evidence, and the forensic engineer (fire investigator), who is a specialist in scenarios of confined fires.

There is a synergy between the forensic biologist, who is a specialist in biological trace evidence, and the forensic engineer (fire investigator), who is a specialist in scenarios of confined fires.

The scientific data on this topic present but not in considerable quantities.

However, it is known that, even following a fire, the biological traces have a fair chance of being preserved and suitable for investigative use.

In fact, the degradation of biological traces is "outbreak-dependent"; in the place where the spread of the fire and/or possibly a subsequent explosion occurs, biological traces, being particularly exposed to the phenomenon of fire, suffer considerable structural damage.

Such damage that biological cells receive cause cellular and genetic lysis phenomena, often leading to the impossibility of obtaining a suitable genetic profile for future investigative use.

Furthermore, some studies describe how, in the event of a fire, the biological traces have a "substrate-dependent," that is, the material on which they deeply affects is deeply affecting the state of the biological traces.

It has been well described that biological traces present on substrates such as "nylon" have a very considerable degradation in the event of a fire.

Furthermore, it is also highlighted that blood traces assume, following a fire, different phenotypic (i.e., visible) characteristics depending on the body tissue considered.

It has been shown that traces of blood take on a color ranging between dark brown and black, following a fire; on the contrary, traces of sperm remain unchanged in their color.

This occurs due to the particular protein composition of the blood. In the blood, cells are rich in the protein "hemoglobin," which contains a red pigment.

In the event, or prolonged exposure over time to high temperatures, a colorimetric change occurs; the red pigment of hemoglobin, due to heat-induced denaturation, goes toward darker colors, such as dark brown.

These colorimetric characteristics are sometimes detectable with the naked eye, but sometimes they are detectable thanks to the use of special forensic lights that are able to enhance the latent traces that are no longer visible to the human eye due to their small size or high degradation.

Furthermore, other sector studies have simulated a fire scenario in a multi-room house. A known series of biological traces have been placed in different rooms.

Following the fire, it was possible to see that in the ignition room, the one from which the flames originated, the biological traces still usable, and identifiable amounted to only 25% of the total traces deposited before the fire.

Otherwise, in the other rooms, also affected by the flames, the biological traces still identifiable and usable for investigations amounted to 80% of the total traces.

This scientific data give considerable confidence to the investigators, as it is highlighted that the total compromising of biological traces does not occur and that in the relevant environments, even if affected by the flames, will be considerable probability of still finding traces using biological products.

This data, from an investigative-scientific point of view, is quite satisfactory and above all confirms that, even if the perpetrators of a crime try to erase the biological traces on a crime scene, even through the use of fire, there is a good chance of finding biological trace evidence of the perpetrators despite the fire scenario used to weaken the investigation.

It should also be emphasized that the degradation of biological traces due to fire damage, from a biological point of view, can always be assessed in two ways: the biological cell is composed, structurally speaking, of nucleus and other intracellular organelles.

In the nucleus, as well as in the mitochondria, there is genetic material which, by practice, is analyzed for investigative purposes. However, it must be specified that the genetic material of the nucleus and the mitochondrion is profoundly different. Indeed, nuclear DNA is the only one of the two that allows, through the use of particular molecular genetic techniques, to define a genotypic profile.

The DNA present in the mitochondrion does not have this peculiarity because it is transmitted intact, from generation to generation, from mother to child.

For this transmission phenomenon, the mitochondrial DNA, speaking, can only give information about the maternal family strain of the subject that released the trace.

Another important difference between these two DNAs, which certainly is applicable in the assessment of the progressive biological damage caused by fire, is the structure of the two DNAs.

Nuclear DNA is a molecule composed of 46 chromosomes, a double-stranded helix of nitrogenous bases. Instead, mitochondrial DNA is a single circular molecule. The spatial conformation, circularity of the molecule in the case of mitochondrial DNA, ensures greater resistance to all external degradation phenomena, including fire or heat.

It should also be added that nuclear DNA is present in a single copy within a biological cell, while mitochondrial DNA is present from one hundred to one thousand copies per biological cell. Therefore, thanks to the numerical presence and conformation factors of the molecule, the damage that is progressively produced by the heat of the fire can be assessed by referring to the nuclear DNA/mitochondrial DNA degradation index.

The following figure represents the biological finds most present on crime (Fig. 10).

The Biological Traces Evidence

Biological evidence in fire investigations can be of various natures. In order to be able to catalog them, they must be part of a standardized investigative methodology that represents a potential aid for the search of evidence in fire investigations activities applicable in all cases of confined fires.

The need to code the operations is strongly felt in the environments of the scientific police force, where the user, who is interested in conducting investigations, often does not have a clear overall picture of the operations and controls to be carried out on the scene to by-pass certain checks and therefore not to find specific traces in the forensic field.

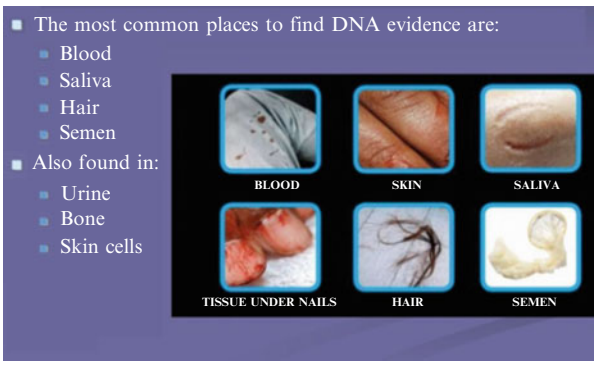


Fig. 10 Overview of the main biological found on the crime scene

Fig. 11 Colorimetric modification of biological traces on the same substrate following a fire

Of considerable importance is therefore the collection of significant events of the scene to be performed immediately after the fire, in order to allow the scene to be frozen at the time of collection and when the state of the crime scene has not yet been compromised by third parties.

On occasion, some details may escape even with the most attentive investigator and, subsequently, become a relevant *unexpected* in the continuation of the investigation. In these terms, for example, the intervention of the Fire Brigade, the presumed duration of the fire, photography supplemented by descriptive reports, etc. can assume proof value at a forensic level.

There are therefore various critical issues in fire scenarios in the search for biological evidence.

By way of example, the phenotypic characteristics induced by fire on biological traces are presented in the following way (Fig. 11).

During the evidence, it is necessary to be attentive to those events that could involve alterations of the scene and clues of a biological nature.

For example, the fingerprints present in a malicious scene can often be compromised due to the necessary shutdown of the fire.

From Scientific Proof to "Beyond Reasonable Doubt"

The investigation presupposes the occurrence of a criminal event and starts with the news crime, with the completion of direct and indirect investigations (Mangione 2017).

Direct investigations are also called objective probative acquisition investigations, since they are carried out directly on things, places, or situations relevant to the crime and involve an analysis of the elements found on the scene. For example, these investigations include planimetric and photographic surveys, laboratory analyzes on finds, and so on (Mangione 2018).

Indirect investigations, on the other hand, or investigations of subjective probative are subsequently carried out and in tandem with direct ones carried out subsequently and in parallel with direct ones.

A different but related problem is whether the judge is able, not being an "expert in the field," to understand if the resulting evidence is altered or is able to perceive any changes.

In reality, the investigations carried out are called into question only in the cross-examination between the parties, and it is at this moment that the issues emerge that will then form the topics for the evaluation of the evidence.

In summary, the judge's control must be exercised both at the beginning, at the moment of admissibility of the test, and at the end, in the evaluation of the result. This judicial control is enacted to ensure a new line of thought in legal doctrine, carried out by judge G. Francione, suggests the establishment of figures of active support in these preliminary stages (Francione 2019).

Specifically, we are talking about the figure of the "pro-unknown consultant," employed by the judicial judiciary who supports the PG in the investigations and has the task of ensuring the quality of the work and submitting observations and requests aimed at guaranteeing maximum efficiency and compliance with all scientific and legal procedures in carrying out the first technical investigative acts.

The causal relationship cannot be considered to exist on the basis of the statistical probability coefficient alone, but must be verified in the same way as a judgment of high logical probability.

The reasonable term indicates insufficiency, contradiction, and probative uncertainty; therefore, the plausible and reasonable doubt is based on specific elements which, based on the available evidence, corroborate it in the specific case.

The search for traces is mainly aimed at identifying, documenting, and also removing fragments of papillary prints, which at the crime scene can be of two types: visible and latent prints.

The visible footprints are those:

- Which are produced by contact of digital surfaces soiled with substances of various kinds (*blood, ink, stain,* etc.) on rigid surfaces, generating fingerprints by overlapping
- Produced by the pressure or sinking of the papillary ridges on malleable substances, such as wax and so on, generating impressions for modeling.

Generally these types of imprints concern surfaces that can be removed with the entire substrate on which they are impressed and must therefore be photographed with the appropriate technical devices, such as filters, polarized light, in order to enhance the contrast with the surface, same on which they are located.

Requests for intervention aimed at enhancing fragments of footprints on the dashboards of motor vehicles, made of plastic material, are increasingly frequent.

Figure 12 shows a technique for removing soot in order to identify the presence of papillary prints and traces of blood left on the surface of a door before the fire.

Figure 13 shows instead some impressions on ceramic, highlighted by vacuum metal deposition (VMD), after the surfaces have been respectively exposed to the following:

Fig. 12 Extrapolation of footprints from a door

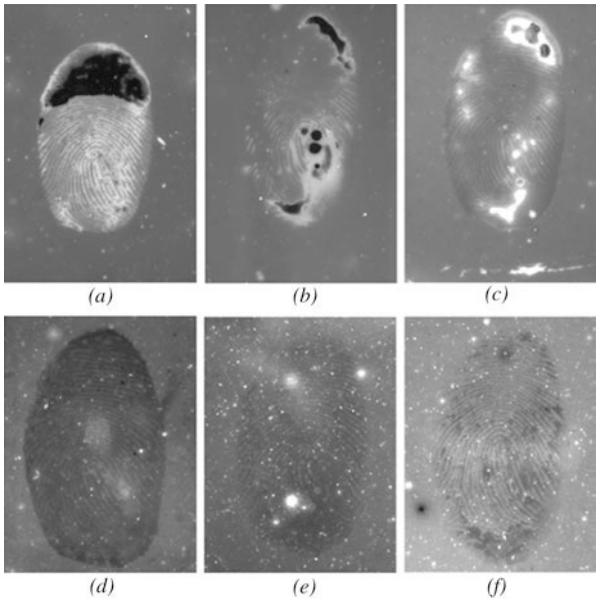


Fig. 13 Imprints on ceramic material as the temperature varies

- A) 500 °C 5 min
- B) 500 °C 15 min
- C) 700 °C 5 min
- D) 700 °C 15 min
- E) 900 °C 5 min
- F) 900 °C 15 min

Even in the extreme conditions of a fire, it is therefore possible to identify traces of interest for forensic purposes.

In this regard, the work carried out by Harper in 1938 showed that there is the possibility of identifying papillary traces on objects exposed to a temperature between 100 and 200 °C more clearly demonstrated in Fig. 14.

DNA and Fire

The effects of heat on DNA do not simply just create damage. Scientific research on DNA evidence and fire investigation is undergoing enormous growth, especially abroad.

Currently, scientific research in the field of forensic genetics is tending more and more frequently toward the accuracy and assurance of the robustness of analytical data. In this case, studies on compromised DNA and studies on the interpretation of mixed profiles are currently underway in many universities.

The study of the methodologies for the analysis and interpretation of degraded DNA also closely concern the fire investigation.

In fact, with the implementation of analytical methodologies, genetic information from fire scenarios will certainly have a greater yield and, consequently, also greater application in the investigative field. In other words, these methodologies are aimed at recovering, and making ~~it~~ usable for analysis, the genetic traces that today are "lost" because they are too degraded.

Figure 15 shows heat denaturation process of nuclear DNA. It is noted how double-stranded DNA can be denatured.

Physiologically speaking, DNA is a molecule composed of a double helix held together by non-covalent bonds between the nitrogenous bases that make the genetic structure.

In order to carry out normal vital functions, the cell continuously "packs" and "unpacks" the DNA, thanks to the use of particular enzymes and a highly controlled biochemical process of reactions.

In fact, just think that biological cells normally replicate.

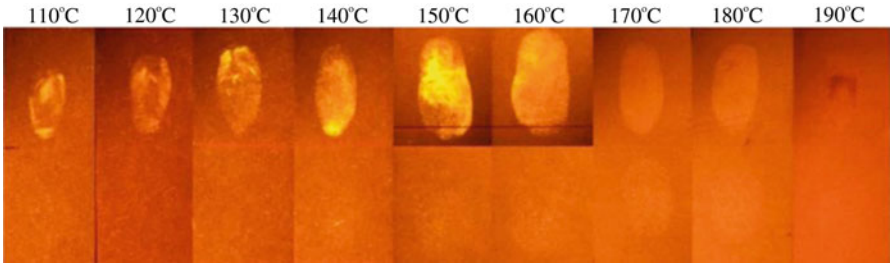


Fig. 14 Papillary traces as the temperature changes

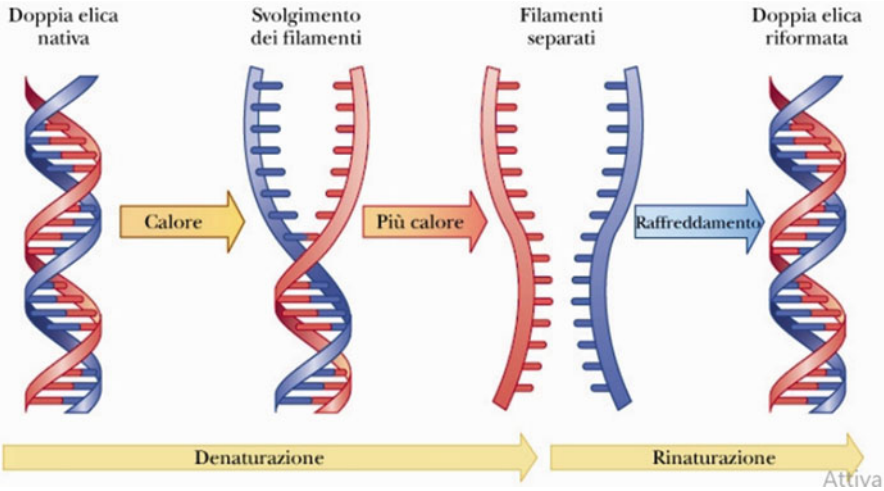


Fig. 15 Heat denaturation process of nuclear DNA

When cell replication occurs, the DNA double helix must open (and it does so, thanks to specific enzymes) and must replicate bi-directionally.

This phase, the cd, called "DNA synthesis," is the basic phase of the cell division process. Thanks to this base, the two daughter cells (which will be equal to each other and equal to the mother cell that generated them) will have the same DNA.

The modern technologies of molecular and genetic studies have led to the discovery, years ago, of the famous PCR (Polymerase Chain Reaction) technique, according to which DNA, due to controlled heat administration, first denatures and later realigns itself later.

This process was used for research in the health diagnostics field. It was later seen that the same technique was also useful in applications in an investigative-forensic context.

DNA can denature at different temperatures based on its internal composition (A-T; C-G) (the nitrogenous bases that make up the DNA are four, namely adenine, thymine, cytosine, and guanine); these comprise the DNA structure by pairing with each other through the formation of non-covalent bonds. Specifically, adenine binds with thymine, and cytosine binds with guanine.

Other bonds between these molecules are not chemically possible.

Figure 16 below shows how A-T-rich DNA denatures at a medium temperature compared to C-G-rich DNA. This can be explained, thanks to elementary principles of biophysics; in fact the pair of nitrogenous bases C-G is united by three non-covalent bonds, while the pair A-T is united only by two bonds. This is why heat-induced denaturation is relatively simpler for A-T-rich DNA stretches.

Obviously, all this confirms the principle according to which heat, simulated or coming from a fire, has effects on the DNA structure and, having reached certain

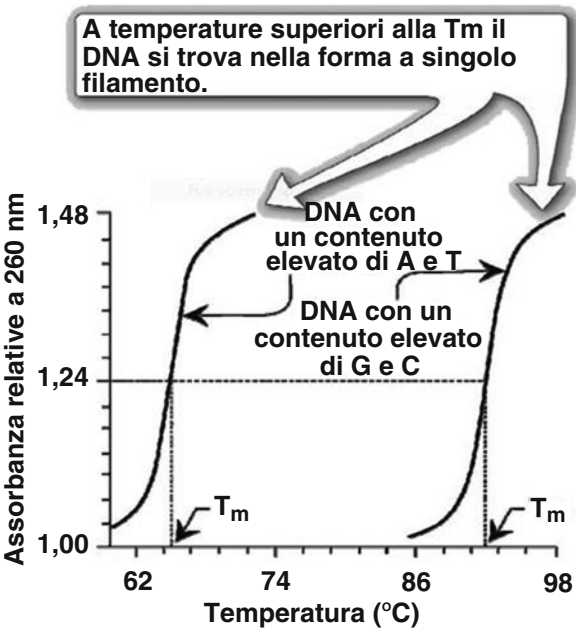


Fig. 16 Average temperatures at which the denaturation of the nitrogenous bases making up the DNA double helix takes place

levels, causes alterations of the genetic structure that are initially repairable, then irreparable depending on the time and temperature reached.

There are two types of DNA in nature: nuclear and mitochondrial. These types of DNA are found in practically all cells of the human body (with the exception of erythrocytes, which have no nucleus).

These types of DNA have very different characteristics and peculiarities. Physiologically, nuclear DNA is present in duplicate within the nucleus of human cells (this is why the human being is genetically called "diploid"); the nuclear DNA is inherited directly from the biological parents of the individual with equal distribution (50% of the genetic material comes from the biological father, and 50% comes from the biological mother); nuclear DNA is present in a "double helix" conformation.

Mitochondrial DNA, on the other hand, is present in the cellular cytoplasm; it is present, depending on the cell type, in a number ranging from 100 to over a thousand copies per cell; it has a totally circular conformation; the mitochondrial DNA is inherited by the individual solely through the mother (this happens because, at the time of fertilization, meaning the fusion between the paternal sperm cell with that of the maternal oocyte, the sperm nucleus enters the maternal egg cell, while the paternal mitochondria, present in the sperm tail, remain outside the new cell that is formed by fusion).

From the point of view of the peculiarities of analysis, nuclear DNA and mitochondrial DNA are profoundly different.

Due to the nuclear DNA physiological characteristics and particular analytical kits, is able to be typed (or analyzed) in a highly specific way.

In other words, nuclear DNA allows us to identify a particular individual in a unique and univocal way and to discriminate it from all the others present in the human population.

Mithocondrial DNA, due to its physiological process that occurs at the time of fertilization (*briefly illustrated above*), does not allow a specific individual to be typified in a unique and univocal way, but allows to go back only to the maternal family of origin (*what is referred to as "maternal stock of biological origin*") (Fig. 17)

Mitochondrial DNA is the most resistant. Often, it represents the DNA object of forensic investigation if the biological material is particularly compromise.

Due to the characteristics of the conformation of the genetic molecules and their relative number of presence that is weird within each single biological cell, nuclear DNA is much more "fragile" than mitochondrial DNA (Fig. 18).

The textile substrate affects the degradation of DNA

The following table shows the results from a study which analyzed the damage caused to biological/genetic material following a fire. In this study, different biological fluids were compared to evaluate whether the biological cells of different tissues (*in this case blood and sperm cells*) are equally damaged by heat or have a non-homogeneous denaturation.

The study provides a chemically evaluation, while the chemically different substrates with the aim of assessing the composition of the substrate can affect biological/genetic damage.

Genetic markersa widely used, such as CSF and TH01 were used to evaluate the genetic damage; of these two markers, the reciprocal relationship was used for the evaluation of genetic damage (Fig. 19).

The ratio of the selected genetic loci does not significantly change between blood and sperm following high temperatures.

The following histograms (Fig. 20) show the relationship between specific gene loci (or markers, to be understood) called "CSF" and "TH01," specially chosen for

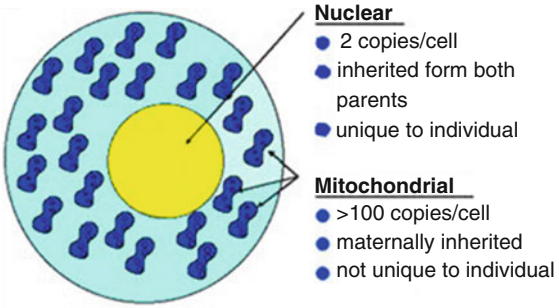


Fig. 17 The biological cell. Relationship between nuclear and mitochondrial DNA

Marcatori "uniparentali" nelle scienze forensi
DNA mitocondriale (mtDNA)

TABLE 14.1 Comparison of Human Nuclear DNA and Mitochondrial DNA Markers.

Characteristics	Nuclear DNA	Mitochondrial DNA (mtDNA)
Size of genome	≈3.2 billion bp	≈16,569 bp
Copies per cell	2 (1 allele from each parent)	Can be >1000
Percent of total DNA	99.75%	0.25% content per cell
Structure	Linear; packaged in chromosomes	Circular
Inherited from	Father and mother	Mother
Chromosomal pairing	Diploid	Haploid
Generational recombination	Yes	No
Replication repair	Yes	No
Unique	Unique to individual (except identical twins)	Not unique to individual (same as maternal relatives)
Mutation rate	Low	At least 5–10 times nuclear DNA
Reference sequence	Described in 2001 by the Human Genome Project	Described in 1981 by Anderson and co-workers

Fig. 18 Comparison of the characteristics of nuclear and mitochondrial DNA

the study of the comparative degradation of genetic material. It is shown that, following damage from fire, biological traces from blood or semen suffer genetic damage in equal measure.

If we associate this result with the fact that we are sometimes in a condition of minimal trace, such as in the case of traces of semen, this information regarding environmental damage is particularly useful because it ensures that the trace undergoes a "controlled" degradation that is not different from what occurs with traces from other biological tissues.

This, from an investigative point of view, strengthens the probability of carrying out successful genetic sampling, even in critical environmental conditions due to fire.

The comparison and discussion of the above data, of both the tabular and histogram examples, confirms that the genetic damage that the cells suffer as a result of heat/fire is "homogeneous," meaning that the biological traces are all denatured at the same way. In this case, it is emphasized that biological traces from different biological tissues (in this case, blood and seminal fluid) have the same degree of degradation calculated through the ratio between the selected genetic markers CSF and TH01.

Again, this data clearly demonstrates that, for what has been studied, cell denaturation and genetic damage are absolutely not "biological-dependent tissue."

Moreover, this data show how much the chemical nature of the substrate on which the biological traces are adhered influences the degradation.

52	Blood	Nylon	1	CSF	33.26	30.79	N/A	32.03	1.24	0.68
				TH01	49.15	45.08	N/A	47.12	2.04	
53	Blood	Nylon	1	CSF	31.77	26.25	N/A	29.01	2.76	0.638
				TH01	48.77	42.19	N/A	45.48	3.29	
54	Blood	Nylon	1	CSF	35.22	29.37	N/A	32.3	2.93	0.665
				TH01	53.91	43.25	N/A	48.58	5.33	
55	Semen	Nylon	1	CSF	156.23	133.95	N/A	145.09	11.14	0.583
				TH01	270.76	227.3	N/A	249.03	21.73	
56	Semen	Nylon	1	CSF	105.38	97.95	N/A	101.67	3.72	0.555
				TH01	183.36	182.93	N/A	183.15	0.22	
57	Semen	Nylon	1	CSF	143.25	134.7	N/A	138.98	4.28	0.49
				TH01	288.57	279.23	N/A	283.9	4.67	
59	Blood	Nylon	1	CSF	15.58	12.23	N/A	13.91	1.67	1.188
				TH01	12.56	10.84	N/A	11.7	0.86	
60	Blood	Nylon	1	CSF	1.51	1.46	N/A	1.49	0.03	2.015
				TH01	0.73	0.74	N/A	0.74	0.01	

Fig. 19 Summary data of biological/genetic damage from fire (CSF; TH01-target)

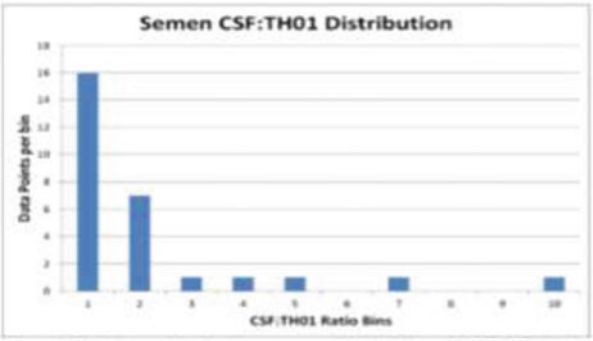


Fig. 20 Relationship between CSF and TH01 gene markers following fire/heat-induced damage in blood and sperm cells

Specifically, it is evident that all the biological traces present on the "nylon" substrate occur in greater denaturation than those present on the "polyester" substrate.

In other words, this shows that the biological/genetic damage is absolutely dependent on the type of chemical composition of the substrate on which the biological traces are adhered (Fig. 21).

Comparison	t	t_{crit}	Sig. Dif.	More Degraded
	Blood vs. Semen			
Blood vs. Semen	0.11	2.021	No	N/A
Donor 1 Blood vs. Semen	0.347	2.131	No	N/A
Donor 2 Blood vs. Semen	0.359	2.069	No	N/A
Nylon Blood vs. Semen	1.497	2.08	No	N/A
Polyester Blood vs. Semen	1.899	2.11	No	N/A
	Nylon vs. Polyester			
Nylon vs. Polyester	3.191	2.06	Yes	Nylon
Blood Nylon vs. Polyester	2.503	2.131	Yes	Nylon
Semen Nylon vs. Polyester	2.411	2.306	Yes	Nylon
Donor 1 Nylon vs. Polyester	1.725	2.201	No	N/A
Donor 2 Nylon vs. Polyester	2.761	2.179	Yes	Nylon
	Donor 1 vs. Donor 2			
Donor 1 vs. Donor 2	1.02	2.026	No	N/A
Blood Donor 1 vs. Donor 2	0.421	2.131	No	N/A
Semen Donor 1 vs. Donor 2	0.712	2.08	No	N/A
Nylon Donor 1 vs. Donor 2	1.429	2.08	No	N/A
Polyester Donor 1 vs. Donor 2	0.069	2.131	No	N/A

Fig. 21 comparison of the denaturation of biological traces placed on different textile substrates

For the sake of completeness, it is important to specify that, when we talk about the evaluation of genetic damage following a fire, the standard to which we refer to in the evaluations is the electropherogram that arises from the typing of DNA through the use of a number of markers between 16 and 24.

Below is an example of an electropherogram over 23 loci (Fig. 22) :

Final Remarks

With the advent of sophisticated technologies, nowadays scientific proof becomes certain proof; this is by virtue of both modern methodologies and the control figures in charge of checking compliance with the rules, procedures, and scientific protocols.

In science, therefore, as well as in the investigation and in the process of investigation, the method is mainly aimed at discovering the error that leads to the truth.

The presence of genetic traces in a fire scene certainly makes it more complex, but the investigation, if well planned, leads to a solid conclusion.

While, in fact, the inquisitorial system proceeded from the claim to already know the whole truth by lowering it from above in the process in the form of already established evidence, the accusatory system takes into consideration the evidence regularly formed (i.e., made less uncertain) in the cross-examination between the

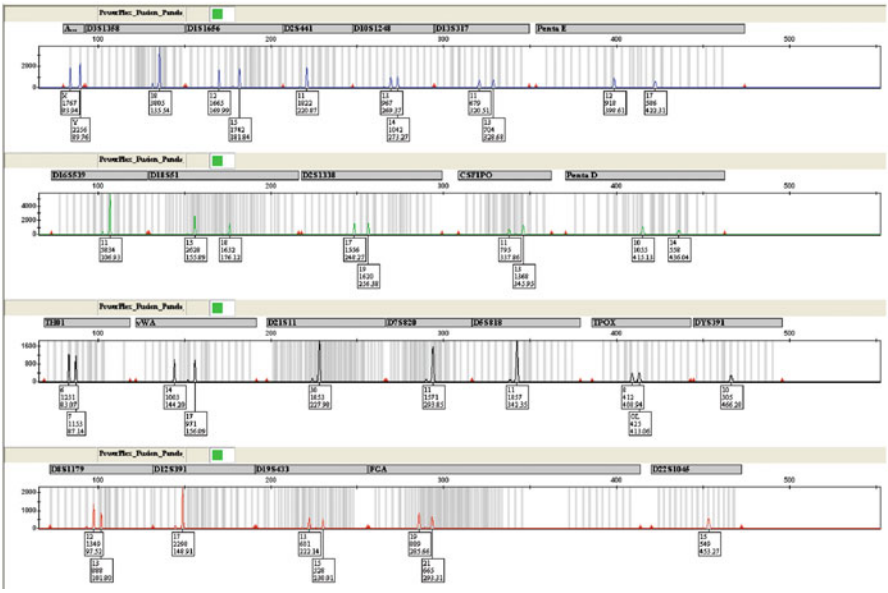


Fig. 22 Example of a genetic electropherogram for forensic purposes

parties before a third judge; therefore the investigation, conducted according to precise rules, constitutes one of the founding ideas of the new criminal trial.

Under these circumstances therefore, there is no investigation that is not naturally included in a system of checks, weights, and counterweights.

The examination of forensic trace evidence in a fire scenario is currently the subject of scientific research and undoubtedly deserves targeted studies for the resolution of numerous judicial cases (Stella 2000).

The New Popperian Epistemology of the Criminal Process: Strong Scientific Evidence and Reduction of the Clues to Conjectures

With the movement for the new renaissance of Justice, we go all over Italy holding conferences on various topics exposed by the avant-gardes to improve the applied law. What are the remedies for a truly fair new justice (My design of a new justice led me to found the MOVEMENT FOR THE NEW RENAISSANCE OF JUSTICE (MOV.RIN.GIU). The early Renaissance was represented by the Enlightenment, which virtually crushed the inhumane righteousness of the Inquisitors. The seeds for a revolution of Themes are still waiting to be realized with our second Renaissance. Even today, the indicative process is underway with the risk of condemning innocent people, subverting Voltaire's maxim: "It is better to run the risk of not condemning a guilty person than of condemning an innocent person." The HUMANIST AND FRATERNAL JURIST is the final target of the movement in order to realize a RIGHT JUSTICE.)?

In this period, some striking judicial cases (e.g., Meredith Kercher, Melania Rea, Elena Ceste, Guerina Piscaglia, Roberta Ragusa, Yara Gambirasio, Sara Scazzi, Chiara Poggi, etc.) have brought to the limelight the suspects who continue, although arrested, to proclaim their innocence. The deficiency of certain evidence and the founding of processes on purely clues have generated on the network and social media opposing groups of innocentists and guilty parties.

The peninsular tour has implemented a series of conferences (in Rome, Caserta, Crotone, Viterbo, Palaia-Agliati, Naples, Milan, Verona, Torre Annunziata), and others are planned. There have been interventions of insiders, emblematic characters (such as Raffaele Sollecito, attorney Piero Tony, lawyer Giuseppe Lipera defender of Contrada), and finally institutions particularly interested in the construction of a fair justice (Basilicata Region, Province of Crotone, Municipalities of Caserta, Palaia, Corsico) (https://www.bbc.com/news/world-europe-21938080). Conferences tend to verify the problems related to clues, preferably having investigators go in search of very strong and crossed proofs, which only establish a fair process to be sure of putting in prison the guilty and not the innocent.

The Revolution of the Justice Method: Popper's Way for Strong Proofs

The first investigation that must be carried out by a fair judge in the search for a procedural truth is that on the method used and on its effectiveness. Here modern epistemology, in particular, the philosophy of Karl Popper, helps us (https://www.simplypsychology.org/Karl-Popper.html).

In science, conjectures based on clues are valid to create a scientific thesis, but this must be submitted to the scientists to experiment in the laboratory. The thesis is valid only if all the scientists reach the same conclusion. Mutatis mutandis also applies to judges. If a conjecture leads to different results on the part of the analyzers, then that conjecture is fallacious or at least it is not known to what extent it is true.

The judge in the analysis of evidence must merge with the traditional criterion of verification, based on the search for data confirming the incriminatory conjecture, the most modern devised by Popper in the epistemology of falsification, i.e., going to research, even beyond the evidence sometimes, facts that could contradict the main statement. "The criterion of falsifiability maintains that an assertion, to be empirically informative, that is to say scientific, must be falsified principally and not denied in fact, despite the most severe attempts to make it fall."

We must abandon the lethalness principle of the "free conviction of the judge." It is necessary, therefore, that the magistracy models a new scientific methodology, avoiding confusion as it has sometimes happened in the past. Only by distinguishing legal science as a conjecture (based on clues) and legal science as a result, based only on strong evidence of proofs, can we have a real guarantee of a criminal justice free from prejudice and truly egalitarian.

Using these principles, as a judge of the Court of Rome, on June 13, 2000, I raised, in vain, the question of unconstitutionality of the process based on the clues, but the Constitutional Court with Ordinance no. 302 of 2001 rejected my request in a

brusque way. A noted journalist, Gigi Trilemma, wrote in his article "The Constitutional Court has lost an opportunity to abandon permanently the literary processes and give definitive space to the scientific process based on certain evidence and not on clues. I am sorry the hasty system with which the Constitutional Court has solved the epistemological question, avoiding to tackle the crucial matter about the so called war on the proof versus the clues. The criminal judge, on the other hand, demanded just to do this, that is to decide not with the tautological criteria of legal formalism but based on the principles of modern epistemology, which can only define what is certain and what is false in any proceedings to collect evidence on facts."

Trials are made for strong proofs not for clues that only serve to create conjectures, invalidated if no evidence is found. This is the Popperian scientific process, not a mediaeval novel. The clues only serve to open investigative tracks, but then if there is no strong evidence, the process falls. A thousand clues do not form a single proof not like 1000 rabbits which form a warren and certainly not a lion!

Discovering the authors of the crimes is anything but simple. Detective stories say that no crime is perfect. Indeed perfect crime does exist! A big number! And justice enjoys finding culprits at all costs to show that it works.

To limit the judicial freedom of the judges in a scientifically way, together with the late professor Imposimato, we came up with a list of legal evidence to be followed. In this regard, the judges must demand not only confession and/or smoking pistol but also unequivocal telephone tapping, crisscrossed testimonies, reconstructed paths with CCTV cameras, post delictum markings with bugs, applications antistalking as Mytutela, and scientific surveys done properly and 100% safe. Certainly not as in the cases of Cogne, Melania Rea, Meredith, Bossetti, not to mention the case of Ceste where you do not even know how the woman died, or Guerina Piscaglia and Roberta Ragusa whose bodies were not even found not knowing if they died or not, if they were killed, and how and by whom. If you do not go through strong proofs, all you can do is trigger indictment trials against alleged perpetrators, keeping them out of jail anyway. If then the clues do not lead to proofs, this serious, precise, and concordant process has failed.

Now trial based on clues is required by law, but it is irrational because in itself it always creates reasonable doubt so much so that these striking cases create the faction of the guilty and that of the innocentists, thus lacking upstream certainly of the final verdict. We continue to fight to make the declaration process unconstitutional. Also because against the expression of the norm, what was supposed to be an exceptional process has become the rule by putting the weaker subject in jail and setting him up as a scapegoat. According to statistics, 90% of the processes today on a clue basis would be wiped out, remaining only 10% of processes to be carried out until the possible sentence. A quick but right way to dispose of the backlog.

The Neutrality of the Researcher and the Rigorous Acquisition of the Proofs

After the examination of judicial science as a method, we pass to that of science *strictu* sensu, asking ourselves the question of who and how to collect the traces of a crime and examine them.

We use DNA as a model of study, which is even considered in some striking processes as a proof, being, instead, a simple clue. Both the facts of the media and the positions of various "insiders" show us that the genetic test is not infallible as is believed. With the rigorous Popperian method, the first notation is that DNA sampling and analysis must be guaranteed by the creation *de iure condendo* of a national sampling service and investigations with super partes experts, depending on the magistracy (we believe to exhume the investigating judge) and not of the prosecutor. At the time, those delicate acts of investigation must be guaranteed by the presence of a defense counselor also for the unknown murderer; otherwise all is nill (article 111 of the Constitution). It is necessary to provide a legal defender and a legal advisor for the unknown to avoid the formal flaw of the control and verification procedure. It is not pure theory given the problems created by the scientific police in the Meredith Kercher case, which ended with the acquittal of Amanda Knox and Raffaele Sollecito.

Besides the criterion of detector neutrality, it is necessary to guarantee a supervisor in the key stages of the collection of exhibits, the correct chain of custody, the laboratory analysis to ensure the right assumption (procedures, instruments, etc.), and conservation and analysis of data.

The currently dominant static criminology is Aristotelian and apodictic and is aligned with the clue process. The dogmatic omnipotence of DNA is part of it.

Dynamic criminology, on the other hand, requires a rigorous answer to the questions: "Quis quid ubi quibus auxiliis cur quomodo." This is of course a Latin phrase, which literally means "who, what, where, by what means, why, how, when?" It is a hexameter elaborated by Cicero (quoted by St. Thomas Aquinas) which contains the criteria to be observed in the conduct of a literary composition: to consider the person acting (quis); the action, what he does (quid); the place where it happens (ubi); the means that he uses in executing it (quibus auxiliis); the purpose it has (cur); the way it is done (quomodo); and the time it takes him to execute it (when) (https://crimsonpublishers.com/fsar/pdf/FSAR.000521.pdf).

So, we use the brocardo with the addition of the "quantum" to implement the reconstructive sequence of a crime in a criminal key.

Therefore, we have built a complete sequence of a crime in terms of dynamic criminology and strict response to every single question in verification and falsification of data according to the teachings of Popper.

The scheme described above in the Bossetti case is admitted and not granted that the DNA is his (https://www.theguardian.com/world/2016/jul/02/yara-gamirasio-murder-massimo-bossetti-dna-evidence-italy-guilty-verdict). This element is not

enough to attribute the crime to him. It is necessary to establish precisely "how," but it must be considered that with the possible homicidal action, it cannot rule out accidental or artful contamination. It is possible that the suspect has left traces not because he is the murderer himself but because he has touched the corpse post delictum accidentally or concealed the dead body.

Before wrapping up, the fair trials are done by science and strong proofs, certainly not by fictional clues.

References

Al Aboud DM, Gossman W (2020) Wood's lamp. StatPearls

Augenti N, Chiaia BM, Flaccovio D (2011) Ingegneria Forense

Baranowska I (2016) Handbook of trace analysis. Springer, Cham

Barni F, Lewis SW, Berti A (2007) Forensic application of the luminol reaction as a presumptive test for latent blood detection. Pubmed

Beckman JS, Weber JL (1992) Survey of human and rat microsatellites. Genomics 12:627–631

Bieber F.R, Buckleton J.S, Budowle B.,. Butler J.M,. Coble M.D, (2016) Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion, BMC Genet 17:125

Biondo R, Spinella A, Montagna P, Walsh PS, Holt S, Budowle B (2001) Regional allele frequencies at nine short tandem repeat loci. Forensic Sci Int 115:95–98

Buckleton J (2009) Validation issues around DNA typing of low level DNA. Forensic Sci Int Genet 3:255–260

Buckleton JS, Bright JA, Gittelson S, Moretti TR, Onorato AJ, Bieber FR et al (2019) The probabilistic genotyping software STR mix: utility and evidence for its validity. J Forensic Sci 64:393–405

Chaitanya L, Breslin K, Zuniga S, Wirken L, Pospiech E, Kukla-Bartoszek M et al (2018) The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation. Forensic Sci Int Genetics 35:123–135

Cowen S, Debenham P, Dixon A, Kutranov S, Thomson J, Way K (2011) An investigation of the robustness of the consensus method of interpreting low-template DNA profiles. Forensic Sci Int Genet 5:400–406

D'Orio E, Mogenesen HS, Larsen HJ, Morling N (2018) Detection of biological stains using the Lumatec Superlite 400 Alternative Light Source. HIDS conference

D'Orio E, Mangione M, Montagna P (2020) Evaluation of the performance of the blood evidence tests in a fire scene. HIDS conference

Fowler JC, Burgoyne LA, Scott AC, Harding HWJ (1988) Repetitive deoxyribonucleic acid (DNA) and human genome variation-a concise review relevant to forensic biology. J For Sci 33:1111–1126

Francione G (2019) Criminologia Dinamica. La via di Popper al DNA. Nuova Editrice Universitaria

Gardner RM, Krouskup DR (2018) Crime scene processing and investigation, 3rd edn. CRC Press, Boca Raton

Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints'. Nature 318:577–579

Gill P, Whitaker J, Flaxman C et al (2000) An investigation of the rigor interpretation of rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 112:17–40

Gill P, Kirkham A, Curran J (2007) LoComatioN a software tool for the analysis of low copy number DNA profile. Forensic Sci Int 2:128–138

Graham EA (2008) DNA reviews: low level DNA profiling. Forensic Sci Med Pathol 4:129–131

Hearne CM, Ghosh S, Todd JA (1992) Microsatellites for linkage analysis of genetic traits. Trends Genet 8:288–294

Kayser M (2015) Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genetics 18:33–48

Mangione M (2017) Dalla progettazione antincendio all'investigazione sugli incendi. Rivista Antincendio

Mangione M (2018) *Investigazione su una scena d'incendio: aspetti forensi*. EPC Rivista Antincendio

Mangione M, Bontempi F, Crosti C (2015) Atti del convegno IF CRASC'15. Structural Fire Investigation e Ingegneria Forense. Università La Sapienza, Roma

McCord B, Lee SB (2018) Novel applications of massively parallel sequencing (MPS) in forensic analysis. Electrophoresis 39(21):2639–2641

Mullis K (1987): US patent 4,683,202 process for amplifying nucleic acid sequences

NDAA Position Statement on Rapid DNA, January 30, 2018. https://dps.alaska.gov/getmedia/fb933229-8e52-4cf8-8fe0-cb72d5e039e3/NDAA%20Statement-on-Use-of-Rapid-DNA-Technology.pdf

Norsworthy S, Lun DS, Grgicak CM (2018) Determining the number of contributors to DNA mixtures in the low-template regime: exploring the impacts of sampling and detection effects. Legal Med 32:1–8

Ottaviani E, Vernarecci S, Asili P, Agostino A, Montagna P (2014) Preliminary assessment of the prototype Yfiler® plus kit in a population study of Northern Italian males. Int J Legal Med 9:1078–1079

Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–1724

Spinella A, Marsala P, Biondo R, Montagna P (1997) Italian population allele and genotype frequencies for the AmpliType PM and HLA-DQ-alpha loci. J For Sci 42:514–518

Stella F (2000) Leggi scientifiche e spiegazione causale nel diritto penale. Giuffrè Editore

Swaminathan H, Qureshi MO, Grgicak CM, Duffy K, Lun DS (2018) Four model variants within a continuous forensic DNA mixture interpretation framework: effects on evidential inference and reporting. PLoS One 13

SWGDAM position statement on rapid DNA, October 23, 2017. https://docs.wixstatic.com/ugd/4344b0_f84df0465a2243218757fac1a1ccffea.pdf

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using polymerase chain reaction. Am J Hum Genet 44:388–396

Welch LA, Gill P, Phillips C, Ansell R, Morling N, Parson W, Palo JU, Bastisch I (2012) European network of forensic science institutes (ENFSI): evaluation of new commercial STR multiplexes that include the European standard set (ESS) of markers. Forensic Sci Int Genet 6:819–826

Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFISTR SGM plus multiplex system for both standard and low copy number (LCN) DNA analysis. Forensic Sci Int 123:215–223

Tools and Techniques Used in Forensic DNA Typing

6

Akanksha Behl, Amarnath Mishra, and Indresh Kumar Mishra

Contents

Abstract

The chapter starts with a basic overview of the history of DNA and its use in human identification. Then, it discusses the fundamental information about DNA, its structure, and its function. Further sections of this chapter explain the various techniques used for forensic DNA analysis. It covers the processes involved in the preparation of the samples for DNA amplification through the polymerase chain reaction or PCR methods. Separate sections are dedicated to the explanation of commonly used STR markers, Y chromosome markers for specifically

A. Behl · A. Mishra (✉)
Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, Uttar Pradesh, India
e-mail: amishra5@amity.edu; amishra5@amity.edu

I. K. Mishra
Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, Uttar Pradesh, India

Forensic Science Laboratory, Rohini, Government of NCT, Delhi, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_4

identifying the male contributor of a sample, and mitochondrial DNA, which is maternally inherited and utilized in cases where highly degraded DNA is found. Finally, in the last section, the topic of nonhuman DNA is touched, which describes how the "other" DNA can aid forensic investigations. This DNA examination constitutes the investigation of animal, plant, and microbial DNA testing. The technology portion of the chapter includes the description of the separation of DNA molecules using slab gel and capillary electrophoresis. Fluorescent detection methods are generally used. It also mentions the use of digital DNA databases to solve crimes and identify suspects. It likewise explores issues that are one of a kind to the scientific DNA analysis such as sample mixtures, in particular combinations, debased DNA tests, PCR restraint, and tainted, all of which sway criminological casework since numerous examples don't come from a sterile, controlled climate. The national DNA information bases that consist of the genetic data will profit the law requirement by connecting various crimes.

Keywords

Forensic DNA typing · Polymerase chain reaction · Restriction fragment length polymorphism · Forensic protein profiling · STR markers

Introduction

DNA typing is also known as DNA fingerprinting or DNA profiling. It is a widely used technique in forensic DNA analysis. It was first introduced during the 1980s; initially, highly polymorphic regions were discovered in human DNA by Wyman and White. This particular discovery has changed forensic science forever and empowered law enforcement services to match criminals with crime scenes. It became the foundation for the study of banding patterns which was specific to an individual, and it was studied after restriction fragment length polymorphism or RFLP analysis of repeated DNA sequences This study was conducted by Professor Sir Alec Jeffrey at the University of Leicester. He observed that certain regions of DNA contained DNA sequences that were repeated over and over again next to each other. Along with that, he also observed that the number of repeated sequences present in a sample could differ among different individuals. These particular "repeated" regions are called as variable number of tandem repeats or VNTR. restriction fragment length polymorphism (RFLP) technique was used by Dr. Jeffreys to examine the VNTRs.

This technique employs the use of a restriction enzyme, which is used to cut the specific regions of DNA surrounding the VNTRs. In the RFLP technique, at first, the DNA is extracted and then purified. RFLP technique consists of dividing the DNA strands into wanted lengths utilizing limitation chemicals, isolating the DNA sections through the gel medium as per their sizes (atomic loads), moving the parts on to the strong backings (nylon or cellulose film) hybridizing the particular DNA parts

with reciprocal DNA pieces (tests labeled with radioactive or nonradioactive material) and taking the pictures or impressions of the ideal sections. RFLP methods were first used for DNA typing, after that they were subsequently replaced by PCR methods (Jamieson and Bader 2016).

Basic Principles of DNA

DNA is a deoxyribose nucleic acids are composed of nucleotide units. DNA is a large, polymeric molecule. These units consist of three parts: a nucleotide base, a sugar, and a phosphate. The nucleobase is responsible for the variation among each nucleotide unit; the phosphate (PO_4) and sugar units, meanwhile, form the backbone structure of a DNA molecule. It consists of four nucleobases: A (adenine), T (thymine), C (cytosine), and G (guanine). The structure is of a DNA double helix. The sides of the ladder are a linked chain of 5-carbon sugars and phosphate groups. The rungs connected to the 5-carbon sugars are known as bases. In all living beings, numerous combinations of these four bases lead to diversity in biological characteristics. Humans have approximately 3 billion nucleotide positions in their genomic DNA. Thus, with four possibilities (A, T, C, or G) at each position, innumerable combinations are possible. In a cell, DNA (deoxyribonucleic acid) is within the nucleus of chromosomes. Gene is the basic hereditary unit and it's defined as a segment of DNA molecule of a chromosome. Genes are responsible for the physical characteristics of an organism. All inherited characteristics are dependent on one or several genes (Fig. 1).

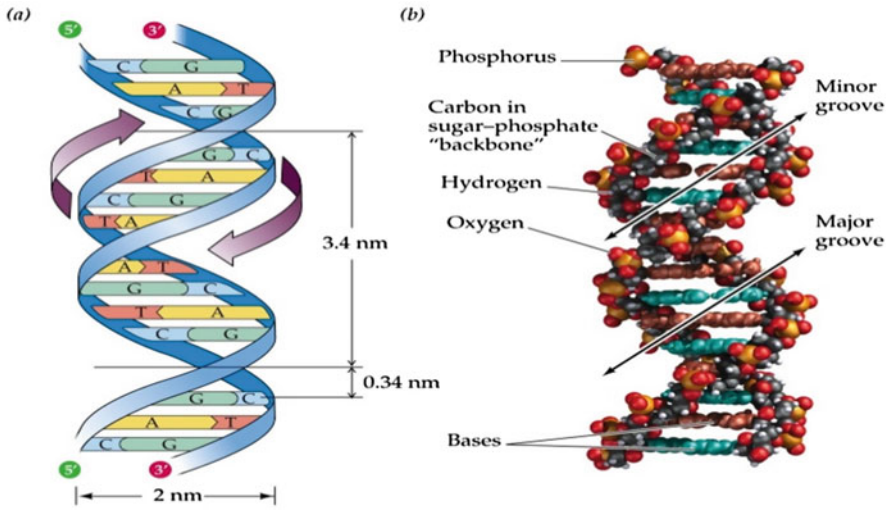


Fig. 1 DNA structure

DNA Typing Methods

A variety of techniques have been discovered and employed for the forensic DNA typing such as single-locus probe and multi-locus probe restriction fragment length polymorphism (RFLP) methods and polymerase chain reaction (PCR)-based assays. Numerous advances have been made in the last quarter of a century in terms of sample processing speed and sensitivity. Various advances have been made in the last quarter of a century as far as to test preparing pace and affectability. Rather than requiring enormous bloodstains with very much protected DNA, minuscule measures of the sample, as meager as a couple of cells in some criminological cases, can yield a helpful DNA profile (Alamoudi et al. 2018) (Fig. 2).

Blood Group Testing

Karl Landsteiner was a researcher at the University of Vienna, Austria. In 1900, he discovered that blood agglutination, which happens when blood belonging to different people, was mixed. He identified four blood types: O, A, B, and AB. According to his research, type O was noticed 43%, type A 42%, type B 12%, and type AB 3%. For the blood transfusion process to work, the donor and recipient require compatible ABO blood types. Otherwise, transfusion reactions will occur, such as agglutination of incompatible cells, and it can also cause death.

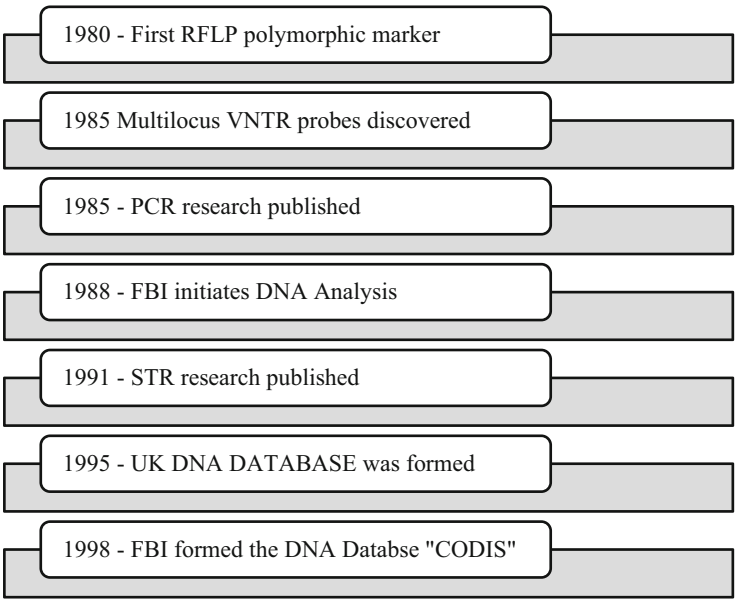


Fig. 2 Short bullet points depicting the history of DNA typing

Table 1 Blood types, antigens, and antibodies

Blood groups :	Blood Group A	Blood Group B	Blood Group AB	Blood Group O
Blood types :	A	B	AB	O
Antibodies :	Anti-B	Anti-A	None	Anti-A, Anti-B
Antigens :	A-Antigen	B-Antigen	A and B Antigen	None

Blood groups are identified with antigen polymorphisms present on the outside of red platelets. These antigens may be protein, carbohydrate, glycoprotein, or glycolipid differences that exist between people. The antigens are acquired from the parents of a person and thus can be utilized to check the paternity. Antibody-based blood tests are employed for the detection of various blood group antigenic alleles (Butler 2010) (Table 1).

Forensic Protein Profiling

In forensic science presently, data about the potential biological origin of forensic samples is generally ascertained by utilizing protein-based possible testing. As of late, mRNA-profiling or messenger RNA profiling has arisen as a system to analyze the natural origin. The improvement of a solitary multiplex mRNA-based framework is explained for the separation of the most well-known serological body liquids and skin cells. A DNA/RNA co-segregation convention was set up that outcomes in DNA yields comparable to our norm in-house approved DNA extraction system which utilizes silica-based sections. An endpoint RT-PCR test was built up that at the same time duplicates 19 (m)RNA markers. In forensic biology and genetics, mRNAs have progressively obtained fame with respect to their capabilities to recognize human body liquids and other forensically important tissues. Alternative strategies for cell-profiling incorporate miRNAs that are tissue-specific, DNA methylation, and

microbial marker. miRNAs are known as 20–24 nucleotides which are small in size, and they are regulatory in nature. They are unequivocally connected with the class of proteins called Argonautes, which makes them entirely steady and beneficial when debased forensic samples are considered. Likewise, epigenetic DNA methylation markers have been introduced that can be separated between some tissue types. Both miRNA and DNA methylation markers appear to be encouraging, yet in their early stages with respect to the occurrence, more markers are expected to segregate the measurable forensic range of body liquids. The utilization of microbial markers has been proposed for the identification of the vaginal mucosa. Notwithstanding, it isn't yet settled whether similar organisms likewise happen on skin surfaces that are in the nearness of, or contact with the vaginal microbial vegetation, for example, skin surfaces of the hands, crotch, or penis. Thus, we see tissue-explicit mRNA examination as the most adaptable cell-profiling approach.

mRNA-profiling has developed from a singleplex PCR procedure to a multiplex RT-PCR stage, giving articulation of information on numerous qualities all the while. mRNA-profiling is promptly joined with DNA-genotyping since RNA and DNA can be gotten from precisely the same examples. The diverse multiplexes that have been created incorporate markers for venous blood, salivation, semen, vaginal epithelia, and feminine emission, and their determination was essentially founded on the capacity portrayed in writing or the tissue-explicit articulation as detailed in articulation information bases. Devoted entire genome articulation exhibit investigation in examples from forensically significant body liquids that were put away for different time stretches has recently appeared to convey stable mRNA markers valuable for legal tissue distinguishing proof. Skin is an extra forensically significant cell type. As of late three mRNA records (LOR, CDSN, and KRT9) were accounted for to show high articulation in skin tests comparative with other forensically pertinent cell types. The expansion of skin markers to an RNA-based cell-typing multiplex would expand the scientific estimation of the test for two reasons: (1) a more complete view on all cell types present in an evidentiary follow is set up which is significant in light of the fact that skin cells are required to happen in numerous wrongdoing scene tests, and (2) a sign for the presence of contact DNA can be gotten. The regularly utilized strategy to show the presence of contact follows is through dactyloscopy unique mark examination, yet additionally other microscopical and immunohistological strategies have been depicted which can recognize skin cells. Unique mark representation techniques don't make a difference to a wide range of substrates, and some can effectively affect the nucleic acids in the skin cells, while others can present pollution. To effectively and unbiasedly evaluate the organic starting point of a measurable evidentiary follow, a solitary profiling examination was built up that tests both the forensically pertinent body liquids and skin cells (Butler et al. 2001).

The amino acid sequences of some proteins vary from person to person. The utilization of multi-protein polymorphisms can lead to some differentiation when there is a coincidence probability of two unattached persons one of the hundreds of hundreds. Before DNA testing became available, protein profiling was the basic technique that was performed in forensic laboratories as a raw way to differentiate

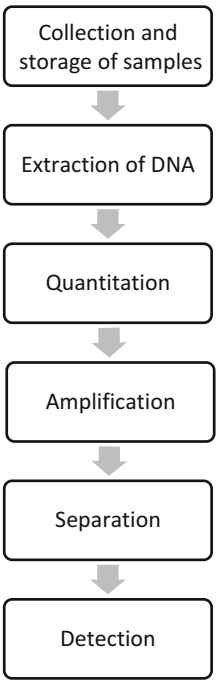


Fig. 3 Overview of basic steps involved in DNA Analysis

between the samples (Q and K), and the serum contains several isozymes. These isozymes are the different types of a protein enzyme that can catalyze the same biochemical reaction even though it has slightly different amino acids acid sequences. However, there are usually only two or three forms of each isoenzyme, which make them quite weak at distinguishing humans. Electrophoresis of starch gel, agarose gel, and polyacrylamide gel divide these proteins into differential alleles. In the early 1980s, many laboratories have started isoelectric-focused polyacrylamide gel electrophoresis (IEF), which has a higher resolution than protein electrophoresis because it creates smarter stripes (Fig. 3).

RFLP-Based DNA Testing

When the RFLP process was first introduced, it was lengthy and took several weeks to complete. For DNA extraction, a biological sample was collected such as blood. The extraction of the DNA was done from the cells by disintegrating the cell membranes and removing the protein layers around DNA. To slice the to cut the long, extracted DNA molecules into smaller pieces, a restriction enzyme was utilized. These enzymes possess the capability of discovering and cleaving specific DNA sequences. In the next step, agarose gel is used to separate e these DNA fragments. These fragments are put into the deep areas called, “wells” in a gel that

floats in a buffer solution. This solution is placed in a chamber between two electrodes, and an electric current is passed through it, which causes the movement of negatively charged fragments toward the positive end. DNA fragments of smaller size (shorter fragments) move faster than the bigger ones. This is how the separation of DNA fragments is done into bands, on the basis of their size. Southern blotting was done after the separation of these fragments, on the basis of their size. In this technique, the DNA fragments that were separated before were transferred to a nylon membrane, after bringing the membrane in contact with the gel. These DNA strands were turned into single-stranded strands by the action of alkaline solution. UV light is then fixed onto the membrane of one of the strands by cross-linking the DNA onto the membrane with UV light. This "UV light" is also called a radioactive or chemiluminescent probe. These probes consist of a VNTR sequence, which hybridizes the DNA attached to the nylon membrane (Butler 2015a).

Hard binding occurred at the appropriate hybridization temperature and ionic strength, allowing the classified probe to detect its complementary set with complete accuracy. Strong binding occurred at acceptable hybridization temperatures and ionic strength, allowing the labeled probe to search for its complementary sequence with complete accuracy. The additional probe is then removed after repeated washing of the membrane. In the last step, the position of the probe was determined by inserting a membrane-bound to an X-ray film. Relative molecular analysis in the human genome and sets of sequences of bases are repeated varied times. Such repeated sequences are known as minisatellites. Minisatellites show a really high degree of cistron variation within the range of repeat units and consequently in their length. Hence, they type the idea of differentiating or distinctive people supported this length polymorphism at the molecular level by mistreatment either multi-locus or single-locus probes.

The use of just one multi-locus probe will offer adequate numbers of variable bands that establish the positive identification of an individual. Thus, it's one powerful take a look at positive matching of body tissues and determination of parentage. The chance of getting identical patterns of bands from the deoxyribonucleic acid of two people mistreatment multi-locus probe is of the order of one in 1014 to 1030 that is over 5X109, the overall world population. Thus, the DNA prints obtained from multi-locus probes are extremely unique. Single-locus tests or a combination of at least two tests are utilized for distinguishing variety at a particular minisatellite locus. A solitary locus test uncovers an example of up to two bands though a combination of a few tests uncovers multiple bands. These tests are profoundly delicate and thus can be utilized for little and even incompletely debased examples of DNA. These are helpful in distinguishing proof of blended examples as in instances of various assaults.

PCR-Based Tests

PCR or polymerase chain reaction is an enzymatic process in which a specific piece of DNA is simulated at several times to obtain multiple copies of the specified

sequence. PCR was first described by Kary Mullis in 1985. PCR is a molecular process, and it consists of "photocopying" which refers to heating and cooling of the sample to the exact thermal cycle pattern for ~30 cycles. One copy of the target DNA sequence is generated for each cycle, and the target sequence contains the molecule. It proved to be more suitable than the previously used MLP and SLP systems. There was a time when SLP and PCR techniques were used in parallel to identify a person. PCR technology has been used to enter most of the VNTR region for forensic purposes. Several polymorphic loci were used for analysis, such as D1S80 and HLADQ (Butler 2015) (Fig. 4).

DNA process technology took a dramatic modification in the early 1990s. With the arrival of a new technique the polymerase chain reaction (PCR), the tiny amounts of template DNA fragment got amplified. The strategy was quick, dependable, and was proficient to work even with deteriorated examples In the PCR-based procedure the time needed to investigate an example was essential. For the recognition of various molecules from one another, separation needs to be performed so as to pull the fragments on the basis of their sizes. Electrophoresis is the most commonly used separation method, and it is conducted in a gel or capillary manner.

The word "electrophoresis" originates from the Latin word "phore" which means bearer and the Greek word electron which means charge. In this manner, the cycle of electrophoresis alludes to electrical charges (positive and negative ions) transported by the molecules. On account of In the DNA structure, the phosphate groups are basically the backbone and consist of the negative charge. In nucleic acids, the phosphate groups donate H+ ions, and it turns them into negatively charged acid in buffers. When an electric field is applied to it, these DNA molecules will move away from the negatively charged electrode (a cathode), and it will move toward a

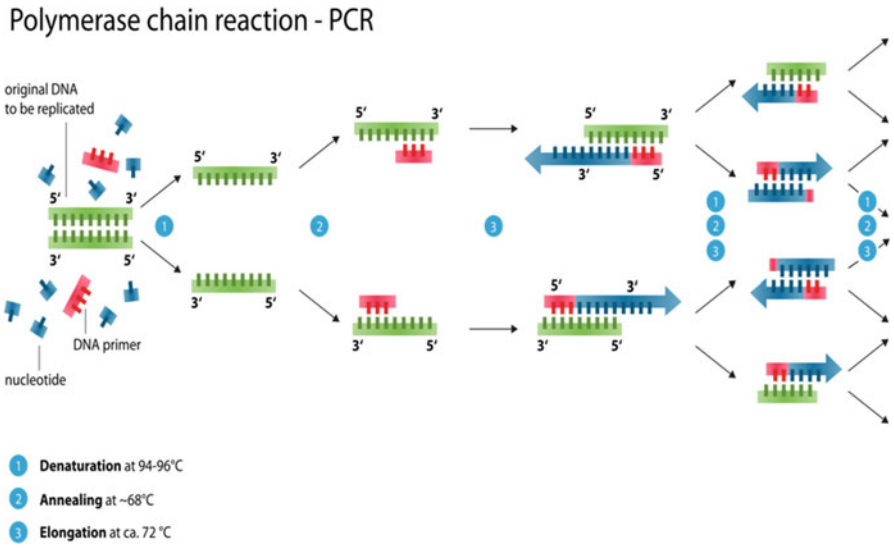


Fig. 4 Steps in PCR process

positively charged electrode (an anode). The speed of movement of DNA fragments depends upon the voltage. If it's high, then the fragments will move faster as they will feel more force. The migration of particles in an electric field leads to the production of heat. This heat that is generated must be dispersed or it will be consumed by the system. Extreme heat can cause a gel to soften and self-destruct. Two kinds of gels are regularly utilized in molecular and forensic DNA analysis for the purpose of separation of DNA. Agarose gels are used as they have generally huge pore sizes and are utilized for isolating bigger DNA particles, while polyacrylamide gels are preferred to be utilized to acquire high separation of smaller DNA particles, generally under 500 or 1000 bp. In Forensic DNA typing, usually techniques the two kinds of gels are utilized. Restriction fragment polymorphism (RFLP) techniques use agarose gels to isolate DNA pieces going in size from ~600 bp to ~23,000 bp. Low-atomic-weight DNA particles are not all well isolated with agarose gels. Then again, PCR-amplified STR alleles, which range in size from ~100 bp to ~400 bp, are better served by polyacrylamide gels. On account of some STR loci that contain micro variants, the high-goal capacity of polyacrylamide gels is fundamental for isolating firmly measured DNA atoms that may just contrast by a solitary nucleotide (Carracedo Angel 2005).

In multiplex PCR process, more than two primer units designed for amplification of different objectives are protected inside the equal PCR reaction. Using this approach, more than one goal sequence in a medical specimen can be amplified in a single tube. As an extension to the realistic use of PCR, this technique can save effort and time. The primers used in multiplex reactions ought to be decided on cautiously to have similar annealing temperatures and should be no longer complementary to every different. The amplicon sizes must be able to form bands that can be seen through gel electrophoresis. Multiplex PCR can be employed in two different ways, one being a single-template PCR response that uses several sets of primers to make bigger specific areas within a template, and the second one multiple-template PCR response, which uses more than one templates and several primer sets inside the identical reaction tube. Although the usage of multiplex PCR can lessen charges and time to concurrently hit upon two, three, or more pathogens in a specimen, multiplex PCR is greater complex to expand and frequently is much less sensitive than single-primer-set PCR. The benefit of multiplex PCR is that a set of primers can be used so that false positives or negatives should not be there. Multiple regions can be copied through the polymerase chain reaction as it lets in concurrently by truly adding one or more primer set to the response aggregate. This simultaneous amplification of DNA and its units is generally referred to as multiplexing or multiplex PCR. The primers should be similar so that a multiplex response should work properly. The primer-annealing temperatures should be similar, and excessive areas of complementarity must be avoided so that the primers bind to each other and not to the template DNA. Any addition of a new primer in a multiplex PCR reaction leads to an increase in the complexity of feasible primer interactions (Butler et al. 2001). Practically, multiplex PCR optimization is better than singleplex reactions due to the fact such a lot of primer-annealing events should occur without interference with each other. For the purpose of achieving a balance

among the amplicons, various loci should be amplified. Primer sequences and concentrations alongside magnesium concentrations are normally the maximum essential to multiplex PCR. Obtaining successful co-amplification with nicely-balanced PCR product helps in identification (Cavanaugh and Bathrick 2018) (Table 2).

Capillary Electrophoresis

Capillaries used for separation can generate results quickly in minutes as opposed to hours. This is because of the fact that higher voltages are allowed with better dissipation of the heat from the vessels. One of the major advantages of CE is that the quantitative data is promptly accessible in an electronic configuration following the finish of a run. No additional means such as filtering the gel or snapping a photo of it are required. Path following isn't important since the example is contained inside the narrow, nor is there dread of the traverse from contiguous wells with CE. The one significant disservice of CE instruments is throughput. Because of the reality that examples are investigated successively each in turn, single fine instruments are not effectively fit for preparing high quantities of tests or test throughputs.

Capillary electrophoresis (CE) is one of the most prominent improvements in the measurable DNA profiling work process, maybe second just to the creation of the polymerase chain response (PCR). The event of locales containing rehash successions inside the DNA particle and their polymorphic nature was found in the mid. In the 1990s, highly polymorphic (STR) markers supplanted VNTRs. At first, intensified STR parts were isolated utilizing piece gel electrophoresis and distinguished by silver recoloring. Creations in fluorescent colors, fluorescence location, and multiplex PCR made ready to the as of now utilized STR composing conventions. A CCT trio framework was effectively evolved in the mid-1990s for synchronous enhancement of CSF1PO, TPOX, and THO1 loci. Multiplexing permitted concurrent partition and location of numerous intensified pieces in a single path and now in one capillary (Sanger et al. 1977).

The main features of CE are its adaptability; inorganic particles, natural atoms, and macromolecules can be isolated on a similar instrument – and much of the time a

Table 2 Depiction of contrasts between the two main DNA typing methods, i.e., RFLP and PCR methods

Properties and conditions	PCR method	RFLP method
DNA sample form	Double or single-stranded DNA can work	Only double stranded DNA will work
DNA amount required	0.1–1 ng	50–500 ng
Analysis time taken	1–2 days	Generally longer, about 1 week or 6 weeks (depending on different probes)
Sample mixtures	Can be analyzed	Can be analyzed
DNA condition	Degraded DNA can be analyzed	Only intact DNA can be analyzed

similar slender – while changing just the piece of the running support and detachment medium. Consequently, CE is the most generally utilized diagnostic technique supplanting gel electrophoresis, elite fluid chromatography (HPLC), gas chromatography, and other partition strategies. CE has a very high settling ability because of its fitting stream and insignificant dissemination. CE offers various focal points over section gel electrophoresis in scientific DNA investigation applications.

Automated software for the purpose of data collection and its handling for a mechanized assortment of information is accessible. Real-time information can be conceivable as well as quantitative data can be inferred. Separation is always high-resolution. Results are exceptionally reproducible and quite precise with this technology. In this method, only a little segment of an example is utilized, which can be, so it tends to be retested/reinjected as well, if necessary. And deleterious end waste products are much lesser. It is essential to take note that quantitative contrasts seen in the CE information may not actually relate to the quantitative contrasts in the first examples. Particular amplification may influence the extents of enhanced DNA identified on CE, explicitly with respect to the minor supporter extent. The CE results include height and peak which are used for improvement of multiplex PCR responses (Giltay and Maiburg 2010).

In theory, about a billion copies of the target areas on the DNA template are generated after 30 cycles. This PCR product is referred to as "amplicon" and therefore insufficient quantity which can be easily measured by various methods. PCR is usually performed with a sample volume of 5 to 100 µl. With such small volumes, evaporation can be problematic and result in precise dosing components that react can become problematic. There was a period when both SLP- and PCR-based techniques were utilized to distinguish an individual. PCR innovation was used to type numerous VNTRs locale for measurable purposes. PCR innovation was utilized to kind a few VNTRs district for logical purposes. Numerous polymorphic loci like D1S80 and HLADQ_ were utilized for investigation.

The PCR reaction is prepared by combining various specific components. Purified water was added to it to obtain the desired volume and concentration of all ingredients. Commercial kits containing premixed ingredients are also utilized for PCR. Such kits are easy to use for forensic purposes. The two most important components of a PCR reaction are short DNA sequences that precede or "spin" the copied region. The primer is used to identify or "target" a portion of the DNA template used. It is a chemically synthesized oligonucleotide concentration versus DNA template for PCR induction.

Some VNTR loci have moderately short size alleles, and they can be PCR-intensified. Locus D1S80 was utilized in a criminological arrangement for the Amp-FLP investigation. In this locus, the fragments are in the scope of 14–42 recurrent units (16 bp per unit). The enhanced sections were isolated by size utilizing polyacrylamide gel electrophoresis and distinguished utilizing silver stain. The discrete alleles are contrasted in a direct manner. The Amp-FLP procedure requires less DNA than the RFLP strategy and functioned admirably for corrupted examples. This locus can be investigated in multiplex style with an amelogenin locus for sex determination. Short tandem repeat (STR) a subclass of VNTR is a locale of human

DNA containing 2–7 base sets (bps) rehash unit. These STRs are simple tandem repeats, or they are also referred to as microsatellites. A total of 105 STRs exist in the human genome out of which some STRs have been described explicitly for scientific DNA profiling. Priorly the STR profiling was completed utilizing polyacrylamide gel electrophoresis of the PCR item and recoloring the gel with silver stain.

Economically accessible nylon strips to which DNA tests are appended are treated with PCR items. Consistently, longer expansion times can increment the yield of longer PCR items since less items are synthesized in long range PCR. The preliminaries (test) utilized in the response are named with an organic tag called biotin. On responding with a protein streptavidin/horseradish peroxidase (HRP) and tetramethylbenzidine (TMB) with hydrogen peroxide, blue shading spots show up at the site of an official of DNA succession with the test. These dot patterns correspond to alleles of the sample. The outcomes are recorded photographically.

PCR amplification is done, either of the whole sample or a part of it with different primer sets. Residual dNTPs and primers are removed from PCR through turn filtration or enzymatic assimilation. After that PCR amount is calculated. DNA sequencing is executed in response to consolidate fluorescent ddNTPs with every response containing an alternate primer to direct which stand is sequenced. Expulsion of unincorporated fluorescent dye eliminators is done after the finished sequencing response. Partition through capillary electrophoresis instrument was carried out, and sequence analysis of every reaction is done.

Single-nucleotide polymorphisms (SNPs) are utilized for forensic identification of humans, paternity testing, hair tone, and eye color ID in scientific analysis. Throughout the long term, various strategies have been created for SNP examination, for example, estimation of fluorescence, iridescence, and subatomic mass. Most measures are done in arrangements or on strong framework backing, for example, glass slide, chip, or dab. Autosomal SNPs can be utilized for some sort of scientific testing including investigation of corrupted examples. SNP loci on the Y chromosome are an additional likely marker for paternity testing due to the low transformation rate (Harbison and Fleming 2016).

Y Chromosome DNA Testing

Y chromosome DNA testing is significant for various utilizations of human hereditary qualities including scientific proof assessment, paternity testing, and verifiable examinations. It is important in terms of contemplating human movement patterns from the beginning of time and genealogical exploration. There are preferences and restrictions relating to Y chromosome testing when it comes to forensic examination and application. The essential estimation of the Y chromosome in legal DNA testing is that it is discovered as it were in guys. For the identification and determination of "male," the SRY (sex-deciding area of the Y) quality is considered. Since a great number of violations and crimes where DNA proof is useful, cases such as rapes, include men as the culprits, DNA tests intended to just analyze the male part can be significant (Fig. 5).

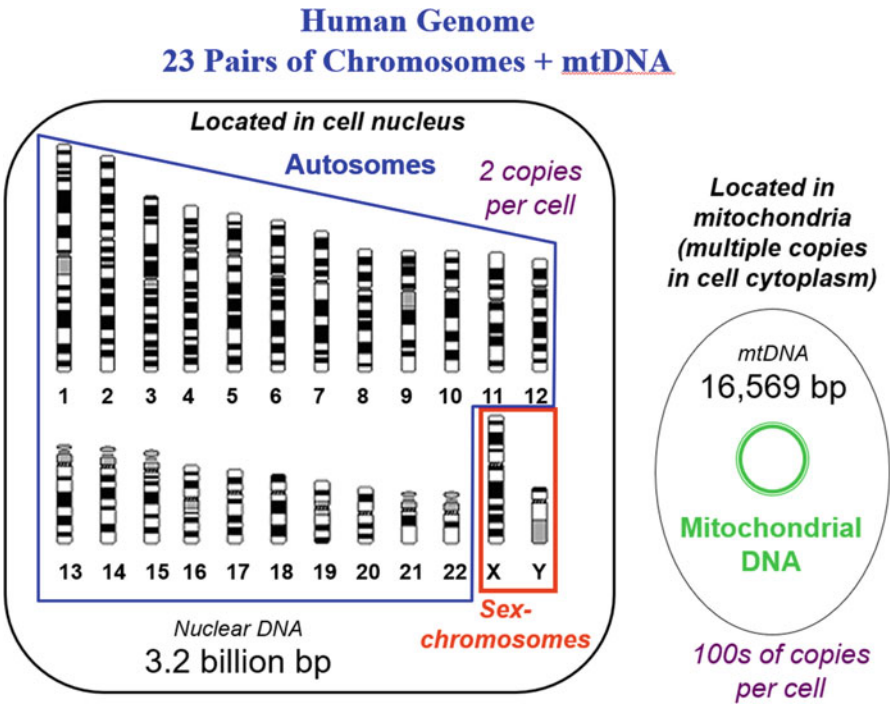


Fig. 5 Human genome and its classification

To completely use the Y chromosome for scientific purposes, it is important to see decisively what makes it quite a remarkable chromosome. The standard unavailability of Y chromosomes in females permits the utilization of the Y chromosome as a marker for human sex recognition, which can add accommodating data in forensic examinations. The precise male-specific inheritance of the NRY gives to explicitly investigate DNA segments that were only given by men and separate them from those given by females, which can be profoundly significant in blended stain examination in legal sciences, for example, in instances of rape. Simultaneously, inheritance free from recombination, from fathers to children, joined with low to direct change paces of most NRY-DNA (non-recombing portion of Y) polymorphisms, implies that male family members typically share similar NRY polymorphisms. This component has both favorable properties and drawbacks for legal applications of Y chromosome DNA. Impediments come in the manner that resolutions from Y chromosome DNA examination ordinarily can't be made on an individual level, as wanted in the criminological examination. This is on the grounds that in case of a matching DNA profile between exhibits from a suspect and a crime scene the speculations that either the suspect or on the other hand, any of his paternal male family members, has left the crime scene sample to have the estimated possibility. Benefits are credited to mutual Y-DNA profiles between male family members; a close male relative (paternal) of an expired alleged father can be utilized

to replace the father in paternity testing of a male posterity utilizing Y-DNA investigation in inadequacy cases, where autosomal DNA profiling regularly isn't enlightening. A similar rule can likewise be utilized in disaster casualty ID of men utilizing close or father's side male family members in situations where autosomal DNA profiling doesn't work. The haploid quality of the NRY likewise prompts the Y chromosome to have a lower populace size than the autosomes, with four duplicates of autosomal loci comparative with each Y locus. This lower compelling populace size results in the Y chromosome showing the least hereditary variety of any chromosome. As a result of the lower compelling populace size, Y polymorphisms can be all the more unequivocally influenced by hereditary float or populace level occasions, for example, bottlenecks or author impacts than autosomal loci. Moreover, the uneven spread of unmistakable polymorphisms is supported by the patrilineal transmission of the Y reflecting certain social practices, for example, patrilocality (where guys hold their familial lands, with females migrating) or polygyny (low quantities of guys having the most noteworthy conceptive achievement) (Hedman et al. 2009).

Drawbacks come within the way that conclusions from Y-chromosome DNA examination more often than not cannot be made on an individual level, as required in forensics. A larger part of the Y chromosome is moved legitimately from father to child without recombination to rearrange its qualities and give a more noteworthy hereditary assortment to people in the future. Arbitrary transformations are the main systems for variety after some time between paternal relations. Subsequently, while rejections in Y chromosome DNA testing results can help in measurable examinations, a match between a suspect and proof just implies that the person being referred to might have contributed the measurable stain, could it be a sibling, father, child, uncle, fatherly cousin, or a long distant cousin from his fatherly genealogy. Y-STRs change all the more quickly contrasted with Y-SNPs; Y-STR results show greater changeability and accordingly have more prominent use in legal applications. Commonly Y-STRs are portrayed as characterizing haplotypes, while Y-SNP characterizes haplogroups. As will be examined toward the finish of the part, Y-SNPs can be helpful in DNA parentage considers.

With heredity markers, the lineage data from every marker is alluded to as a haplotype instead of a genotype on the grounds that there is typically just a solitary allele for each person. Since Y chromosome markers are connected on a similar chromosome and are not rearranged with every generation, the statistical measures for an arbitrary match likelihood can't include the product rule. Hence, haplotypes acquired from genealogy markers can never be as successful in separating between two people as genotypes from autosomal markers. These autosomal markers are unlinked and isolate independently from one generation to another. Then again, the presence of family members having a similar ChrY grows the number of possible reference tests in missing people identification and mass calamity casualty ID proof. ChrY testing additionally helps in finding the family. Insufficient paternity tests where the father is dead or not available for the examination, can be conducted if ChrY markers are utilized. In any case, an autosomal DNA test is generally favored whenever possible since it gives a higher intensity of separation. The Y chromosome

has additionally become a well-known technique for following recorded human movement patterns through male genealogies. YHRD is also known as the Y-STR Haplotype Reference Database is the largest and most widely used forensic and general population genetics Y-STR database; it was developed by Lutz Roewer and colleagues at Humboldt University in Berlin, Germany. US Y-STR database which is a population-specific Y-STR database (US Y-STR) was launched in December 2007 to enable haplotype frequency estimates on five different US groups using the 11 SWGDAM recommended loci. The original version of US Y-STR contained 4796 African American profiles, 820 Asian, 5047 Caucasians, 2260 Hispanics, and 983 Native Americans.

In the rape cases where there are mixtures of DNA samples (both male and female), where the DNA levels of a female are high, male-specific testing can help in the assessment of a male culprit's profile. Extra mixtures may perhaps be examined (e.g., fingernail scrapings, salivation on the skin, and so on) by this technique. Transmission through the paternal line from a father to the entirety of his male children expands possible reference exhibit suppliers and, furthermore, helps in finding the family ancestries. Since patrilineal family members are indistinguishable, Y-STR composing can't be utilized to recognize among siblings or even far off male paternal family members. The product rule can only be used with the recombination between loci; without it, subsequently, the separation power of Y-STRs is restricted by the size of the populace information base utilized. Chromosomal duplication and deletions can muddle the examination (Kumar et al. 2016).

STR Typing

STR profiling is a widely utilized technique for human identification in forensic applications such as criminal profiling, paternity testing, mass catastrophes survivors' identification, legal casework forensic investigation, distinguishing proof of missing people, etc. According to this methodology, DNA is amplified from its sources for various human-specific polymorphic STR loci in a solitary PCR utilizing fluorescently marked preliminaries. These amplified regions are then isolated by electrophoresis, the crude information is examined utilizing programming that decides the size of each enhanced section, and genotypes by correlation with alleles in an allelic ladder that are run on a similar plate (Fig. 6).

For STR loci amplification, the primers in the multiplex frameworks are carefully planned to accustom the alleles of the loci amplified with the primers labeled with a similar dye and prevent their covering over one another. Utilizing primers marked with various dyes, it is conceivable to investigate loci creating amplified alleles with similar sizes of the fragments.

STR sequence groupings are named on the basis of the length of the recurrent units. In dinucleotides, the two nucleotides are repeated; similarly, in trinucleotides, three nucleotides are repeated and so on. Tetranucleotide repeats have become the major STR markers for personal identification proof. STR repeats differ in the length

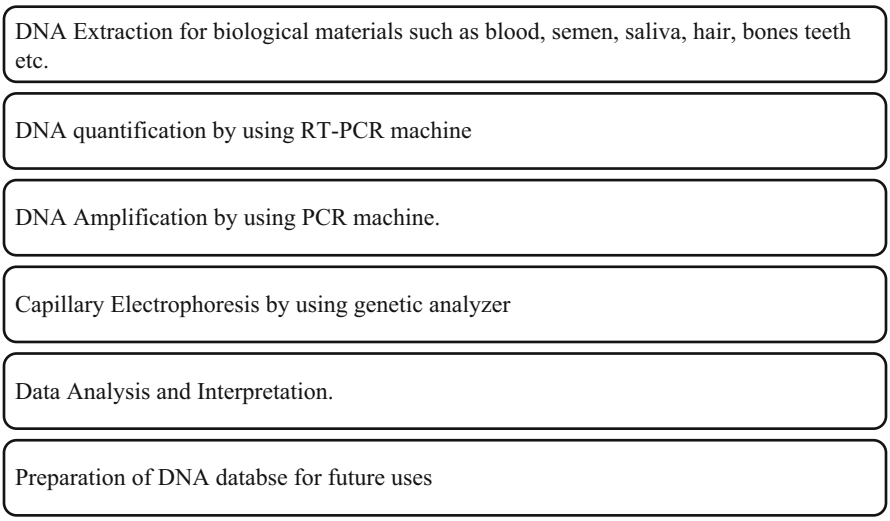


Fig. 6 STR profiling process

of the recurrent unit and the quantity of the. STRs are categorized further on the basis of their pattern of repetition. Basic STR units consist of units that have identical sequences and the same length. In compound repeats, two or more simple repeats are included, whereas, in complex repeats, various repeat units of variable length are included (Lindenbergh et al. 2012).

Tetranucleotide STR loci are preferred over VNTR minisatellites for a variety of reasons. It provides a narrow allele size range which allows multiplexing, and it also lessens the allelic dropout. With the use of tetranucleotide STR loci, stutter product formation lessens as well. As of 2017, the CODIS loci consist of the following: CSF1PO, FGA, THO1, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, and D22S1045. Of the first 13 CODIS STR loci, the 3 most polymorphic markers are FGA, D18S51, and D21S11. TPOX, CSF1PO, and TH01 normally display a minimal measure of variation between people. CODIS was designed to examine a target DNA record against the DNA facts contained within the database. Once a suit is diagnosed via the CODIS software program, the laboratories exchange records to affirm the match and set up coordination among their two companies. The comparison of the forensic DNA file against the DNA document in the database can be used to establish probable cause to acquire an evidentiary DNA pattern from the suspect. The regulation enforcement organization can use this documentation to gain a court order authorizing the gathering of a regarded biological reference sample from the culprit. The casework laboratory can then perform a DNA evaluation at the known organic pattern so that this evaluation can be supplied as proof in court.

An average STR profiling kit comprises of the main five parts such as:

1. A primer mixture for PCR which consists of oligonucleotides intended to enhance or amplify STR loci set. These primers are labeled with the one of each pair of fluorescent color.
2. A buffer mixture for PCR which consists of deoxynucleotide triphosphates, MgCl2, and different reagents that are required to perform PCR reaction.
3. An enzyme is known as DNA polymerase, which is usually premixed with the buffer mixture of PCR.
4. An allelic ladder consisting of the basic alleles for the amplification of STR loci for the calibration of allele repeat size.
5. A positive control DNA test to check that the unit reagents are working appropriately.

Various packs are accessible for single or multiplex PCR enhancement of STR markers utilized in DNA composing. An allelic ladder basically refers to the artificial amalgamation of the regular alleles present in the human populace for a specific STR. They are produced with the equivalent primers as tested exhibits and hence give a reference DNA size to every allele in the ladder. Allelic ladders are significant for exact genotype determination. These allelic ladders act like a ruler for every STR locus. They are important to fit the various estimations achieved from various instruments and conditions utilized by different research facilities. Allelic ladders are made by the union of genomic DNA or locus-specific PCR items from various people in a populace, which has alleles that represent the variation of a specific STR marker. The tests are then co-enhanced to create an artificial exhibit containing the basic alleles for the STR marker. The amount of these alleles is adjusted by changing the information measure of each part with the goal that the alleles are represented equally in the ladder (McElfresh 1991).

According to various researches, miniSTRs are highly sensitive and strong for the evaluation of low template and degraded DNA. These miniSTRs are extremely useful for the analysis of forensic samples and show the proportion of samples that advantage from evaluation with additional miniSTR loci in terms of ensuing in a user profile. Degradation of DNA leads to the fragmentation of DNA into smaller units known as template fragments. The smaller amplicon length of miniSTRs results in a better amplification of fragmented DNA. In addition to the abovementioned advantages, analysis of additional (mini)STR loci affords additional discriminative capability. Analysis of miniSTRs in forensic DNA samples has major advantages over the bigger-sized STRs. DNA that is degraded to fragments smaller than the medium- and massive-sized STR amplicons could nonetheless be detected with the aid of miniSTR analysis. Forensic DNA exhibits contain DNA that remains when someone comes into touch with an object. Typically, those samples comprise handiest minute contents of DNA. Many of those samples include degraded DNA due to publicity to the surroundings. For those motives, forensic DNA exhibits are few of the toughest samples to get usable DNA profiles. Its predicted that the utilization of miniSTRs for the examination of touch DNA samples outcomes in a better percent of DNA profiles due to the advantages of miniSTRs. The amplification of smaller amplicons typically

produces greater efficient, growing the signal and sensitivity of the PCR technique. The impacts of the miniSTR evaluation implementation of can be studied by several methods (Menotti-Raymond et al. 2003).

Mitochondrial DNA

Mitochondrial DNA or mtDNA refers to the circular genome found inside the mitochondria which are different from the nuclear DNA. Researches have uncovered that mtDNA is a round genome and it's around 16,569 nucleotides long. A very much characterized, noncoding part of the genome endures the aggregation of mutations, which can be questioned to build up a forensic mtDNA profile. The profile data will frequently be useful when distinguishing missing people also unraveling various crimes. For instance, a correlation of mtDNA characterizations permits the relation of biological proof with one individual in a criminal case, while barring people who are not related in a similar case as the origin of DNA obtained from the scene of the crime. Although the maternal inheritance of the mitochondrial genome lessens the general separation capability of the mtDNA testing framework, family members from a similar maternal genealogy usually have the same profiles. In any case, the genome's higher mutation pace can bring about the incidents of heterogeneous pools of mtDNA types that can fundamentally expand the intensity of segregation and can be communicated in contrasting proportions across maternal lines. The material introduced in this section gives a review of the forensically important attributes of the mtDNA genome and how these qualities can be utilized to respond to questions brought up in legal examinations. Recovery of DNA data from destructed DNA is possible in some cases with mitochondrial DNA or mtDNA. While a nuclear DNA test is normally more significant, a mtDNA result is superior to no outcome by any means. There are around hundreds of duplicates of mtDNA in every cell; the likelihood of getting a DNA composing result from mtDNA is higher than that of polymorphic markers found in nuclear DNA, especially in situations where the measure of extricated DNA is minuscule, as in tissues, for example, bone, teeth, and hair. At the point when remains are very old or seriously desecrated, bone, teeth, and hair are the main organic sources left from which DNA can be extracted. The essential trademark that leads to the recovery of mitochondrial DNA (mtDNA) from damaged examples is the higher duplicate number of mtDNA in cells comparative with the nuclear DNA from which STRs are enhanced. To put it plainly, however atomic DNA contains substantially more data, there are just two duplicates of it in every cell (one maternal and one fatherly), while mtDNA gives a touch of helpful hereditary data many occasions per cell. As a result of their higher numbers, some mtDNA atoms are bound to get by than nuclear DNA (Prinz and Lessig 2014).

DNA molecules are situated inside the cell core, as chromosomes are one of the sources of DNA in present-day eukaryotic cells. Mitochondria found in the cytoplasm of most cell types contain a second intracellular DNA genome. As per the broadly acknowledged endosymbiotic hypothesis of mitochondrial emergence,

mitochondria were gotten from α-*Proteobacteria* that lived around 2 billion years prior inside pre-eukaryotic cells looking like old protists. Throughout the course of evolution, the endosymbiont lost its capacity to live outside the eukaryotic cell, and parts of its related DNA were retained. The endosymbiont is converted into the mitochondria, and its genome is turned into mtDNA. The mitochondrion went on with the essential function of the creation of cell energy by methods for oxidative phosphorylation and the union of molecules of ATP. As the evolution process continued, the mitochondrion preserved various qualities associated with this significant biochemical cycle, while losing huge numbers of the tasks expected to stay alive as a totally independent solitary organism. The strands which are complementary, belonging to the mtDNA sequence, are precisely unique in their structure. The "heavy strand" consists of purine nucleosides (adenosine and guanosine), though the "light strand" consists of the wealthy in pyrimidine nucleosides (thymidine and cytidine). The coding locale includes roughly 93% of the genome, with the quality successions thickly orchestrated. There are 37 qualities found in the coding area grouping: 22 qualities for transport RNAs (tRNA), 2 qualities for ribosomal RNAs (rRNA, the 12S and 16S subunits), and 13 qualities for catalysts in the respiratory chain associated with the cycle of oxidative phosphorylation and, also, ATP creation.

By far most of the human genome is situated inside the core of every cell. In any case, there is a little, roundabout genome found inside the mitochondria, the energy-delivering cell organelle living in the cytoplasm. The number of mtDNA atoms inside a cell can go from hundreds to thousands. Normally there are 4–5 duplicates of mtDNA atoms per mitochondrion with a scope of 1–15. Since every cell can contain several mitochondria, there can be up to a few thousand mtDNA particles in every cell as on account of ovum or egg cells. The usual number has been assessed at around 500 in the majority of cells. It is this enormous number of mtDNA particles in every cell that leads to more prominent success as compared to the nuclear DNA markers, with the examples that may have been harmed or degraded with heat or moistness. Mitochondrial DNA consists of around 16,569 base pairs with the complete number of nucleotides in a particular mtDNA genome shifting because of little insertions or deletions. For instance, there is a dinucleotide rehash at positions 514 to 524, which in most people is ACACACACAC or (AC)5, however has been seen to change from (AC)3 to (AC)7. Note that with two duplicates of atomic DNA (3.2 billion bp from each parent) and in any event, expecting that there are 1000 duplicates of mtDNA or 16,569 bp per mtDNA in a cell, mtDNA makes up just about 0.25% of the all-out DNA content of a cell.

The majority of the concentration in scientific DNA studies to date has included two hypervariable locales inside the control district generally alluded to as HVI (HV1) and HVII (HV2) (Slatko et al. 2018). Every so often a third part of the control area, known as HV3, is analyzed to give more data concerning tried example human mitochondrial DNA is only inherited from the maternal side. During the process of conception, just the sperm's core enters the egg and joins with the egg's core. When the zygote cell partitions and a blastocyst is created, the cytoplasm and other cell parts spare the nucleus with the mother's unique egg cell. Mitochondria with their mtDNA cells are passed straightforwardly to all posterity free of any male impact.

Subsequently, the mutation does not occur, a mother passes along her mtDNA type to her children, and so the siblings and the maternal family members have an indistinguishable mtDNA sequencing. Henceforth, a person's mtDNA type isn't one of a kind to the person in question. Mitochondrial DNA variety is broadly studied in different branches of science. Clinical researchers have connected various sicknesses to changes and mutations in mtDNA. Evolutionary scientists look at human mtDNA arrangement variety comparative with different species with an end goal to decide connections. A genuine illustration of this application is the assurance that Neanderthals are not the immediate predecessors of current people dependent on control area groupings decided from antiquated bones. Anthropologists study contrasts in mtDNA groupings from different worldwide populace gatherings to look at inquiries of the family line and movement of people groups all through history. Several papers have been distributed in these fields over the recent many years. Hereditary genealogists are currently utilizing mtDNA and Y chromosome markers while trying to follow lineage where paper trails run cold. In the previous few years, various intriguing recorded IDs have been performed with the guide of mtDNA testing. Steps in mitochondrial DNA analysis are as follows:

The mtDNA is extracted from the sample, and then it is amplified by PCR and HV1 and HV2 regions. These HV1 and HV2 amplicons are sequenced, and this sequence is confirmed with both forward and reverse strands. Differences are noted from the reference sequence. Meanwhile, another process must be carried out separately which includes the extraction of DNA from the reference sample. The same abovementioned steps are repeated, and the differences are noted from the Anderson or reference sequence. These sequences that are obtained by both analyses are then compared with the database to determine the haplotype frequency estimate. Steps for assessment of mtDNA exhibits are as follows : The evidence or questioned exhibit comes from a crime scene or a mass calamity. The reference or known exhibit might be a maternal family member or the suspect in a criminal examination. In a criminal forensic examination, the victim may likewise be examined and compared with the questioned and known outcomes (Sullivan 1994).

PCR duplication of mtDNA is normally completed with 34–38 cycles. Conventions for profoundly desecrated DNA examples even call for 42 cycles. In some cases, more Taq is added to beat PCR inhibitors, for example, melanin. It is significant to remember that sensitivity is boosted with mtDNA testing since it is the last attempt to acquire DNA results from an exhibit. The higher the affectability of any assay, the more prominent the possibility for spoiling and consequently more care and consideration are normally needed with mtDNA work than with regular STR composing. Major mtDNA varieties between the human populace are found inside the control locale or displacement loop (D-loop). Two points inside the D-loop are called hypervariable points I (HV1, HVI, or HVS-I) and hypervariable locale II (HV2, HVII, or HVS-II) which are usually analyzed by PCR enhancement followed by sequence examination. Roughly 610 bp are ordinarily assessed – 342 bp from HV1 and, furthermore, 268 bp from HV2.

For DNA sequencing, the Sanger technique was introduced more than 30 years back (Sanger et al. 1977). This Nobel Prize-winning sequencing procedure is still

broadly utilized, and the technique includes the polymerase fusion of dideoxyribonucleotide triphosphates (ddNTPs) as chain terminators and then by a separation step equipped for single-nucleotide resolution. There is no hydroxyl bunch at the 3'finish of the DNA nucleotide with a ddNTP, and, subsequently, chain development ends when the polymerase joins a ddNTP into the synthesized strand. Extendable dNTPs and ddNTP eliminators are both present in the response blend so that a few bits of the DNA particles are broadened. Toward the finish of the sequencing response, a progression of molecules is available that vary by one base from each other. In the Sanger sequencing measure, every DNA strand is sequenced in independent responses with a solitary groundwork. Frequently either the forward or reverse PCR primers are utilized for this reason. Four diverse shaded fluorescent colors are connected to the four distinctive ddNTPs (Yang et al. 2014).

Heteroplasmy is characterized as a combination of more than one mtDNA genome sequence inside a cell or between cells of a solitary individual. As they gather, these heteroplasmic variations can prompt illnesses and can profoundly affect the process of aging. In a scientific analysis procedure, these variations can give extra layers of segregation potential to help distinguish human remaining parts and have become a standard for the examination procedure when investigation on exhibits is conducted in forensic case analysis. Accordingly, in light of the fact that the presence of heteroplasmy can affect the evaluation of mtDNA examinations among evidentiary and exhibits that are known experienced in criminological casework, it is imperative to comprehend the biological premise of heteroplasmy, how regularly it is noticed comparable to the sequence identification strategy being utilized, and how best to decipher the results. Since most eukaryotic cells contain hundreds to thousands of duplicates of mtDNA, it is conceivable and without a doubt likely that irregular mutations exist in various duplicates of the mitochondrial genome all through a person. The mtDNA has a higher rate of mutation than the nuclear DNA genome. Despite the fact that it is regularly estimated that the high rate of mutation is because of lesser fidelity in the γ-polymerase utilized for the genome replication, it ends up that this adaptation of the replicative DNA polymerase is very loyal. All things considered, the probable origin of mutation is from DNA destruction instigated by responsive oxygen species (ROSs) produced during oxidative phosphorylation, trailed by the activities of a not exactly satisfactory fix framework.

X Chromosome Analysis

The X chromosome (ChrX) has potential criminological scientific and human identification applications because of its pattern of inheritance as compared with other hereditary markers. Usually, males have one X chromosome and one Y chromosome, while females have two X chromosomes despite the fact that there are infrequently some sporadic karyotypes, for example, XXY (Klinefelter disorder; Giltay and Maiburg 2010), XXX, and XYY. More than 40 STR markers have been described from the X chromosome, and populace examinations have been performed with huge numbers of these X chromosome STRs (X STRs). X chromosome STR

composing can be useful in some familial relationships examination circumstances especially with insufficient paternity situations where a DNA exhibit from one of the guardians isn't accessible for testing. For instance, if a father/girl child parentage relationship is being referred to, X STRs might be useful because of the 100% transmission of the dad's X chromosome to his little girl. Then again, in a father/boy child parentage question, Y chromosome results would be useful records a few applications for X chromosome DNA testing. ChrX testing can be particularly useful in some missing people or catastrophe casualty distinguishing proof circumstances where direct reference tests are not accessible; what's more, organic family members must be tried to help human identification proof.

Autosomal short tandem repeat or STR examination has become the foundation of forensic personal identity testing since the 1990s when the first fluorescently marked STR markers were portrayed. From that point forward, a large number of polymorphic STR markers have been explained, and at present, a number of combinations of autosomal markers permit the strict ID of people to the level of one out of trillions. In missing people and insufficient paternity cases, Y chromosomal STRs grow the pool of family reference tests that can be utilized to affirm character and have demonstrated valuable in circumstances where the male DNA is mixed with female DNA and female DNA is present in much larger quantity as compared to male one as would be the situation with a vaginal swab from a rape case (Weedn 2007).

In the past few years, STR markers situated on the X chromosome have arisen as an extra tool in this forensic scientific analysis. X chromosomal STRs can be utilized to enhance STR profiling due to their inheritance pattern, and, correspondingly, the expansiveness of distributed writing regarding the matter has extended incredibly lately. STR markers on the X chromosome might be valuable for forensic purposes. To start, missing people cases generally require the examination of family members due to an absence of direct reference material. Regularly, mitochondrial DNA (mtDNA) composing can be utilized to address the potential for destructed or damaged exhibits, for example, skeletal remaining parts, especially in shut populaces also, when a direct maternal reference is accessible, as a result of its moderately high duplicate number and ensured area inside the mitochondria of the cell. Nonetheless, mtDNA is maternally acquired; accordingly, where maternal references are inaccessible or where the unidentified individual matches one of the most well-known mtDNA haplotypes, mtDNA testing alone might be lacking. In such cases, markers on the X chromosome may give extra data.

X chromosomal STRs can be especially helpful for any parent-offspring relationship that includes one female or more (e.g., father-little girl, mother-child, or mother-little girl), for instance, the related family of parents with one son who is the father of a little girl. In this situation, if that son and his wife are unable to come for testing, it very well might be important to utilize the grandparents' DNA profiles to reassociate their granddaughter. In this particular situation, autosomal STRs by and large give a low probability proportion of a relationship since there is on normal just one-quarter sharing of alleles between a grandparent and a grandkid. X chromosomal STRs, then again, end up being more helpful since the X chromosome of the girl child was

acquired altogether from his mom's genome with no commitment from his dad. This X chromosome was then passed in full to the granddaughter. In this model, one would anticipate seeing one allele from every X STR marker of the grandma present in the granddaughter's X STR profile; hence, X chromosomal STRs will no doubt beat autosomal STRs. Other maternally related situations, for example, distinguishing cousins or auntie niece connections utilizing X chromosomal STRs to replace or expand autosomal STR testing, have been proposed.

Mutations generally happen because of strand slippage during DNA replication, and they are responsible for the significant system of the serious level of polymorphism found in human microsatellites. Single-step mutations such as deletions or insertions of one recurrent unit are most occurring, and they influence longer alleles more often than shorter ones. This pattern is noticed for markers on the Y chromosome, and no distinction is normal on the X chromosome. Mutation rates can be varied and in the comparison, among men and women, with one approximation of the proportion of mutations from paternal to maternal changes at 17:3. Mutation rates can likewise fluctuate with the populace. As the quantity of new markers depicted in the writing expands, the potential for classification contrasts additionally increments. There are 36 X STRs that are ordinarily utilized by distinctive scientific research centers, and, sometimes, contrasts in allele terminology make correlations of distributed populace information between the research centers, best case scenario monotonous and to say the least unimaginable (Yang et al. 2014).

Microbial and Animal Forensics

Forensic DNA profiling is usually done for the investigation of perpetrators which basically involves human DNA and that DNA is important in proving the innocence or guilt of the suspect. But there can be other DNA also, apart from the human ones that might be valuable in exhibiting the innocence or guilt of an individual associated with a crime. Pets and domestic animals', for example, felines (cats) and canines (dogs), hair might be utilized to establish a link between the suspect and a crime scene. Showing that an herbal exhibit belongs to a particular plant from a specific plant can help in linking of a suspect to the scene of the crime. It may also help in showing that the body of an expired casualty may have been moved from the homicide site. DNA testing would now be able to be utilized to connect the origins of cannabis. One of the major applications of forensic DNA typing in the future scenario would include the examination of materials such as anthrax (bacterium *Bacillus anthracis*) which can be used for bioterrorism activities (Slatko et al. 2018).

Such an attack of bioterrorism happened in the United States. Quite simply, *Bacillus anthracis* bacterium spores were sent to media and offices, through the mail. It caused about 22 contaminations and, also, 5 deaths. As the aftereffect of this assault, the world got mindful of a weakness in which numerous in the field of counterterrorism was very discerning. Bioterrorism is a genuine danger. The utilization of the US mail as a dispersal vehicle uplifted concerns since it showed that a mail delivery could be utilized to uncover individuals to a fatal microbe (an infection

causing specialist. The Federal Bureau of Investigation (FBI) of USA, is the organization with the essential duty to explore the crimes, were not well set up to attempt the criminological examinations identified with the examination of the Bacillus anthracis assault; supporting criminological science examination of microbial proof was amazingly restricted at that time. Animal and plant testing in the forensic analysis is relatively new as compared to conventional analysis. Illegal trafficking and trade of endangered species have led to the realization that such crimes are not of any less important than the others.

A cat has 18 sets of autosomes (nonsex chromosomes) and the sex chromosomes X and Y. For each of the *Felis catus* chromosomes, genetic markers have been established. MeowPlex is a group of STR markers, and it consists of 11 STRs on 9 distinct autosomes. In this assay, an extra marker was added for the identification of the gender by the augmentation of PCR primers. These primers are specific for the SRY gene on the feline (cat) Y chromosome. For the process of amplification, the PCR items are in between the size scope of 100 bp to 400 bp and utilize three dye color tones. Cat STR allele frequencies from pet felines have been studied to show the distinctive feature of DNA profiles in forensic examination.

An ongoing quantitative polymerase chain response (PCR) test (see ▶ Chap. 6, "Tools and Techniques Used in Forensic DNA Typing") for assessing the DNA yield extricated from homegrown feline examples has been created (Menotti-Raymond et al. 2003). This test is fit for distinguishing down to 10 femtograms of cat genomic DNA and uses high duplicate number short interspersed atomic components (SINEs). A large number of these pets and domestic animals usually shed hair, and hence these hairs could be gotten or abandoned at the scene of the crime by a culprit. The perpetrator may unconsciously have cat or dog hairs on their clothes or shoes, from a victim's pet. They could have carried it away, and this might help in connecting the perpetrator to the victim and the crime scene. Animal DNA pieces of evidence can indicate to three different explanations: the animal could be a victim, a culprit, or a witness. DNA testing can help in cases of animal abuse or animal trafficking cases. The remaining evidence of a lost pet can be recognized through genetic or DNA examination. Normally genetic markers such as short tandem repeats (STRs) and mitochondrial DNA (mtDNA) are analyzed just like how human DNA is examined. In cases the point when creatures are associated with an assault on an individual, DNA composing might be utilized to recognize the creature culprit. In case that the victim has perished, at that point DNA proof might be the main evidence that the animal has done a particular crime. DNA testing of animals can "absolve" innocence of living beings (other than humans) so that they are not unnecessarily decimated.

Animal DNA has been utilized effectively to interface suspects to wrongdoing scene. Research has been conducted on the exchange of animal hair during stimulating criminal conduct found that several feline hairs or canine hairs could be moved from the homes of casualties to a thief or an attacker. The number of hairs discovered was high to a great extent that it is practically impossible to find a house where a domestic pet lives, without being "polluted" by hairs of the pet (cats or dogs), when the proprietor depicts their animals as not a good source of the hair. The issue with

the hairs which are shed regularly is that they don't contain roots, so nuclear DNA may not be available in adequate amounts for STR typing. Mitochondrial DNA might be a more suitable option for a considerable lot of these sorts of shed hair to get transferred.

Previously, forensic analysis of animals or animal hair was based upon morphological models' species-level identification. Isoelectric centering of keratins were the principal endeavors utilized toward the molecular identification of animal samples. Genetic individualization was done by utilizing profoundly recurring minisatellite loci to create singular explicit "fingerprints" of human DNA. Jeffreys et al. were additionally the first to illustrate the capability of DNA fingerprinting of animals by utilizing human minisatellite DNA to produce multi-locus DNA fingerprints of canine and feline DNA. The Locard's Exchange Principle permits forensic investigators to interface unique source and target surface and has a point of view on primary and secondary movement. The animal proof isn't excluded from Locard's standard and has become important for recognizable proof and individualization of follow or move biomaterial from crime scenes. The DNA examination of animal's biomaterial, nonetheless, just gives crucial analytical leads and remaking of the scene of the crime; the examination of animal DNA can't add to the individualization of any human suspect, and as a rule, this sort of proof should be enhanced by different types of actual proof. For the production of the forensic DNA typing model, 49 tri-and tetranucleotide STR loci were disengaged from felid STR-enhanced genomic libraries. Tetranucleotide STRs have been utilized for human profiling since they limit the generation of "stutter band" items created during PCR by amplification which can convolute the translation of genotypes from mixed DNA exhibits. The loci were included in the genetic maps of the domestic feline comparative with 579 coding qualities and 255 STR loci to choose unlinked markers. Thusly, the loci were screened in a little board of outbred felines and 28 feline varieties (3–10 creatures/breed, $n = 213$), to distinguish a board of markers with the most elevated separating power. A bunch of 11 exceptionally polymorphic loci was at first chosen for the typing process. The loci were unlinked, exhibited high heterozygosity over various feline varieties, and, furthermore, demonstrated the absence of cross-species amplification. A multiplex amplification procedure was composed so the loci could be amplified with as meager as a nanogram of DNA. It also incorporated a gender recognizing STS on the Y chromosome, enhancing a part of the SRY quality. The PCR results of the 11 loci were planned in a size ranging from 100 to 415 bp, named with one of 4 fluorescent labels, with no allele cover with adjoining loci, and the SRY item noticeable at 96 bp. Approval investigations of the multiplex exhibited that total item profiles could be created with as meager as 125 pg of genomic DNA, with a nonappearance of "allele dropout." A database of the genetic information of the domestic cat varieties has been created from the multiplex with which to register composite match probabilities. In contrast to felines, which can be forceful yet infrequently dispense critical injury, canines can be culprits that produce critical injury; a canine scientific genetic testing unit was created. This kit is known as the Finnzymes Canine Genotypes™ 1.1 Multiplex STR Reagent pack which was created and approved utilizing a board of canine-explicit STRs utilizing

species testing utilizing loci on the mitochondrial genome which has become a standard technique in preservation biology and phylogenetic examinations (Harbison and Fleming 2016).

The danger of a bioterrorist assault has been the subject of worldwide concern. Various US public intelligence estimates and reports have inspected the issues. Because of current innovation and an enormous expansion in information gathered about atomic science of microbial specialists, researchers can all the more viably and proficiently distinguish microorganisms from an assortment of lattices to incorporate those from the climate. Progress around there has been important to the customary general well-being network for an extremely long time. It is likewise important to those networks that will be liable for attribution following a criminal demonstration or psychological oppressor assault utilizing a bioagent. To eventually distinguish the people or associations liable for such an assault will probably require thorough utilization of current microbiological logical apparatuses combined with conventional measurable orders. Proof gathered as a component of an organic attribution examination will yield exceptional kinds of microbiological proof that might be explicit to the idea of the assault. As instances of such microbiological proof, they referred to the accompanying: feasible examples of the microbial specialist, protein poisons, nucleic acids, clinical examples from casualties, research facility gear, scattering gadgets and their substance, natural examples, tainted apparel, or follow proof explicit to the cycle that created as well as weaponized the organic specialist. This gathering recognized that there was a requirement for research and, also, advancement endeavors to improve current capacities in microbial crime scene investigation and that this would require exertion among numerous components of the public government. The proposals found in this public procedure might be applicable to numerous nations endeavoring to all the more likely plan for the danger of bioterrorism. Regardless of whether for a general well-being or a bioterrorism occasion, the distinguishing proof, assortment, and investigation of suitable examples are a basic essential for effective discovery of organisms. Not exclusively can clinical material from tainted patients be used for the recognition; however, regularly creature or natural examples will yield basic data. For fruitful ID of dubious specialists, it is critical to utilize both exemplary microbiological methods and current atomic science procedures microbial forensics and, likewise, with other customary legal controls, requires execution and also use of conventions and practices that will eventually yield results that can be utilized either by leaders in law implementation or potentially public/global security. Recognizable proof of fitting examples, “chain of care” records (documentation that tracks actual control of tests), the utilization of legitimate logical conventions via prepared staff, adherence to quality confirmation gauges, and guaranteeing the protected capacity and safeguarding of tests – all are of basic significance for effective attribution examinations. Fairly extraordinary to this sort of proof is guaranteeing the well-being of anybody associated with the assortment and treatment of microbial scientific proof. When managing pathogenic microorganisms, there are extraordinary dealing with techniques that should be continued to satisfactorily forestall extra mischief to staff and the encompassing network. All staff engaged with reaction or treatment of these

bioagents should go through specific preparation to guarantee the well-being and security of the people included and the general climate.

Next-Generation Sequencing

NGS is primarily a high-throughput system for DNA sequencing, also considered as cutting edge sequencing. It has quickly advanced in the course of recent years, and new strategies are consistently being popularized. As the innovation grows, so do increments in the quantity of comparing applications for essential and applied science. This NGS innovation, with its high-throughput limit and minimal effort, has quickly become popular, and it has turned into a significant insightful instrument for some genomics scientists. In the field of forensic scientists, new opportunities have been emerged crediting to the NGS technology. NGS innovation can be applied various areas of scientific interest regarding genetic contexture, for example, mitochondrial, autosomes, and sex chromosomes. Besides, NGS innovation can likewise have expected applications in numerous different parts of exploration. These incorporate DNA information base development, phenotypic derivation, monozygotic twin examinations, body liquid and species recognizable proof, and criminological creature, family line/ancestry, and plant and microbiological investigations. Here we audit the use of NGS innovation in the field of legal science with the point of giving a reference to future criminology studies and practice. This technology alludes to non-Sanger-based high-throughput DNA sequencing innovation. Millions or billions of DNA particles can be sequenced in equal, consequently expanding the throughput considerably and limiting the requirement for the fragment cloning technique that is frequently utilized in Sanger sequencing. It incorporates second-age sequencing innovation dependent on circle exhibit sequencing, which can examine countless examples at the same time, just as third-age sequencing innovation, which can decide the composition of single DNA particles.

Roche presented the 454 Genome Sequencing System, the world's first pyrosequencing-based high-throughput sequencing framework, in 2005. The initial 454 Genome Sequencer was equipped for creating around 200,000 peruses of 110 base sets (bp) long. To begin with, these advancements don't need bacterial cloning of DNA parts; all things being equal, they depend on the readiness of NGS libraries in a sans cell framework. This technique, rather than many sequencing responses, can parallelize the thousands-to-many-a large number of sequencing response. The sequencing yield is generated thorough NGS straightforwardly identified with no requirement for electrophoresis. The tremendous number of peruses produced by NGS empowered the sequencing of whole genomes at an exceptional speed, and subsequently it came to be generally utilized in different fields of life sciences. Be that as it may, one disadvantage of second-age sequencing innovation is their moderately short understand lengths, which has brought about challenges in ensuing succession grafting, gathering, interpretation, and bioinformatics examination. Besides, standard PCR was used to arbitrarily intensify genomic sections during library planning. Due to the complex structure of genomes, factors, for

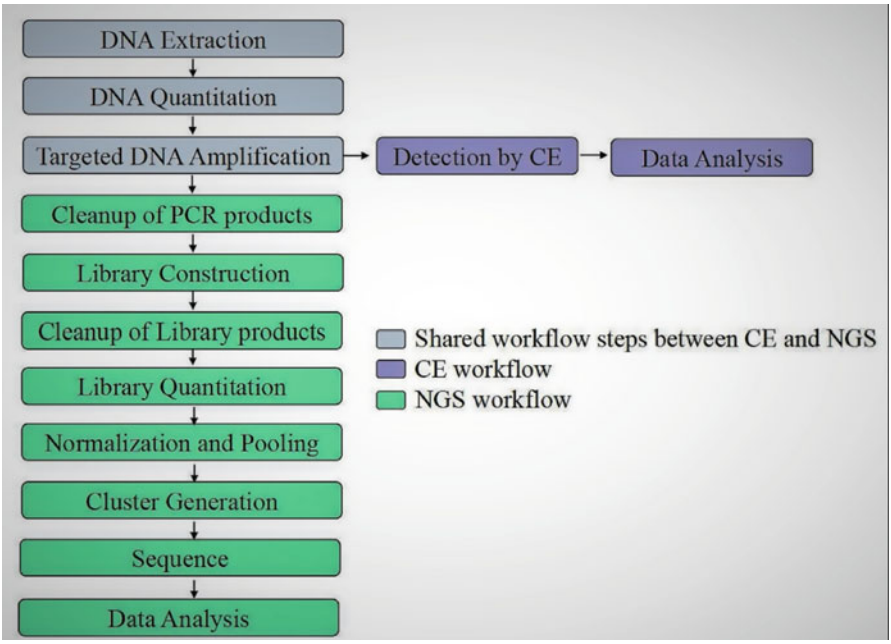


Fig. 7 Next-gen sequencing steps

example, auxiliary structure and warm dependability, will influence the productivity of PCR enhancement.

The utilization of DNA techniques in forensic examinations has delivered DNA investigation a significant device in forensics. In the forensic DNA investigation, contrasted with different fields of life sciences, DNA analysis is based onto the low copy number, exceptionally debased and defiled examples, so there is the requirement for high precision and reproducibility, and time and cost examination. Today, most of scientific DNA tests utilize PCR- and narrow electrophoresis (CE)-based part investigation techniques to distinguish length variety in short tandem repeat (STR) markers. The CE-based Sanger sequencing has been utilized to examine explicit locales of mitochondrial DNA (mtDNA) (Fig. 7).

Conclusion

This chapter analyzed the study of DNA typing methods and techniques used at present for the examination of DNA. It focused on the biology basics, technology, innovation, and genetic characteristics of short tandem repeats markers or (STR) markers. These markers incorporate the most well-known forensic DNA investigation techniques utilized today. The materials in this chapter are intended to provide information basically for forensic scientists and law professionals, and it describes the complex examination methods with simple language, for a clear understanding

of DNA profiling. This content ought to likewise straightforwardly profit understudies studying criminological DNA investigation in a scholastic climate. These principles aims to improve the nature of work performed in scientific research centers by requiring specialized supervisors and DNA analysts to have preparing in organic chemistry, hereditary qualities and atomic science in request to pick up a fundamental comprehension of the establishment of criminological DNA examination. The research publications have developed significantly on the subject of STR composing and its utilization in criminological DNA testing. New points, for example, single-nucleotide polymorphisms (SNPs) and Y chromosome testing include picked up more noteworthy acknowledgment inside the legal network. A very thorough gander at mitochondrial DNA and its application to legal DNA examination is explained. There is refreshed data on new DNA extraction methods, ongoing PCR for DNA evaluation, and multi-narrow electrophoresis instruments that are currently utilized in numerous forensic DNA research facilities.

References

Alamoudi E, Mehmood R, Albeshri A (2018) DNA profiling methods and tools: a review. 1:216–231

Butler JM (2010) Fundamentals of forensic DNA typing. In Elsevier Inc

Butler JM (2012) Advanced topics in forensic DNA typing: methodology. In Elsevier Inc

Butler JM (2015) The future of forensic DNA analysis. Royal Society Publishing. https://doi.org/10.1098/rstb.2014.0252

Butler SL, Hansen MST, Bushman FD (2001) A quantitative assay for HIV DNA integration in vivo. Nature Medicine 7(5):631–634. https://doi.org/10.1038/87979

Carracedo Angel S-DP (2005) Forensic DNA-typing technologies: a review. In: Methods in molecular biology. Humana Press. https://www.researchgate.net/publication/8157674_Forensic_DNA-typing_technologies_a_review

Casey E, Rose CW (2010) Forensic analysis. In: Handbook of digital forensics and investigation. Elsevier Ltd., pp 21–62. https://doi.org/10.1016/B978-0-12-374267-4.00002-1

Cavanaugh SE, Bathrick AS (2018) Direct PCR amplification of forensic touch and other challenging DNA samples: a review. In: Forensic science international: genetics, vol 32. Elsevier Ireland Ltd., pp 40–49. https://doi.org/10.1016/j.fsigen.2017.10.005

DNA Fingerprinting: Purpose, Procedure, and How It's Used (n.d.) Retrieved November 25, 2020, from https://www.webmd.com/a-to-z-guides/dna-fingerprinting-overview#1

Giltay JC, Maiburg MC (2010) Klinefelter syndrome: Clinical and molecular aspects. Expert Review of Molecular Diagnostics 10(6):765–776. https://doi.org/10.1586/ERM.10.63

Harbison S, Fleming R (2016) Forensic body fluid identification: state of the art. Research and Reports in Forensic Medical Science 11. https://doi.org/10.2147/rrfms.s57994

Hedman J, Nordgaard A, Rasmusson B, Ansell R, Rådström P (2009) Improved forensic DNA analysis through the use of alternative DNA polymerases and statistical modeling of DNA profiles. BioTechniques 47(5):951–958. https://doi.org/10.2144/000113246

Issues with DNA Fingerprinting in Forensic Lab: A Review (n.d.) Retrieved November 25, 2020, from https://escientificpublishers.com/issues-with-dna-fingerprinting-in-forensic-lab-a-review-JMRCR-01-0002

Jamieson A, Bader S (2016) A guide to forensic DNA profiling. In: Bader S, Jamieson A (eds) Wiley, 1st edn. John Wiley & Sons Ltd.

Kumar S, Verma AK, Singh P, Singh R (2016) Current scenario of forensic DNA databases in or outside India and their relative risk. Egypt J Forensic Sci 6(1):1–5). Egyptian Forensic Medicine Authority. https://doi.org/10.1016/j.ejfs.2015.03.002

Lindenbergh A, De Pagter M, Ramdayal G, Visser M, Zubakov D, Kayser M, Sijen T (2012) A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. Forensic Sci Int Genet 6(5):565–577. https://doi.org/10.1016/j.fsigen.2012.01.009

McElfresh KC (1991) DNA fingerprinting. Sci Am 264(1):12. https://doi.org/10.5005/jp/books/14210_8

Menotti-Raymond M, David VA, Chen ZQ, Menotti KA, Sun S, Schäffer AA, Agarwala R, Tomlin JF, O'Brien SJ, Murphy WJ (2003) Second-generation integrated genetic linkage/radiation hybrid maps of the domestic cat (Felis catus). Journal of Heredity 94(1):95–106. https://doi.org/10.1093/JHERED/ESG008

Prinz M, Lessig R (2014) Forensic DNA analysis. In: Handbook of forensic medicine. Wiley-Blackwell, Hoboken. https://doi.org/10.1002/9781118570654.ch63

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74 (12):5463–5467. https://doi.org/10.1073/PNAS.74.12.5463

Schneider PM (n.d.) From blood and forensic specimens. Methods 98

Slatko BE, Gardner AF, Ausubel FM (2018) Overview of next-generation sequencing technologies. Curr Protoc Mol Biol 122(1):e59. https://doi.org/10.1002/cpmb.59

Sullivan KM (1994) Forensic applications of DNA fingerprinting. Mol Biotechnol 1, pp 13–27

Walsh SJ (2005) Legal perceptions of forensic DNA profiling: Part I: a review of the legal literature. Forensic Sci Int 155(1):51–60. https://doi.org/10.1016/j.forsciint.2004.11.001

Weedn VW (2007) Forensic DNA typing. Mol Pathol Clin Pract. https://doi.org/10.1007/978-0-387-33227-7_44

Yang Y, Xie B, Yan J (2014) Application of next-generation sequencing technology in forensic science. Genomics Proteomics Bioinformatics 12(5):190–197. https://doi.org/10.1016/j.gpb.2014.09.001

Evaluation of the Autosomal STR Markers and Kits

7

Vikash Kumar

Contents

Abstract

DNA profiling is one of the most important and ultimate techniques for identification of the individuals, biological dispute, family history, pedigree chart, etc. DNA profiling was developed by Dr. Alec Jeffrey and was first used to assist police in identifying a suspect in a rape case. Initially, it was based on RFLP-based VNTR markers. Later on, STR markers was widely accepted by scientific community to analyze the specific region of DNA for more discriminatory power and more precise results. Forensic Science Services, England started using STR markers to analyses specific region of DNA. Later on, a CODIS system was developed for creating national DNA databases of DNA profile by FBI under NDIS program. The CODIS and FSS markers were widely accepted by other countries and forensic science laboratories and became a standardized benchmark

V. Kumar (✉)
Centurion University of Technology and Management, Bhubaneswar, Odisha, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_5

for minimum analyzing criteria of the DNA profiling. On the basis of STR markers, many organizations developed different variant of autosomal kit which includes multiple markers in one mixture to analyze different region of DNA in one process. They have also added new markers in the kit to provide more discriminatory power and probability index to compare two DNA profile. Nowadays, many advanced autosomal STR kits are commercialized and used by forensic science laboratories, which are overcoming vast area of challenges and producing more precise and accurate results. Autosomal STR multiplex kit has become a concrete tool to analyze DNA profile and plays a vital role in criminal justice system. In this chapter, the different autosomal STR kits accepted widely among scientific and forensic community and used frequently in forensic science laboratory had been described.

Keywords

DNA Profiling · STR markers · Autosomal STR markers · Autosomal STR Multiplex Kits · CODIS system · Nomenclature of Markers · Identifiler · Globalfiler · Verifiler · Powerplex 16HS · Powerplex Fusion 6C · Forensic Validation

Case Study: Narborough Murders: The Identification of Colin Pitchfork

The first forensic use of DNA profiling dates back to Narborough Murders also referred as Enderby Murders committed by Black Pad Killer – "Colin Pitchfork," who was sentenced to life imprisonment for the murder of two young girls and also was sentenced 10 years for the two rapes along with 3 years for two indecent assaults. He was also given 3 years for the conspiracy to pervert the course of justice by avoiding giving a DNA samples at Leicester crown court on 22 January 1988.

On 21 November 1983, a 15-year-old school girl Lynda Mann was raped and strangled in Narborough, Leicestershire. The body of Lynda Mann was found the next morning in the grounds of Carlton Hayes hospital next to Black Pad footpath. Semen sample recovered from Lynda Mann's body was analyzed which indicated that culprit had blood group 'A' and enzyme profile shared by only 10% of the male population. With no further leads, the investigation never progressed further.

In 31 July 1986, another 15-year-old school girl Dawn Ashworth was raped and murdered in the same area. The body of Dawn Ashworth was found 2 days later. The sample were recovered from body shows same characteristics as was found in Lynda Mann. These finding led the police to believe that both crimes were committed by the same person.

On 8 August 1986, police arrested a 17-year-old local kitchen porter named Richard Buckland. He confessed the murder of Dawn Ashworth but he denied

(continued)

the murder of the first girl Lynda Mann. The police were convinced that they arrested the culprit of both murder but denying of first murder by Richard Buckland led the police to seek help from Dr. Alec Jeffrey of Leicester University as he had developed a unique technique that can individualize two people apart called DNA profiling. After analyzing all the questioned samples and reference sample of Richard Buckland, Dr. Alec Jeffrey's findings indicate that both crimes were committed by same person but the profile did not match with Richard Buckland and the police released him on 21 November 1986.

This emphasizes the need for caution even when a suspect confesses to a crime.

With their chief suspect exonerated (another first for DNA profiling), the police undertook the world's first DNA mass intelligence screening in which all the men in the area, 5000 in all, were asked to provide DNA either as a blood sample or a saliva swab. Of these samples, only those exhibiting the same blood group and enzyme pattern as the murderer were subjected to DNA profiling. This was a major operation, not least because the profiling techniques were much more time consuming than those in use today, and took 6 months to complete.After completing the entire DNA profiling, none of the profile matched with the questioned evidence sample. The operation meets a dead end without any lead and a year passed.

In August 1987, a woman reported that she overheard a man named Ian Kelly in a pub saying that he had provided a DNA sample in place of his friend. When police questioned Ian Kelly he confessed that he took the blood test instead of his friend named Colin Pitchfork who was a local baker. On 19 September 1987, after the confession of Ian Kelly, the police arrested Colin Pitchfork who was then aged 27. The sample was taken from Colin Pitchfork and was sent to Dr. Alec Jeffrey for DNA profiling. The DNA profile of Colin Pitchfork was found to match of the semen samples recovered from the two murdered girls.

Colin Pitchfork, who was a local baker, was therefore arrested for raping and murdering two girls.

Colin Pitchfork was brought to trial on 22 January 1988 at Leicester Crown Court where he was found guilty and sentenced to life imprisonment.

Introduction

As we all know that cell is the basic unit of life and a miniature factory having necessary capabilities to sustain life. Each cell of an individual contains same genetic programming. Within the nucleus of our cells the genetic material resides known as DNA and referred as nuclear DNA. Nuclear DNA is referred as genetic blueprint

because it contains different types of genes having genetic traits necessary for genotypical character and phenotypical character. DNA is composed of nucleotides unit that are made up of three segments: a nitrogenous base, a sugar, and a phosphate group. The nitrogenous base plays a vital role to form different combination of nucleotides sequence in DNA. Humans have approximately 3 billion nucleotide position in their genomic DNA with zillions of combinations (Butler 2005).

Within the nucleus, DNA is found in a specialized structure called chromosomes. Human cell contains 23 pairs of chromosomes among which 22 pairs are autosomal chromosome and 1 pair is sex chromosome. Autosomal chromosomes contain the information for the development of the body. Remaining sex chromosomes control the development of internal and external reproductive organ.

Each chromosome contains a strand of tightly coiled DNA which is divided into small units called gene. Gene is the sequence of nucleotides having basic physical and functional unit of heredity. Each gene resides on a specific physical location referred as locus. A gene usually ranges from few thousands to tens of thousands base pairs in sizes. Genes consist of exons and introns and make up approximately 5% of human genomic DNA. Remaining noncoding region of DNA is called junk DNA.

Identification of an individual has been an issue in the scientific community as well as in forensic community for several years. With the advancement of scientific knowledge of gene, genomic sequence and different types of markers plays a vital role for identification of the individuals, biological dispute, family history and pedigree chart, etc.

In 1985, Dr. Alec Jeffrey of the University of Leicester, England, developed a unique technique known as DNA profiling. Dr. Alec Jeffrey first used DNA technology to assist police in identifying a suspect in a rape case. DNA profiling is basically involved RFLP/RAPD or AFLP. Although 99.9% of human DNA sequences are the same but approximately 0.1% DNA is different in every individual from each other. This 0.1% DNA is the basic concept of DNA fingerprinting that includes repetitive sequence of STR and VNTR. In court, an expert presenting DNA-based evidence talks about probabilities of a match between two samples. They do not state that that it is a direct match or not match. This fascinating technique is often explored and represented as a simple, exact, and infallible method of finding criminals and convicting the culprit to bring the justice. However, determining the match between a suspect and a crime scene is a complicated process that relies on probability of matching the profile.

Autosomal STR Marker

DNA profiling for human identity testing is performed using markers on the autosomal chromosomes and gender determination is done with markers on the sex chromosomes. From a forensic aspect there is very little rationale in analyzing the 99.9% of human DNA that is common between individuals (Goodwin et al. 2012). Fortunately, the variable regions residing in DNA vary between individuals who

become the center of attention for forensic community. Markers are exclusive small region of DNA which shows sequence and length polymorphism in different individuals referred as easily identifiable traits.

A **short tandem repeat** (STR) is a unique tool that analyze and evaluate specific regions found in DNA when a pattern of two or more nucleotides are repeated and the repeated sequences are directly adjacent to each other. The repetitive pattern of nucleotide can range in length from 2–7 bp (Butler 2005). These repetitive patterns are highly polymorphic, which may be detected using PCR. Fluorescence detection by electrophoresis separation of the alleles of STR loci is distinguished from one another by the numbers of copies of the repeat sequence contained within the amplified region.

With advancements in molecular biology and genetics techniques it is now possible to analyze any region within the 3.2 billion bases that make up the human genome. DNA loci that are to be used for forensic genetics should have some key properties; they should ideally:

- Be highly polymorphic (varying widely between individuals)
- Be easy and cheap to characterize
- Be easy to interpret and compare between laboratories
- Be not under any selective pressure
- Have a low mutation rate

The markers are capable of revealing the genetic variation at DNA sequence level and enable to distinguish two individuals from each other to discriminate between one individual's DNA profile to another. The characteristics of the markers are following (Butler 2018):

- Inexpensive to develop and apply
- Unaffected by environmental and developmental variation
- Highly robust and repeatable across different tissue types and different labs
- Polymorphic, i.e., reveal high levels of allelic variability
- Codominant in its expression

The STR markers are widely used in forensic DNA typing for human identification and forensic DNA analysis. There are certain criteria for autosomal STR marker as follows (Butler 2011):

- High discriminating power with observed heterozygosity 70%
- Separate (or widely spaced) chromosomal locations to ensure that closely linked loci are not chosen
- Robustness and reproducibility of results when multiplexed with other markers
- Low stutter characteristics
- Low mutation rate
- Predicted length of alleles that fall in the range of 90–500 bp with smaller sizes better suited for analysis of degraded DNA samples

Terminology

- **Locus or Loci**: The specific location of a gene or DNA sequence or position on a chromosome.
- **Alleles:** One of a number of alternative forms of the same gene or same genetic locus.
- **Polymorphism:** The occurrence of more than one form or morph.
- **Heterozygous:** If the two alleles at a genetic locus on homologous chromosomes are different they are termed as heterozygous.
- **Homozygous**: If the alleles are identical at a particular locus, they are termed homozygous.

Nomenclature for DNA Markers

The nomenclature for DNA markers is fairly straightforward. If a marker is part of a gene or falls within a gene, the gene name is used in the designation. For example, the short tandem repeat (STR) marker **TH01** is from the human tyrosine hydroxylase gene located on chromosome 11. The "**01"** portion of **TH01** comes from the fact that the repeat region in question is located within intron **1** of the tyrosine hydroxylase gene (Butler and Hill 2012). Sometimes the prefix HUM is included at the beginning of a locus name to indicate that it is from the human genome. Thus, the STR locus TH01 would be correctly listed as HUMTH01. Nomenclature of the DNA markers that fall outside of gene regions are designated on the basis of the chromosomal positions, chromosome number, their copy sequence, and physical location of the marker found on a particular chromosome (Butler 2012). On the basis of these criteria the nomenclature of the markers may be designated as follows:

- "D" stands for DNA by their chromosomal position.
- The next character refers to the chromosome number 5 for chromosome 5 and Y for the Y chromosome.
- The "S" refers to the fact that the DNA marker is a single-copy sequence.
- The final number indicates the order in which the marker was discovered and categorized for a particular chromosome.
  Ex: D16S539
  - D – DNA
  - 16 – Chromosome 16
  - S – Single-copy sequence
  - 539 – 539th locus described on chromosome 16

Types of Autosomal STR Marker

On the basis of participation number of nucleotides, there are generally four types of autosomal STR marker such as:

Table 1 Representation of tandem nucleotide repeats (AGAT or GATA motif is the most common motif for STR loci used by forensic scientist)

S. No.	Types	Representation
1.	Mononucleotide repeats	A, C
2.	Dinucleotide repeats	AC, AG, AT, CG
3.	Trinucleotide repeats	AAC, AAG, AAT, ACG, ACT, ACC, AGC, AGG, ATC, CCG
4.	Tetranucleotide repeats	AAAC, AAAG, AAAT, AACG, AACC, AACT, AAGC, AAGG, AAGT, AATC, AATG, AATT, ACAG, ACAT, ACCC, ACCG, ACCT, ACGC, ACGG, ACGT, ACTC, ACTG, AGAT, AGCC, AGCG, AGCT, AGGC, AGGG, ATCC, ATCG, ATGC, CCCG, CCGG

1. **Mononucleotide repeats**: When one nucleotide is repeated and repeated sequences are directly adjacent to each other.
2. **Dinucleotide repeats**: When two nucleotides are repeated and repeated sequences are directly adjacent to each other.
3. **Trinucleotide repeats**: When three nucleotides are repeated and repeated sequences are directly adjacent to each other.
4. **Tetranucleotide repeats**: When four nucleotides are repeated and repeated sequences are directly adjacent to each other.

The most popular STR repeats among different types of nucleotides repeats which drew attention of forensic community are tetranucleotide repeats. Tetranucleotide repeats are choice of STR types frequently used in different STR kits. Advance autosomal STR kits also contain trinucleotide and tetranucleotide repeats (Table 1).

A narrow allele size range is more suitable to reduce allelic dropout from preferential amplification of smaller alleles and also permits multiplexing. The capability of generating small PCR product sizes benefits the recovery of information from degraded DNA specimens and reduced stutter product formation compared to dinucleotide repeats that benefit the interpretation of sample mixtures. For DNA typing markers to be effective across a wide number of jurisdictions, a common set of standardized autosomal markers must be used (Butler and Hill 2012).

On the basis of repeating unit it is further classified into four types as follow (Peter et al. 2020):

1. **Simple repeats**: Units of identical length and sequence are repeated.
2. **Compound repeats:** Comprise two or more adjacent simple repeats.
3. **Complex repeats**: Several repeat blocks of variable unit length as well as variable intervening sequence.
4. **Microvariant**: When the tandem repeat ends with incomplete repeating sequence.

CODIS Markers

The commonly used STR loci today were initially characterized and developed either in the laboratory of Dr. Thomas Caskey at the Baylor College of Medicine or at the Forensic Science Service (FSS) in England. The Promega Corporation (Madison, WI) initially commercialized many of the Caskey markers, while Applied Biosystems (Foster City, CA) incorporated the FSS STR loci and also developed some new markers.

The UK started using STR loci way before the USA with the help of Forensic Science Service. After utilization of STR by the UK, USA started a DNA project in early 1996 and in 1997 specific STRs were selected by FBI and established core STR loci with the help of NIST, and a national DNA database was created known as CODIS (Combined DNA Index System), a computer software program that helps law enforcement agency to compare DNA profile digitally with the help of 13 autosomal STR markers and one sex marker (Hares 2012). It is also referred as CODIS13.

In 1994, US Congress passed an act called DNA identification act to create a national database of DNA profile. In 1998, FBI created the National DNA Index System referred as NDIS program (Christian et al. 2001), a part of CODIS to be implemented at national level to contain the DNA profiles submitted by local and state law enforcement agency and also by different forensic laboratories. As per a report provided by FBI, till 2013 CODIS contains over 10 million DNA profiles of different category which included to form a DNA database (CODIS Operating policies and procedures manual 2009).

DNA profiles were recorded in different category of indexes which were included in CODIS system as database to generate investigative leads in crimes where biological evidence is recovered from the crime scene to find the victim, suspect, or culprit. CODIS uses computer software to search these indexes for matching the DNA profiles whenever required by law enforcement agency. The searching and comparing the DNA profile is automated by computer software, but can also be manually analyzed by users of respective authority to analyze the different probability required by law enforcement agency to get desirable perspective. The different category of CODIS index is represented as:

- **Forensic index**: The forensic index contains all the evidence profile created from biological evidence collected from crime scene.
- **Convicted offender index**: The convicted offender index contains DNA profile and data of the culprit who has been convicted of any crime.
- **Arrestee index**: The arrestee index contains DNA profile and data of the suspect who has been arrested but not convicted.
- **Missing or unidentified index**: The missing or unidentified index contains DNA profile and data of the individuals who went missing or unidentified.

The DNA profile of different individuals and species are stored in CODIS on the basis of core STR markers and each DNA profile is given a tagline to identify an

individual. Once any DNA profile is generated and analyzed with CODIS for a potential match, the respective authority is responsible to identify the individual with index category and obtain additional information about the individual such as name, number, address, case details, previous criminal records (if any), previous felony or conviction (if any), and any other information required. CODIS does not store any criminal history information, any case information, social information, or social numbers of any individual.

The 13 autosomal STR markers used in CODIS are:

1. **TPOX**: It is a simple repeat of $(AATG)_n$ found on 10th intron of human thyroid peroxidase gene of 2nd chromosome. The mutation rate of TPOX STR marker is 0.01%. The physical location of TPOX is 1.436 Mb on chromosome 2. It was discovered and published by Anker et al. in 1991.
2. **D3S1358**: It is a compound repeat of tetrameric STR found on 3rd chromosome. The mutation rate of D3S1358 is 0.12%. It was discovered by Schmidt et al. in 1993.
3. **FGA**: It is a compound repeat STR marker of CTTT/TTCG found in the 3rd intron of human alpha fibrinogen gene of 4th chromosome. The mutation rate of FGA marker is 0.28–0.30%. The physical location of FGA STR marker is 156.086 Mb on chromosome 4 with microvariants for some alleles. It was discovered and published by Milis et al. in 1992.
4. **D5S818**: It is a simple repeat of ATCT tetrameric STR found on the 5th chromosome. The mutation rate of D5S818 is 0.12%.
5. **CSF1PO**: It is a simple repeat of ATCT tetrameric STR found in 6th intron of c-fma proto-oncogene of 5th chromosome. The mutation rate of CSF1PO is 0.16%. It was discovered by Hammond et al. in 1994.
6. **D7S820**: It is a simple repeat of TATC tetrameric STR found on 7th chromosome. The mutation rate of D7S820 is 0.10%.
7. **D8S1179**: It is a compound repeat of tetrameric STR found 8th chromosome. The mutation rate of D8S1179 is 0.13%.
8. **TH01**: TH01 STR marker is also referred as TC11 OR HUMTH01. TH01 is simple repeat of tetrameric STR found in 1st intron of tyrosine hydroxylase gene on chromosome 11 with physical location 2.156 Mb. It is a tetrameric STR of repeated unit of AATG with tri-allelic pattern. The 3rd alleles of TH01 in some cases may found microvariant (ATG) in nature. Mutation rate of TH01 is 0.01%. It was discovered by Polymeropoulos et al. in 1991.
9. **VWA**: It is a compound repeat of tetrameric STR found in 40th intron of 12th chromosome with physical location 5.963 Mb and responsible for Von Willerbrand factor. The mutation rate of VWA is 0.17%. It was discovered by Kimpton et al. in 1992.
10. **D13S317**: It is a simple repeat of TATC tetrameric STR marker found on 13th chromosome. The mutation rate of D13S317 is 0.15%.
11. **D16S539**: It is a simple repeat of GATA tetrameric STR marker found on 16th chromosome. The mutation rate of D16S539 is 0.11%.

12. **D18S51**: It is a simple repeat of AGAA tetrameric STR marker found on 18th chromosome. The mutation rate of D18S51 is 0.25%.
13. **D21S11**: It is a complex repeat of tetrameric STR marker found on 21th chromosome. It also shows microvariant in some alleles. The mutation rate of D21S11 is 0.21%.

Apart from 13 core STR markers, FBI announced additional 7 STR markers in 2015 to be introduced in CODIS program that become effective from 1 January 2017 and known as CODIS20. CODIS 20 comprises of seven additional STR loci along with 13 core STR loci. The additional STR loci are:

1. D1S1656: It is a compound repeat of tetrameric STR found on 1st chromosome. It has microvariant value of x.3 due to insertion of TGA.
2. D2S441: It is a simple repeat of TCTA tetrameric STR marker found on 2nd chromosome. Some microvariant alleles of x.3 have been also observed.
3. D2S1338: It is a compound repeat of tetrameric STR marker found on 2nd chromosome.
4. D10S1248: It is a simple repeat of GGAA tetrameric STR marker found on 10th chromosome.
5. D12S391: It is a compound repeat of tetrameric STR marker found on 12th chromosome. It is very highly polymorphic STR marker having microvariant of x.2 for some alleles.
6. D19S433: It is a complex repeat of tetrameric STR marker found on 19th chromosome.
7. D22S1045: It is a simple repeat of ATT trimeric STR marker found on 22nd chromosome.

Autosomal STR Multiplex Kits

Designing primer, ladders, or optimization PCR multiplex is hard and problematic to most of the laboratories due to lack of time, lack of proper resources, as well as high costing equipment to synthesize these materials and monitor quality control of the product. To overcome this issue, many commercialized organization developed a ready-made kit based on different core autosomal STR loci with other required materials and exported to respective agencies. These kits are referred as "Autosomal STR multiplex kit." It improved the accuracy of DNA profiling as well as probability of matching two DNA samples.

Autosomal STR multiplex kits have been widely accepted by forensic community to generate DNA profiles and compare with the help of selected autosomal STR marker. These kits have been routinely used in all forensic laboratories and private detective agencies in human identity testing for various purposes to support the criminal justice system.

Autosomal STR multiplex kits increase discrimination powers along with increasing the number of loci. All the loci are combined in one kit to produce analysis of multiple STR loci at one time. It also reduces the time of analysis and produces result faster than conventional method.

In commercial autosomal STR multiplex kits there are usually five components included for DNA profiling such as:

1. **Primer**: It contains PCR primer mixture of oligonucleotides labeled with dye to amplify desired STR loci.
2. **Buffer**: It contains deoxynucleotide triphosphate, MgCl2, and other desired respective PCR buffer.
3. **Polymerase**: It contains DNA polymerase, the molecular precursor to synthesize the targeted DNA.
4. **Ladder**: It contains allelic ladder with respect to STR loci alleles.
5. **Control sample**: It contains positive DNA control sample to verify if kit is working properly and with standard protocol (Table 2).

Earlier STR Multiplex Kits

STR multiplex kit contains number of different STR along with specific loci used to analyze genome. The first autosomal STR multiplex kit is developed by forensic science services comprised of quadraplex STR loci, also referred as "first generation multiplex." The probability of matching two DNA profiles by first generation multiplex (FGM) is 1 in 10^4. The first commercial autosomal STR multiplex kit was developed by Promega Corporation and become available in 1994. This kit was referred by CTT triplex and had a matching probability of approx. 1 in 500. Promega and Applied Biosystem is the two frontier organization which started commercial STR multiplex kit with different variant of STR multiplex kit that contains specific STR markers. Both Promega and Applied Biosystem have commercialized many multiplex kits since 2000 that facilitate co-amplification of all 13 STR markers in one single reaction. Apart from that they commercialized different variant with extra STR markers added in kit.

During starting period of adaptation of autosomal STR multiplex kit three cases were ruled out by the court of law where it was stated that DNA result cannot be admissible in court as evidence because primer sequence and validation of commercialized autosomal STR kit is not public information. After statement by the court, question mark arises on the credibility and reliability of the commercialized product of autosomal STR kit by both Promega and Applied Biosystem. After the statement made by court, the Promega Corporation decided to publish the sequence of their autosomal STR multiplex kit on 24 July to public and obtained several patients of autosomal STR multiplex kits. Meanwhile Applied Biosystem has refused to make the sequence publicly. But since 2002, Applied Biosystem provided the sequence of primer under protective court order.

Table 2 Variant of autosomal STR marker multiplex kit commercialized by different organization

Name of the kit	Organization	Published/Released date
AmpFlSTR Blue	Applied Biosystem	Oct 1996
AmpFlSTR Green	Applied Biosystem	Jan 1997
CTTv	Promega	Jan 1997
FFFL	Promega	Jan 1997
GammaSTR	Promega	Jan 1997
Powerplex 1.1	Promega	Jan 1997
Powerplex 1.2	Promega	Sept 1998
AmpFlSTR Profiler	Applied Biosystem	May 1997
AmpFlSTR Profiler Plus	Applied Biosystem	Dec 1997
AmpFlSTR COfiler	Applied Biosystem	May 1998
AmpFlSTR SGM Plus	Applied Biosystem	Feb 1999
Powerplex 2.1 (For Hitachi FMBIO users)	Promega	June 1999
Powerplex 16	Promega	May 2000
Powerplex 16 BIO (For Hitachi FMBIO users)	Promega	May 2001
AmpFlSTR Identifiler	Applied Biosystem	July 2001
AmpFlSTR profiler plus ID (Extra unlabeled D8-R primer)	Applied Biosystem	Sept 2001
Powerplex ES	Promega	Mar 2002
AmpFlSTR SEfiler	Applied Biosystem	Sept 2002
AmpFlSTR MiniFiler	Applied Biosystem	Mar 2007
AmpFlSTR SEfiler Plus	Applied Biosystem	Nov 2007
AmpFlSTR sinofiler	Applied Biosystem	Mar 2008
Powerplex 16HS	Promega	Mar 2009
Powerplex ESX 16 & ESX 17	Promega	Sept 2009
Powerplex ESI 16 & ESI 17	Promega	Sept 2009
AmpFlSTR Identifiler Direct	Applied Biosystem	Nov 2009
AmpFlSTR Identifiler Plus	Applied Biosystem	Jan 2010
AmpFlSTR NGM	Applied Biosystem	Jan 2010
Investigator ESSplex	QIAGEN	April 2010
Investigator decaplex SE	QIAGEN	April 2010

(continued)

Table 2 (continued)

Name of the kit	Organization	Published/Released date
Investigator triplex AFS QS	QIAGEN	April 2010
Investigator triplex DSF	QIAGEN	April 2010
Investigator ID plex	QIAGEN	Aug 2010
Investigator HDplex	QIAGEN	Sept 2010
Investigator hexaplex ESS	QIAGEN	Sept 2010
Investigator Nanoplex SE	QIAGEN	Sept 2010
Investigator ESSplex SE	QIAGEN	Oct 2010
AmpFlSTR NGM Select	Applied Biosystem	Dec 2010
Powerplex 18D	Promega	Feb 2010

Identifiler Kit

Identifiler Kit is Multiple PCR STR amplification Kit for the amplification of multiple autosomal STR markers. Identifiler kit is first developed and commercialized by Applied Biosystem as "AmpFLSTR Identifiler™" in July 2001. Identifiler kit was based on 13 CODIS STR markers and sex-typing marker. Apart from 13 CODIS STR markers, 2 additional autosomal STR markers were included in a single tube which was developed in collaboration with Forensic Science Service (FSS).

The Identifiler Kit can amplify 15 autosomal STR loci and amelogenin gender determining marker in a single PCR amplification. Due to 16-Locus multiplex, Identifiler Kit gain advantages over previous and other kits that were accepted by forensic community. It reduces the sample preparation time, amplification time, laborious process, and analysis time in half as it completed the amplification of the entire locus in single multiplex PCR followed by one capillary electrophoresis.

All 15 autosomal STR markers and one sex marker is included in one single tube. The autosomal STR markers are CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, D2S1338, and D19S433.

Identifiler Kit as stated by Thermofisher is the single most discriminating and widely used STR-based kit for human identification with two technologies. The first one is five-dye fluorescent detection system. The spectral detection range was expanded with the help of five dyes set which enable high throughput analysis and maintain small amplicons. In five-dye detection system four different dyes are used to label the PCR product obtained by Identifiler kit STR markers. These four dyes are 6FAM™, VIC™, NED™, and PET™. The fifth dye is used to standardize the internal size to correlate the electrophoretic mobilities. The fifth dye is LIZ™ (User Guide, AmpFlSTR identifiler PCR amplification kit 2018).

So, in Identifiler total five dyes are used unlike four traditional dyes that have been used previously in AmpFlSTR or Powerplex. In Powerplex kits four dyes were used and among those four dyes, three dyes were used to label PCR product, i.e.,

Table 3 15 autosomal markers of Identifiler Kit with respect to labeling dye for each marker (Product Bulletin, AmpFlSTR Identifiler PCR amplification kit)

S. No.	Label dye	STR marker	Chromosome location	Allelic variation
1.	6-FAM	CSF1PO	5q33.3-34	6–15
2.		D7S820	7q11.21-22	6–15
3.		D8S1179	8	8–19
4.		D21S11	21q11.2-q21	24, 24.2, 25–28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36–38
5.	VIC	D2S1338	2q35-37.1	15–28
6.		D3S1358	3p	12–19
7.		TH01	11p15.5	4–9, 9.3, 10, 11, 13.3
8.		D13S317	13q22-31	8–15
9.		D16S539	16q24-qter	5, 8–15
10.	NED	vWA	12p12-pter	11–24
11.		TPOX	2p23-2per	6–13
12.		D18S51	18q21.3	7, 9, 10, 10.2, 11–13, 13.2, 14, 14.2, 15–27
13.		D19S433	19q12-13.1	9–12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2
14.	PET	FGA	4q28	17–26, 26.2, 27–30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2
15.		D5S818	5q21-31	7–16

5-FAM, JOE, and JOE, and one dye was used to standardize the internal size, i.e., ROX (Table 3).

The data generated by Identifiler Kit was widely accepted and approved by forensic community such as FBI, Interpol, and European Network of Forensic Science Institutes (ENFSI) (Fig. 1).

Identifiler expanded the spectral range to 660 nm. The expanded range allows maximum color separation with minimal spectral overlap.

The second technology involves mobility modifying non-nucleotide linkers. The mobility modifying is composed of hexaethyleneoxide (HEO). The unique properties of mobility modifying non-nucleotide linker enables size shift to retain original primer sequence to avoid discovering primer binding mutation while STR allele size ranges may be altered during electrophoresis. These non-nucleotide linkers prevent overlapping of the size range of amplified alleles by providing a shift of approximately 2.5 nucleotides with each HEO unit.

Non-nucleotide linkers are placed between the primers and the fluorescent dye during oligonucleotide synthesis. The non-nucleotide linkers are used in primer synthesis for some of the loci in Identifiler such as: CSF1PO, D13S317, D16S539, D2S1338, and TPOX (Rabelo et al. 2015).

Applied Biosystem released two more variants of the Identifiler kit with different enhancement in each version as follows:

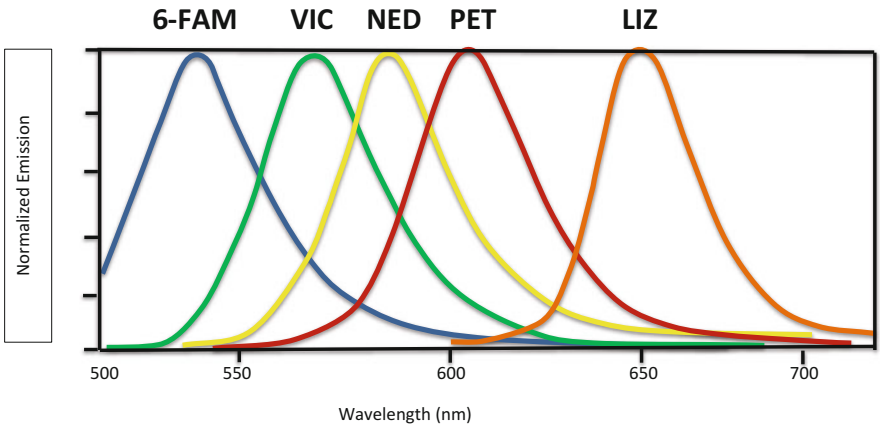


Fig. 1 Emission spectra of five dyes used by Identifiler kit

"AmpFLSTR Identifiler™ Direct"

After 8 year of releasing of Identifiler Kit, Applied Biosystem released Identifiler Direct kit as AmpFLSTR Identifiler™ Direct in Nov 2009, also referred as next generation STR Kit. The main advantage of Identifiler direct kit over Identifiler kit is that Identifiler direct kit allows direct amplification of single source. Identifiler direct kit provides prep n-Go buffer for extraction of DNA. DNA can be extracted from prep n-Go buffer without the need for sample purification (User Guide, AmpFlSTR identifiler Direct PCR amplification kit 2018).

"AmpFLSTR Identifiler™ Plus"

After Identifiler direct, Applied Biosystem released a new variant of Identifiler kit called as AmpFLSTR Identifiler™ Plus in Jan 2010. Identifiler plus kit also referred as next generation STR kit which provides better sensitivity and better robustness than previous Identifiler kit. Identifiler kit uses modified PCR cycling condition for enhanced sensitivity. It also included a buffer mixture which gives better result with inhibited samples and provides cleaner electrophoretic background (User Guide, AmpFlSTR identifiler Plus PCR amplification kit 2015).

The Identifiler kit is compatible with different genetic analyzer such as 310 genetic analyzer, 3100-avant genetic analyzer, 3130xl genetic analyzer, Gene Amp 9600, Gene Amp 9700, 3500 genetic analyzer, 3500xl genetic analyzer, and 3730 DNA analyzer, although optimization of the different process is required as per the standard protocol provided by Thermo fisher, Applied Biosystem.

Powerplex 16HS Kit

Powerplex 16HS system is a multiple STR kit which is used for Forensic DNA typing developed and published by Promega Corporation in March 2009. Powerplex 16HS system is an updated version of the Powerplex 16 kit (Julio et al. 2013). The primer and dyes of Powerplex 16 HS kit is same as Powerplex 16 kit, but in

Powerplex 16HS kit enhanced buffer system is provided that includes hot-start taq DNA polymerase that ensures reproducible result with low quantity of the DNA and robust performance.

Powerplex 16HS kit contains 16 loci in a single tube and amplifies all 16 loci simultaneously in one processing. Among those 16 loci it includes all 13 CODIS STR markers along with amelogenin for gender determination and it also contains two new STR markers such as "Penta E and Penta D." Penta E and Penta D is highly discriminating pentanucleotide STR markers having low stutter product (Jebor et al. 2015).

- Penta D: It is a pentanucleotide repeat STR marker found on 21st chromosome. It is situated about 25Mbp away from D21S11.
- Penta E: It is a pentanucleotide repeat STR marker found on long arm of 15th chromosome. It have very low stutter product with alleles ranging from 5–32 repeats.

Powerplex 16HS kit contains three color detection label dye for 16 loci and one label dye for internal lane standard. The three labeling dyes for STR markers are fluorescein (FL), carboxy-tetramethylrhodamine (TMR), and 6-carboxy-4′, 5-′-dichloro-2′, 7′-dimethoxy-fluorescein (JOE). One labeling dye for internal lane standard is carboxy-X-rhodamine (CXR).

Powerplex 16HS kit can give reproducible results with 0.5–1 ng DNA but some experiment shows that it can also give good results with lesser than 0.5 ng DNA also. Penta D and Penta E allow high discriminating power for STR result with low quantity of DNA.

Powerplex 16HS kit has been engineered to enable STR amplification in the presence of different PCR inhibitors such as heme, humic acid, tannic acid, and other inhibitors. This advantage allows biological samples to be directly amplified without purification (Technical Manual, Powerplex 16HS system 2016).

The Powerplex 16HS kit is compatible with different genetic analyzer such as ABI PRISM 310-, 3100-, and 3100-avant genetic analyzer and applied biosystem 3130, 3130xl, 3500, and 3500xl genetic analyzer (Thatch et al. 2012), although optimization of the different process is required as per the standard protocol provided by Promega Corporation.

New Generation Kit

Forensic DNA typing has been constantly evolving according to the different circumstances and challenges faced by forensic community and scientific community. Wide varieties of commercial STR kits have been designed to meet various needs of a forensic lab as STRs become the markers of choice for forensic DNA typing. Different challenges and requirement drive the scientific and research laboratories as well as commercial organization to overcome the other challenges in earlier generation autosomal STR multiplex kits and meet the

Table 4 15 autosomal markers of Powerplex 16 HS autosomal STR multiplex kit with respect to labeling dye for each marker and allelic variation of each STR loci in kit

S. No.	Label dye	STR marker	Chromosome location	Size of allelic ladder bases	Allelic ladder variation
1.	FL	D3S1358	3p	115–147	12–20
2.		TH01	11p15.5	156–195	4–9, 9.3, 10–11, 13.3
3.		Penta E	15q	379–474	5–24
4.		D18S51	18q21.3	290–366	8–10, 10.2, 11–13, 13.2, 14–27
5.		D21S11	21q11-21q21	203–259	24, 24.2, 25, 25.2, 26–28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36–38
6.	TMR	TPOX	2p24-2pter	262–290	6–13
7.		FGA	4q28	322–444	16–18, 18.2, 19, 19.2, 20, 20.2, 21, 21.2, 22, 22.2, 23, 23.2, 24, 24.2, 25, 25.2, 26–30, 31.2, 43.2, 44.2, 45.2, 46.2
8.		D8S1179	8q24.13	203–247	7–18
9.		vWA	12p13.31	123–171	10–22
10.	JOE	CSF1PO	5q33.3-34	321–357	6–15
11.		D5S818	5q23.3-32	119–155	7–16
12.		D7S820	7q11.21-22	215–247	6–14
13.		D13S317	13q22-q31	176–208	7–15
14.		D16S539	16q24.1	264–304	5, 8–15
15.		Penta D	21q	376–449	2.2, 3.2, 7–17

criteria such as more accuracy, high robustness, simpler/automated processing, speedy process, sensitivity, and other parameters by different forensic science laboratories as well as others organization to produce new generation kits. These improvements enhanced the ability of forensic science laboratories to obtain a DNA profile from more challenging samples which were facing problem in accuracy and sensitivity in earlier generation kits. The new generation STR multiplex kits provide more accuracy with higher robustness and greater discriminatory power.

These improvements and characteristics provide a higher reliability and credibility for differentiating trace DNA alleles from noise-related artifacts. The CODIS increased their core loci from 13 STR markers to 20 STR markers.

The 20 CODIS system become effective from January 2017. All new generation STR kits contain 20 CODIS marker as base markers. Different commercial organizations incorporate these 20 CODIS STR markers as well as some other markers developed by them to provide better discriminatory power. With these improvements, many STR kits have been released by commercial organizations which provide high discriminatory power, robustness, and better performance in harsh conditions.

The 13 autosomal STR markers of CODIS are (Gettings et al. 2015):

1. **TPOX**: It is a simple repeat of $(AATG)_n$ found on 10th intron of human thyroid peroxidase gene of 2nd chromosome. The mutation rate of TPOX STR marker is 0.01%. The physical location of TPOX is 1.436 Mb on chromosome 2. It was discovered and published by Anker et al. in 1991.
2. **D3S1358**: It is a compound repeat of tetrameric STR found on 3rd chromosome. The mutation rate of D3S1358 is 0.12%. It was discovered by Schmidt et al. in 1993.
3. **FGA**: It is a compound repeat STR marker OF CTTT/TTCG found in the 3rd intron of human alpha fibrinogen gene of 4th chromosome. The mutation rate of FGA marker is 0.28–0.30%. The physical location of FGA STR marker is 156.086 Mb on chromosome 4 with microvariants for some alleles. It was discovered and published by Milis et al. in 1992.
4. **D5S818**: It is a simple repeat of ATCT tetrameric STR found on 5th chromosome. The mutation rate of D5S818 is 0.12%.
5. **CSF1PO**: It is a simple repeat of ATCT tetrameric STR found in 6th intron of c-fma proto-oncogene of 5th chromosome. The mutation rate of CSF1PO is 0.16%. It was discovered by Hammond et al. in 1994.
6. **D7S820**: It is a a simple repeat of TATC tetrameric STR found on 7th chromosome. The mutation rate of D7S820 is 0.10%.
7. **D8S1179**: It is a compound repeat of tetrameric STR found 8th chromosome. The mutation rate of D8S1179 is 0.13%.
8. **TH01**: TH01 STR marker is also referred as TC11 OR HUMTH01. TH01 is simple repeat of tetrameric STR found in 1st intron of tyrosine hydroxylase gene on chromosome 11 with physical location 2.156 Mb. It is a tetrameric STR of repeated unit of AATG with tri-allelic pattern. The 3rd alleles of TH01 in some cases may found in microvariant (ATG) in nature. Mutation rate of TH01 is 0.01%. It was discovered by Polymeropoulos et al. in 1991.
9. **VWA**: It is a compound repeat of tetrameric STR found in 40th intron of 12th chromosome with physical location 5.963 Mb and responsible for Von Willerbrand factor. The mutation rate of VWA is 0.17%. It was discovered by Kimpton et al. in 1992.
10. **D13S317**: It is a simple repeat of TATC tetrameric STR marker found on 13th chromosome. The mutation rate of D13S317 is 0.15%.
11. **D16S539**: It is a simple repeat of GATA tetrameric STR marker found on 16th chromosome. The mutation rate of D16S539 is 0.11%.
12. **D18S51**: It is a simple repeat of AGAA tetrameric STR marker found on 18th chromosome. The mutation rate of D18S51 is 0.25%.
13. **D21S11**: It is a complex repeat of tetrameric STR marker found on 21th chromosome. It also shows microvariant in some alleles. The mutation rate of D21S11 is 0.21%.

Apart from CODIS 13 STR marker, other seven STR markers which are part of both the CODIS 20 and ESS are as follows:

1. **D1S1656**: It is a compound repeat of tetrameric STR found on 1st chromosome. It has microvariant value of x.3 due to insertion of TGA.
2. **D2S441**: It is a simple repeat of TCTA tetrameric STR marker found on 2nd chromosome. Some microvariant alleles of x.3 have been also observed.
3. **D2S1338**: It is a compound repeat of tetrameric STR marker found on 2nd chromosome.
4. **D10S1248**: It is a simple repeat of GGAA tetrameric STR marker found on 10th chromosome.
5. **D12S391**: It is a compound repeat of tetrameric STR marker found on 12th chromosome. It is very highly polymorphic STR marker having microvariant of x.2 for some alleles.
6. **D19S433**: It is a complex repeat of tetrameric STR marker found on 19th chromosome.
7. **D22S1045**: It is a simple repeat of ATT trimeric STR marker found on 22nd chromosome.

Global Filer Kit

DNA profiling by targeting STR markers present in human genome proves an ultimate techniques to discriminate between two individuals for human identification in forensic cases. With the ever-growing needs and challenging samples for DNA profiling, different new generation STR kits were commercialized by different organization. Globalfiler kit is PCR amplification kit having comprised of 24 loci (Ludeman et al. 2018). Globalfiler kit was published in 2012 by Applied Biosystem and received approval soon after its release by FBI. The 24 loci comprised of 21 autosomal STR markers and 3 sex determination markers. Among those 21 autosomal STR marker, 13 CODIS STR marker were included and 7 markers were included from European standard set of loci (ESSL) and a highly discriminating locus was also added, i.e., SE33 locus. Sex determination markers comprised of Y-STR locus, Y-Indel, and amelogenin. Globalfiler PCR amplification kit allows for amplification of all 21 autosomal STR loci and 3 sex determination loci in one single reaction.

Globalfiler kit is the 1st kit having six-dye technologies with 24 loci. Thermo Fisher Scientific (Applied Biosystem) is the only company to receive FBI approval for the six-dye technology. The 21 autosomal along with 3 sex determination marker provides maximum information recovery and amplification of challenging degraded samples (Bogas et al. 2015). Some of the key features introduced by Thermo Fisher Scientific company in Globalfiler kit are (Gouveia et al. 2015):

- Reduced amplification time with the help of efficient protocol compared to previous generation kits and a specific enzyme included in optimized buffer system which also enables the expansion of DNA input volume which delivers maximum sensitivity for trace/low-level DNA samples
- Easy interpretation of mixture samples with the help of improved intra-color balance
- Maximize performance from degraded samples with the help of 10 mini-STR included in kit having less than 220 bp size.

- Inclusion of CODIS core loci and FSS core loci increased global data compatibility and make this the only kit which includes all recommended markers by respective organizations.
- The TPOX reverse primer has been redesigned to relocate the amplicon into the higher size range of the multiplex and optimizing marker spacing.
- Addition of eight new SNP-specific primers for the D3S1358, VWA, D16S539, AMEL, D2S441, D22S1045, and D8S1179 loci.

Globalfiler kit is a robust single-amplification STR multiplex kit that provides enhanced buffer system for enhanced sensitivity with inhibited and low concentration DNA samples (Brito et al. 2015). Non-nucleotide linkers are used in primer synthesis for the following loci: D19S433, VWA, CSF1PO, D2S441, TH01, FGA, and D12S391. Globalfiler kit satisfies the new recommendation based by FBI as well as ESS which gives them many advantages over others kits and it also included other STR markers to provide greater discrimination power to identify individuals. The seven markers from ESS were also accepted by CODIS system in 2017 in order to expand CODIS system from CODIS 13 to CODIS 20 (Hares 2012). So, in other words Globalfiler PCR amplification kit includes all 20 CODIS STR markers which were approved and become effective by 2017 and one additional loci of European STR kit. The details of all STR markers included in Globalfiler kit are given below:

- **CODIS 20 markers**: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, VWA, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 and D22S1045.

Apart from these 20 STR markers, one more marker is included, i.e., SE33 which is the part of European standard set only.

- **SE33**: It is a simple repeat of AAAG tetrameric STR marker found on 6th chromosome. It is highly polymorphic STR marker having microvariant for most of its alleles that provide greater discrimination power (Al-Snan et al. 2019). The mutation rate of this STR marker is 0.64% (Table 5).

As mentioned above, Globalfiler kit is the first kit that used six-dye technology for labeling all the 24 loci. These dyes are 6-FAM, VIC, NED, TAZ, SID, and LIZ. Among these six dyes, five dyes are used for labeling the sample and allelic ladders, i.e., 6-FAM, VIC, NED, TAZ, and SID while one dye is used for labeling and standardization the internal lane standard, i.e., LIZ (Fig. 2).

All of the24 loci of Globalfiler kit is having size lesser than 400 bp. The lesser size and allelic variation of all loci provide greater discrimination power to distinguish between two individuals. The matching probability between two individuals after analysis by Globalfiler kit is 10^{24} to 10^{28}.

The Globalfiler kit is compatible with different genetic analyzer such as SeqStudio genetic analyzer, 3500 genetic analyzer, 3500xL genetic analyzer, 3130 genetic analyzer, and 3130 xL genetic analyzer although optimization of

Table 5 24 markers of Globalfiler STR multiplex kit with respect to labeling dye for each marker and allelic variation of each STR loci in kit

S. No.	Label dye	STR marker	Chromosome location	Allelic variation
1.	6-FAM	D3S1358	3p21.31	9–20
2.		TPOX	2p23-2per	6–15
3.		CSF1PO	5q33.3-34	6–15
4.		vWA	12p12-pter	11–24
5.		D16S539	16q24.1	5, 8–15
6.	VIC	Amelogenin	X: p22.1-22.3 y:p11.2	X, Y
7.		Y Indel	Yq11.221	1,2
8.		D8S1179	8q24.13	5–19
9.		DYS391	Yq11.21	7–13
10.		D18S51	18q21.3	7, 9, 10, 10.2, 11–13, 13.2, 14, 14.2, 15–27
11.		D21S11	21q11.2-q21	24, 24.2, 25–28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36–38
12.	NED	D2S441	2p14	8–11, 11.3, 12–17
13.		FGA	4q28	13–26, 26.2, 27–30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2
14.		TH01	11p15.5	4–9, 9.3, 10, 11, 13.3
15.		D19S433	19q12-13.1	9–12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2
16.	TAZ	D5S818	5q21-31	7–18
17.		SE33	6q14	4.2, 6.3, 8, 9, 11–20, 20.2, 21, 21.2, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35, 35.2, 36, 37
18.		D7S820	7q11.21-22	6–15
19.		D13S317	13q22-31	5–16
20.		D22S1045	22q12.3	8–19
21.	SID	D1S1656	10q26.3	9–14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3
22.		D2S1338	2q35	11–28
23.		D10S1248	10q26.3	8–19
24.		D12S391	12p13.2	14–19, 19.3, 20–27

the different process is required as per the standard protocol provided by respective organization. Value obtained for the same sample can differ between instruments platforms, because of different polymer matrices and electrophoretic conditions.

Verifiler Kit

As we previously discussed the challenging samples in forensic work cases, the degrading samples in forensic area poses main threat and the circumstantial

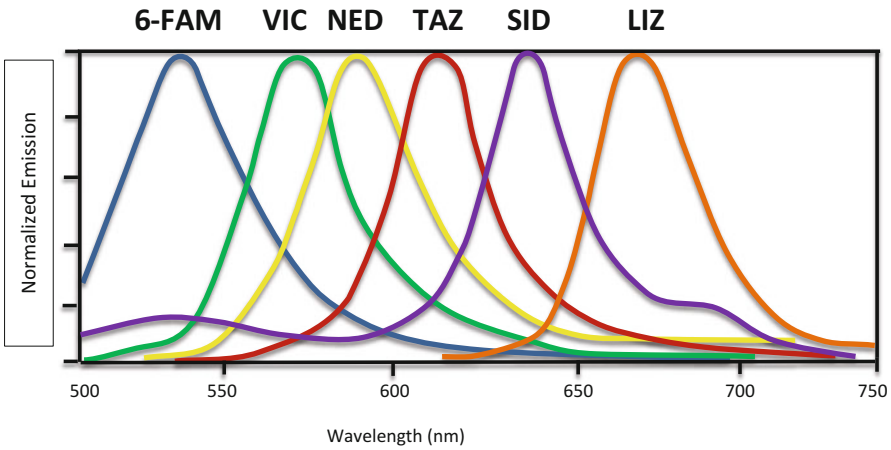


Fig. 2 Spectral graph of six dyes used by Globalfiler kit for labeling all 24 loci

evidence found from different types of environment and places with many inhibition factors in samples arise a lot of problem to get proper DNA profiling. To overcome this problem Applied Biosystem release a focused STR multiplex kit for these forensically challenged samples. Verifiler™ Plus PCR amplification kit was published and commercialized in the market in 2018. Verifiler kit is a 25-loci multiplex assay PCR kit for the amplification of single-source human genomic DNA. This kit is optimized for paternity testing and obtaining DNA profiles from single-source samples. It delivers highest discrimination power compared to previous ThermoFisher Scientific Human identification kit (User Guide, Verifiler Express PCR amplification kit 2017). Verifiler kit uses six-dye technology for the labeling of provided loci and standardization of internal lane standard. Verifiler kit provides amplification of all 25 loci in a single tube that can be amplified in a single reaction. The key features provided by Verifiler Plus PCR amplification kit includes (Pankaj and Dash 2019):

- Very high discriminatory power across global populations
- All CODIS marker and Chinese national database loci required markers
- Concordance with previous thermoFisher scientific kit loci
- 11 mini STR having amplicon size less than 250 base pairs that increased the chances of alleles detection and improved the DNA profiling from degraded samples
- Internal quality control (IQC) system that enables the distinction between inhibited and degraded samples
- Next generation master mix formulation that speeds up the workflow process with 76-minute PCR cycle time for the amplification of the loci
- Optimized buffer system that enhances the sensitivity to recover more alleles, hence recovery of more information for casework

The Verifiler Plus PCR amplification kit loci comprised of:

- **CODIS 20 markers**: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, VWA, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, and D22S1045
- **Three additional loci**: D6S1043, Penta D, and Penta E
- **One Y chromosome marker**: Y-Indel
- **One sex determining marker**: Amelogenin
- **Two internal quality control marker**: IQCS and IQCL (Table 6).

The primers of Verifiler kit have been optimized to maintain concordance with pervious STR multiplex kit published by ThermoFisher Scientific. To simplify the interpretation of the results, primers are manufactured to maximize the assay signal-to-noise ratio. Verifiler Plus PCR amplification kit provide next generation master mix formulation specifically designed for challenging samples such as touch samples, degraded samples, and samples contaminated with inhibition factors. It provides improved sensitivity and enhanced robustness against inhibition factors.

The Verifiler kit use six-dye technology for the labeling of the loci. These dyes are 6-FAM™, VIC™, TED™, TAZ™, SID™, and LIZ™. Among these six dyes, five dyes are used for labeling the samples, allelic ladders, and control, i.e., 6-FAM™, VIC™, TED™, TAZ™, and SID™ while sixth dye is used for size standard, i.e., LIZ™ (Fig. 3).

The combination of six-dye fluorescent system and use of non-nucleotide linkers allow simultaneous amplification and efficient separation of all 25 loci. Non-nucleotide linkers are placed between the primers and fluorescent dye during oligonucleotide synthesis which enable interlocus spacing which is used for primer synthesis for the following loci: D19S433, VWA, CSFAPO, D2S441, TH01 FGA, and D12S391.

All the 25 loci of Verifiler kit have size lesser than 500 bp. The lesser size and allelic variation of all loci provide greater discrimination power to distinguish between two individuals. The matching probability between two individuals after analysis by Verifiler Plus PCR amplification kit of is 10^{28} to 10^{30}.

One of the unique advantages of Verifiler Plus PCR amplification kit is internal quality control (IQC) system. The IQC system has two synthetic targets, one low molecular weight (IQCS) and one high molecular weight (IQCL) (Al Janaahi et al. 2019). The IQC system identifies the presence of inhibitors and enables the distinction between degraded and inhibited samples and act as positive control for PCR amplification in Verifiler Plus PCR amplification kit. The Verifiler Plus PCR amplification kit gives the capability to obtain maximum possible information from the most challenging samples found in forensic casework.

The Verifiler plus pcr amplification kit is compatible with different genetic analyzer such as 3500 genetic analyzer, 3500xL genetic analyzer, 3130 genetic analyzer, and 3130 xL genetic analyzer although optimization of the different process is required as per the standard protocol provided by respective organization. Value obtained for the same sample can differ between instruments platforms because of different polymer matrices and electrophoretic conditions (User Guide, Verifiler Plus PCR amplification kit 2020).

Table 6 25 markers of Verifiler Plus STR multiplex kit with respect to labeling dye for each marker and allelic variation of each STR loci in kit

S. No.	Label dye	STR marker	Chromosome location	Allelic variation
1.	6-FAM	IQCS	Synthetic marker	1, 2
2.		D3S1358	3p21.31	9–20
3.		CSF1PO	5q33.3-34	6–15
4.		D6S1043	6q15	7–25
5.		vWA	12p12-pter	11–24
6.		D16S539	16q24.1	5, 8–15
7.		IQCL	Synthetic marker	1, 2
8.	VIC	Amelogenin	X: p22.1-22.3 y:p11.2	X, Y
9.		Y Indel	Yq11.221	1,2
10.		D8S1179	8q24.13	5–19
11.		D5S818	5q21-31	7–18
12.		D18S51	18q21.3	7, 9, 10, 10.2, 11–13, 13.2, 14, 14.2, 15–27
13.		D21S11	21q11.2-q21	24, 24.2, 25–28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36–38
14.	TED	D2S441	2p14	8–11, 11.3, 12–17
15.		FGA	4q28	13–26, 26.2, 27–30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2
16.		D10S1248	10q26.3	8–19
17.		D19S433	19q12-13.1	9–12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2
18.	TAZ	D1S1656	10q26.3	9–14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18.3, 19.3, 20.3
19.		D7S820	7q11.21-22	6–15
20.		D13S317	13q22-31	5–16
21.		D22S1045	22q12.3	8–19
22.		Penta E	15q26.2	5–26
23.	SID	TPOX	2p23-2per	5–15
24.		D2S1338	2q35	11–28
25.		TH01	11p15.5	4–9, 9.3, 10–12, 13.3
26.		D12S391	12p13.2	14–19, 19.3, 20–27
27.		Penta D	21q22.3	2.2, 3.2, 5–17

Fusion 6C Kit

In forensic casework the evidence often contains a mixture of DNA from more than one individual. These mixtures can be very challenging to analyze and interpret. In most challenging cases where mixture of samples is found from crime scene, these

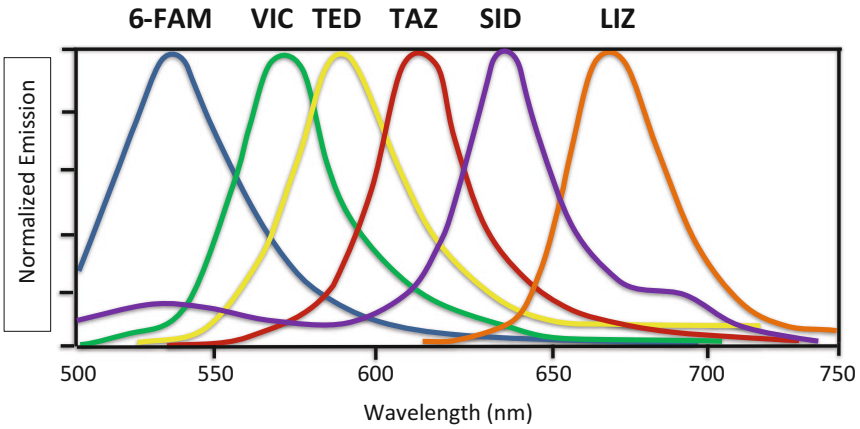


Fig. 3 Emission spectral graph of six dyes used by Verifiler Plus PCR amplification multiplex assay kit for labeling all the 25 loci

types of samples poses challenges to forensic community and scientific community to distinguish and identify the individuals.

In January 2015, Promega Corporation released a powerful STR kit "Powerplex Fusion 6C system" to resolve these types of challenge and provide a solution to forensic community and scientific community. PowerPlex fusion 6C kit amplifies 27 locus multiplex which includes 23 autosomal STR loci, 3Y-STR, and one amelogenin marker for human identification, scientific testing, and research use. The Powerplex Fusion 6C system meets the requirement of CODIS and European standard enabling to achieve the international database compatibility (Ensenberger et al. 2016). The Powerplex Fusion 6C system is manufactured in accordance with ISO 18385:2016. The Fusion 6C kit uses six-dye technology to amplify all 27-locus multiplex. The key features provided by Powerplex Fusion 6C system include:

- It reduces sample processing time by direct amplification protocols and rapid cycling capabilities
- It minimizes the need of re-amplification enhanced inhibitors tolerance and sensitivity to minimize the need of re-amplify samples for casework
- It meets the recommendation of CODIS marker and European standard set marker
- It provides highest PI value for human identification.

The 27 loci comprised of expanded CODIS core loci known as CODIS 20 along with amelogenin and DYS391 for gender discrimination. To increase more discriminatory power it also includes Penta-D, Penta-E, and SE33 loci (Shivani et al. 2019). Apart from these loci two rapidly mutation Y-STR loci are included, i.e., DYS570 and DYS576. The 27 loci are included in Fusion 6C kit as follows (Boavida et al. 2017):

- **CODIS 20**: CSF1PO, D1S1656, D2S441, D2S1338, D3S1358, D5S818, D7S820, D8S1179, FGA, TH01, TPOX, VWA, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, and D22S1045
- **Gender marker**: Amelogenin and DYS391
- **ESS markers**: Penta-D, Penta-E, and SE33
- **Rapidly mutating Y-STR marker**: DYS570 and DYS576

This extended panel of STR marker in PowerPlex Fusion 6C system is intended to satisfy both CODIS and ESS recommendations (Table 7).

The Powerplex Fusion 6C system is designed for amplification of extracted DNA and direct amplification samples. The primers included in the Powerplex Fusion 6C system have been designed to amplify human DNA. DNA samples encountered in forensic cases may contain trace or even excess nonhuman DNA along with human DNA. Hence, it is possible for some primers to recognize sites within nonhuman sources of DNA (Protocol, Use of the powerplex Fusion 6C system to amplify extracted DNA 2020). The possibility of mix source of human and nonhuman DNA may come from different circumstances such as: microbes in samples, environmental contamination, other animal biological fluids, etc. Validation studies of Powerplex Fusion 6C system by different researchers show a threshold RFU peak to determine, identify, and distinguish the RFU peak of human DNA and nonhuman DNA.

Powerplex Fusion 6C system uses six-dye technology for labeling all the loci. These dyes are FL, JOE, TMR, CXR, TOM, and WEN. These six dye hues are blue, green, yellow, red, purple, and orange, respectively. Among these six dyes, five dyes are used for labeling samples, allelic ladders, and control, i.e., FL, JOE, TMR, CXR, and TOM, while one dye is used for internal lane standard, i.e., WEN referred as WEN ILS 500 (Technical Manual, Powerplex Fusion 6C system for the use on the spectrum compact CE System 2020).

All the 27 loci of Powerplex Fusion 6C system have size lesser than 500 bp. The lesser size and allelic variation of all loci and mini-STR provide greater discrimination power to distinguish between two individuals. The additional Y-marker provides greater discriminatory power for gender determination and human identification which increased the sensitivity for forensic cases such as paternity dispute, rape cases, etc. One of the main advantages this kit provides is distinguish between two individuals from the mixture of DNA. Different validation and mixture study shows that mixture of three individuals DNA can be analyzed and determined by Powerplex Fusion 6C system. Powerplex Fusion 6C system yields good DNA profile from touch DNA samples which enable the forensic community to face most challenging sample even in harsh environment and adverse circumstances. Powerplex Fusion 6C system provides highest probability of identity (PI), hence it decreases the likelihood of adventitious match of two individuals. The matching probability between two individuals after analysis by Powerplex Fusion 6C system is 10^{31}.

The Powerplex Fusion 6C system is compatible with different genetic analyzer such as ABI prism 3100 genetic analyzer, ABI prism 3100-avant genetic analyzer

Table 7 27 markers of PowerPlex Fusion 6C STR multiplex kit with respect to labeling dye for each marker and allelic variation of each STR loci in kit (Technical Manual, Powerplex Fusion 6C system for the use on the Applied Biosystem Genetic Analyzers, 2018)

S. No.	Label dye	STR marker	Chromosome location	Repeat sequence	Allelic variation
1.	FL-6C	D1S1656	1q42	TAGA complex	9–14, 14.3, 15, 15.3, 16, 16.3, 17, 17.3, 18, 18.3, 19, 19.3, 20.3
2.		D2S441	2p14	TCTA	8–11, 11.3, 12–17
3.		D3S1358	3p21.31	TCTA complex	9–20
4.		D10S1248	10q26.3	GGAA	8–19
5.		D13S317	13q31.1	TATC	5–17
6.		Penta-E	15q26.2	AAAGA	5–25
7.		Amelogenin	Xp22.1–22.3 and Yp11.2	NA	X, Y
8.	JOE-6C	D2S1338	2q35	TGCC/TTCC	10, 12, 14–28
9.		CSF1PO	5q33.1	AGAT	5–16
10.		D16S539	16q24.1	GATA	4–16
11.		D18S51	18q21.33	AGAA (20)	7–10, 10.2, 11–13, 13.2, 14–27
12.		Penta-D	21q22.3	AAAGA	2.2, 3.2, 5–17
13.	TMR-6C	TPOX	2p25.3	AATG	4–16
14.		D5A818	5q23.2	AGAT	6–18
15.		D7S820	7q21.11	GATA	5–16
16.		TH01	11p15.5	AATG (20)	3–9, 9.3, 10–11, 13.3
17.		vWA	12p13.31	TCTA complex	10–24
18.		D21S11	21q21.1	TCTA complex	24, 24.2, 25, 25.2 26–28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36–38
19.	CXR-6C	SE33	6q14	AAAG complex	4.2, 6.2, 8–12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2, 18, 18.2
20.		D8S1179	8q24.13	TCTA complex	7–19
21.		D12S391	12p12	AGAT/AGAC complex	14–17, 17.3, 18, 18.3, 19–27
22.		D19S433	19q12	AAGG complex	5.2, 6.2, 8–12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2, 18, 18.2
23.		D22S1045	22q12.3	ATT	7–20
24.		FGA	4q28		

(continued)

Table 7 (continued)

S. No.	Label dye	STR marker	Chromosome location	Repeat sequence	Allelic variation
	TOM-6C			TTTC complex	14–18, 18.2, 19, 19.2, 20, 20.2, 21, 21.2, 22, 22.2, 23, 23.2, 24, 24.2, 25, 25.2, 26–30, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 48.2, 50.2
25.		DYS391	Y	TCTA	5–16
26.		DYS570	Y	TTTC	10–25
27.		DYS576	Y	AAAG	11–23

and Applied Biosystem 3500 genetic analyzer, 3500xL genetic analyzer, 3130 genetic analyzer, and 3130 xL genetic analyzer although optimization of the different process is required as per the standard protocol provided by respective organization. Value obtained for the same sample can differ between instrument platforms, because of different polymer matrices and electrophoretic conditions.

Forensic Relevance of Amelogenin Marker

In forensic science DNA analysis plays a vital role in identification of individuals. In forensic community and research community, there is frequent need to determine the sex of individuals based on DNA for the identification of victims, accused, or culprit. The ability to determine the sex of an individual based on DNA evidence can be crucial in forensic investigation. In humans, a specific marker is used for gender determination known as amelogenin. The amelogenin locus has two homologous genes present on both the X and the Y chromosomes referred as Amel-X and Amel-Y. However, there are size differences in this gene between chromosomes enable to distinguish between male and female and have been used for sexing in forensic science. Forensic DNA typing of sex determination in human has been proven to be imperious tool to criminal justice system. The widespread use for amelogenin marker for gender determination or human identification helps forensic community in different type of cases such as: sexual assaults cases, parental dispute, gender identification, missing person investigation, and identification of victims of mass disaster. It also can differentiate the relative contributions and estimate the ratio of male and female DNA in mixed forensic samples. The amelogenin marker for gender determination is discovered by Sulivan et al. in 1993 and published in "A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin, BioTechniques, 15, 637-641." The physical location of amelogenin gene AMELX on distal short arm of X-chromosome is Xp22.1–22.3 and AMELY on near the centromere of the Y-chromosome is Yp11.2. It encodes a matrix protein which is involved in forming of enamel known as amelogenesis.

The British Forensic Science Service was the first to describe the particular primer sets that are used in forensic DNA laboratories today (Butler 2015). Most commonly used primer sets are designed by Sullivan et al. Presenting amplicon of AMELX/AMELY are 106/112 and 212/218 bp, respectively. The PCR product of AMELX and AMELY can be distinguished from one each other by primer flanking of a 6 bp deletion in the 3rd intron AMELX which is absent in AMELY. Both AMELX and AMELY can be amplified and analyzed in a single reaction along with other loci. After amplification of the amelogenin gene, if the shorter fragment is observed which gives a single peak this indicates that the DNA samples contain X-chromosomes and that will represent female phenotype but if both fragments are observed, i.e., longer and shorter fragments exhibit two peaks, then this indicates the presence of chromosome X and Y which represents male phenotype. This size variation of amelogenin gene in chromosome X and Y is used by forensic community to determine the gender of the individuals.

However, there is several challenges that pose threat for a proper DNA profiling for forensic community and research community such as: degraded samples, PCR inhibitors, presence of mixed DNA, mutation, AMELX/AMELY deletion, primer site polymorphism, and aneuploidy. To overcome this problem, advanced next generation STR kit contain multiple marker for gender discrimination which decrease the anomalies that could make the sample falsely appear as a female in gender determination. Addition of multiple marker increases the gender discrimination and identification of an individual DNA samples.

Forensic Validation of Autosomal STR Kits

Whenever a DNA profile is generated, questions always arise about the credibility and reliability of the DNA profile and the process of the DNA typing by jury, lawyers, and criminal justice system. Forensic validation of autosomal STR kits is a vital part of DNA typing, whenever a DNA profile is submitted to court. Validation is required to ensure that the method and process is accurate, efficient, without any error, relevant, and backed by scientific explanation (Jamie et al. 2019).

Validation of autosomal STR kits is categorized into different category to check the efficacy of generated DNA profile such as:

- **Developmental validation**: Extensive efficacy performance by manufacturer of the method and technology before commercializing or publishing (Melody et al. 2009).
- **Internal validation**: Check the effectiveness of the introduced method, process, or technology by an individual's laboratories.
- **Performance checks**: Check the performance of instruments either new or repaired, calibration of the instruments, validation of the glassware and reagents, and assessment of the control samples or reference samples.

Forensic validation of autosomal STR kits ensures proficiency test, standard operating procedure, sensitivities, reproducibility, precision, historical performance if any, limitation, probabilities, and chances of errors.

A forum of technical working group known as scientific working group on DNA analysis methods (SWGDAM) comprised of forensic scientists as well as academic expert and other expert. SWGDAM assess the method, protocols, training, and research and evaluate the quality assurance to provide recommendation to the FBI (Julio et al. 2008). SWGDAM provides guidelines for validation of different autosomal STR kits known as validation guidelines.

The following parameters are accessed for the forensic validation of autosomal STR kits:

- **Specificity**: PCR primers target specific loci in DNA for amplification. These similar loci or sequence may be found in different animal such as chimpanzee, and other mammals. During PCR amplification of human STR marker cross-reactivity may occur with nonhuman DNA which can generate allelic peak in electropherogram. To overcome these problem different parameters is used in autosomal STR kits. These STR kits are manufactured in a way that it will be specific to human DNA samples. Also, RFU is identified and determined to exclude the nonhuman allelic peak in an individual DNA profile. For any peak to be considered in analysis, it should be on or above-threshold RFU. Different STR kits provide the specific data and guideline with respect to RFU peak inclusion and exclusion and for different species after validation studies.
- **Sensitivities**: Efficacy of the material, method, and instruments are very important during analysis. Different autosomal STR kits are PCR-based method which amplifies the target loci in the samples. For amplification of targeted segments, sample should be in enough quantity but it is not always possible to get enough quantity of the samples for analysis. The sensitivity of the STR kit determines the enough quantity is required for amplification of the targeted segments. Different STR kits shows different sensitiveness for the amount of samples and improving the STR kit increases the efficacy of sensitivity which can detect and processed the low amount of DNA samples (Hammond et al. 1994). The amount of required DNA for different STR kits may vary from ng to 0.5 pg.
- **Reproducibility**: Whenever a method or materials is introduced to public, it need to be tested and documented for producing adequate information about the results. The results should be consistent when an experiment is repeated many times. The reproducibility of STR kits provide the information of allele size range, buffer, mixture, and other materials included in STR kits are consistent or not. The biological samples collected from crime scene are often contaminated or degraded which can be overcome by enhanced buffer system and material included in the kits for optimal amplification of the target segments of the DNA (Holt et al. 2002).

- **Inhibition study**: The biological samples collected from the crime scene for DNA profiling is often contaminated and may contain inhibitor factors. The often-encountered inhibitor factors are hematin, humic acid, tannic acid, phenol-chloroform, and other inhibitors. These inhibitors may inhibit PCR amplification which may result in alter peak or lesser peaks of the alleles.
- **Stutter product**: Allelic dropout and allelic dropout is often encountered during PCR amplification of the loci which resulted in loss of peak height. This phenomenon is known as stutter product. During DNA profiling stutter product is often encountered. This problem is generally handled with standard deviation method or threshold point inclusion and the guideline provided by respective manufacturer of STR kit. Although different validation studies show about the problem of stutters during PCR amplification of target sequence of DNA and often assessed by different calculation.

Usefulness of Autosomal STR Kits in Forensic Practice

DNA fingerprinting was first used in forensic science by Sir Alec Jeffreys of University of Leicester to verify a suspect's confession which he confessed for two rape and murders in 1986. With the help of DNA profiling test, it was founded that he had not committed the crimes.

First paper publication on STR was published on 1991 while first STR kit was published by Applied Biosystem is AmpFISTR Blue in October 1996. There are many types of uses of STR kits in different scientific community such as:

- Environmental and health science
- Genetics
- Evolution
- Cell and molecular biology
- Forensic science

In forensic science, DNA profiling is an ultimate tool for forensic investigation based on DNA evidence. It helps criminal justice system to determine the victim, accused, or culprit. It helps in different types of forensic caseworks such as:

- Human identification
- Paternity dispute
- Mass disasters cases
- Historical investigation
- Missing person investigation
- Military DNA (dog tag) case work
- Preparation of DNA databases

Conclusion

The ultimate power of DNA profiling as an identification of individuals provides a tremendous technology which helps in forensic investigation and criminal justice system. Forensic community use the biological resource to obtain the DNA of an individual for the forensic DNA typing based on commonly used STR markers. STR sometimes also referred as microsatellite or simple sequence repeats. Human genome contains thousands of STR markers, only a small core set of loci have been selected for forensic DNA analysis and human identification. These STR loci is authenticated by different agency and organization such as Combined DNA Index System (CODIS) of the USA, core loci managed by FBI and NIST, European Standard Set (ESS), and many more organization to use these loci in forensic DNA profiling. Theses STR loci is amplified with different PCR amplification kit commercialized by different corporation. These organizations provide a variety of autosomal STR kit that meets the requirement of forensic community and used in forensic investigation for different types of casework. The development of new generation STR kit is required to meet the advance era scenario which can overcome limitation of old generation kit and provide result which can be interpreted easily. The harsh condition, environment, contamination, and many more issue are faced during DNA typing from the samples. The continuous advancement and improvement is the need of the hour to overcome all the problem faced by forensic community in DNA typing.

References

Al Janaahi N et al (2019) Forensic evaluation of Verifiler plus 6-dye chemistry kit composed of 23 loci with casework samples. Forensic Sci Int Genet Suppl Ser 7(1):892–896

Al-Snan NR et al (2019) Population genetic data of the 21 autosomal STRs included in Globalfiler kit of a population sample from the kingdom of Bahrain. PLoS One. https://doi.org/10.1371/journal.pone.0220620

Boavida A et. al. (2017), Powerplex Fusion 6C system: Internal Validation study, Forensic Genetics, Forensic Science Research 3(2). https://doi.org/10.1080/20961790.2018.1430471.

Bogas V et al (2015) Testing the behaviour of Globalfiler PCR amplification kit with degraded and/or inhibited biological samples. Forensic Sci Int Genet Suppl Ser 5:e21–e23

Brito P et al (2015) Direct amplification of casework samples in fabrics using globalfiler PCR amplification kit. Forensic Sci Int Genet Suppl Ser 5:e487–e489

Butler JM (2005) Forensic DNA typing, 2nd edition biology, technology and genetics of STR markers. Academic, Elsevier

Butler JM (2011) Short Tandem Repeat Loci and Kits, Science direct. https://doi.org/10.1016/B978-0-12-374513-2.00005-1.

Butler JM (2012) Advanced topics in forensic DNA typing: methodology. Academic, Elsevier

Butler JM (2015) STR Profiles: Multiplex PCR, Tri-alleles, Amelogenin and partial profiles. https://doi.org/10.1016/B978-0-12-405213-0.00005-1.

Butler JM (2018) Short tandem repeat typing technologies used in human identity testing. Biotechniques 43(4):ii

Butler JM, Hill CR (2012) Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Forensic Sci Rev 24:15

Christian RM et al (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Res NCBI 29(1):320–322

CODIS Operating policies and procedures manual (2009), issued by Biology Program Manager, DFS Document 210-D100
Dixit S et al (2019) Forensic genetic analysis of population of Madhya Pradesh with powerplex fusion 6c multiplex system. Int J Legal Med 133:803–505
Ensenberger MG et al (2016) Developmental validation of Powerplex Fusion 6C system. Forensic Sci Int Genet 21:134–144
Gouveia N et al (2015) Direct amplification of reference samples with globalfiler PCR amplification kit. Forensic Sci Int Genet Suppl 5:e135–e137
Gettings KB et. al. (2015) STR Allele sequence variation: Current knowledge and future issues, FSI: Genetics 18:118–130. https://doi.org/10.1016/j.fsigen.2015.06.005.
Goodwin W et al (2012) An Introduction to Forensic Genetics, 2nd Edition, Wiley-Blackwell Publication, ISBN-978-0-470.
Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55:175–189
Hares C (2012a) Expanding the CODIS core loci in the United States. Forensic Sci Int Genet 6:e52–e54
Holt CL, Buoncristiani M et al (2002) TWGDAM validation of AmpFSTR™ PCR amplification kits for forensic DNA casework. J Forensic Sci 47(1):66–96
Jamie AL et al (2019) Interlaboratory study to validate a STR profiling method for intraspecies identification of mouse cell line. PLoS One. https://doi.org/10.1371/journal.pone.0218412
Jasmine W, Julie MA et al (2019) Sensitivity and baseline noise of three new generation forensic autosomal STR kits: Powerplex Fusion, Verifiler Plus and Investigator 24plex QS. Forensic Sci Int Rep 1:100049
Julio MJ et al (2008) Development and Validation of the AmpF'STR MiniFilerTM PCR Amplification Kit: A MiniSTR Multiplex for the Analysis of Degraded and/or PCR Inhibited DNA. J Forensic Sci 53(4):838
Julio MJ et al (2013), Next generation STR Genotyping kits for Forensic applications, Forensic DNA Analysis- Current Practices & Emerging Technology, 1st Edition, ISBN: 9780429254338, CRC Press.
Jebor MA et. al. (2015), Detection Detection of new variant Off-ladder at the D12S391, D19S433 and D1S1656 loci and tri-allelic pattern at the D16S539 locus in a 21 locus autosomal short tandem repeat database of 400 Iraqi Individuals, African Journal of Biotechnology, Academy Journals 14(5):375–399. https://doi.org/10.5897/AJB2014.14103.
Ludeman MJ et al (2018) Developmental validation of GlobalFiler PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. Int J Legal Med 132:1555–1573
Melody D et al (2009) Developmental validation of short tandem repeat reagent kit for forensic DNA profile of canine biological materials. Croat Med J 50(3):268–285
Pankaj S, Dash HR (2019) Forensic effectiveness and genetic distribution of 23 autosomal STRs included in Verifiler Plus multiplex in a population sample from Madhya Pradesh. Int J Legal Med 134:1327–1328
Peter G et al (2020) Forensic Genetics: The basics, Science Direct. https://doi.org/10.1016/B978-0-12-820562-4.00009-2.
Product Bulletin, AmpFlSTR Identifiler PCR amplification kit, Applied Biosystem
Protocol, Use of the powerplex Fusion 6C system to amplify extracted DNA (2020), Promega Corporation
Rabelo KCN et al (2015) Traces samples of human blood in mosquitoes as a forensic investigation tool, Genet. Mol. Res 14(4):14847–14856. https://doi.org/10.4238/2015.Novermber.18.50.
Siddhita G et al. Verifiler direct PCR amplification kit for paternity testing, 78-poster verifiler, Life Technology Corporation. Applied markets-Human identification
Tanya C (2015) Identification of variant alleles at CODIS STR loci in different populations of India. ResearchGate. https://doi.org/10.19070/2332-287X-1500042
Technical Manual, Powerplex 16HS system (2016) Promega Corporation
Technical Manual, Powerplex Fusion 6C system for the use on the Applied Biosystem Genetic Analyzers (2018) Promega Corporation

Technical Manual, Powerplex Fusion 6C system for the use on the spectrum compact CE System (2020) Promega Corporation
Thatch C et al (2012) Internal validation of PowerPlex® 16 HS with the Applied Biosystems® 3500xL Genetic Analyzer, Marshal.edu/Forensic/Files
User Guide, AmpFlSTR identifiler Direct PCR amplification kit (2018) Cat- 4467831, 4408580, 4415125, Pub No.:4415125, Applied Biosystem
User Guide, AmpFlSTR identifiler PCR amplification kit (2018) Catalog Number, 4322288, Publication number 4323291, Applied Biosystem
User Guide, AmpFlSTR identifiler Plus PCR amplification kit (2015) Pub No.:4440211, Applied Biosystem
User Guide, Verifiler Direct PCR amplification kit (2013) catalog number 4479567, Publication No 4479552, Applied Biosystem
User Guide, Verifiler express PCR amplification kit (2017) Catalog numbers A32014, A32070, A33032, Publication No 100043588, Applied Biosystem
User Guide, Verifiler Plus PCR amplification kit (2020) Catalog number A35495, Publication No MAN0017493, Applied Biosystem
Westen AA et al (2014) Comparing six commercial autosomal STR kits in a large Dutch population sample. Forensic Sci Int Genet 10:55–63

Forensic Human Y-Chromosome Markers: Principles and Applications

8

Arash Alipour Tabrizi

Contents

Abstract

The human Y chromosome distinguishes males from females and passes through male lineages as a highly polymorphic data-field linked together which is called

A. Alipour Tabrizi (✉)
Legal Medicine Research Center, Legal Medicine Organization, Tehran, Iran

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_6

Y-haplotype. Out of different types of polymorphisms mainly produced by the mutational process, Y short tandem repeats (Y-STRs) are the most popular markers implemented in commercial multiplex amplification kits and routinely used for forensic genetic casework analysis. The unique specifications of Y-STRs and mini Y-STR markers make them powerful alternative tools in the hands of forensic experts to analyze compromised crime scene evidence and solve complicated sexual crimes even in sperm-free samples or in the presence of minimal amounts of male DNA in a mixture. The exceptional features of rapidly mutating Y-STRs suggested them as the markers of the choice to differentiate among male paternal relatives. The application of national/international Y-haplotype databases and familial searching strategy empowers forensic investigators to narrow investigation domains. The characteristics of Y-chromosome single-nucleotide polymorphisms could be beneficial for haplogroup determination and historical investigations. In addition to the benefits of typing in sexual assault investigation, sex determination, and paternity dispute analysis, various technical limitations should be considered and addressed in standard interpretations and reporting guidelines.

Keywords

Y chromosome · Forensic genetic · Y-STR · Y-SNP · Y-haplogroup · Y-haplotype database

Introduction

In addition to visible phenotypic differences between males and females, which were remarkable for the humans of the distant past, several gender-specific feathers are assigned to modern human males. These gender-specific feathers include the endless use of the television remote buttons, the ability to remember the soccer scores while amnesiac for his wedding anniversary, driving hours without asking the right direction, and the inability to differ a lapis, azure, and blue bag. The Y chromosome as an indicator of maleness is crucial for phenotypic male traits, fertility maintenance, cerebral asymmetry, handedness, and tooth size.

Y-Chromosome Structure

Since 1905, human sex determination has been based on the presence or lack of the Y chromosome. A normal karyotype of a human male has one Y chromosome; however, no Y chromosome is detected in human females. The full human Y-chromosome sequence was announced in 2003 for the first time. About 2% of the total human genome (i.e., approximately 60 million nucleotides) consists of the Y chromosome. This third-smallest human chromosome in a normal male karyotype contains the least number of genes in comparison with any other chromosomes.

The determination of genetic gender in humans is based on the male determining genes (i.e., sex-determining region Y [SRY] genes). The SRY genes are located in the male-specific region of the Y chromosome (MSY) in the non-recombining region of the Y chromosome (NRY) nearby the distal end of the short arm (p) of the Y chromosome encoding the developmental factors of the testes. The MSY contains 78 functional protein-coding sequences encoding 27 different proteins. In addition to the coding regions, the MSY consists of repetitive sequences.

Different from autosomes, only about 5% of the Y chromosome could have recombination with the X chromosome. These three portions are located at the distal tips of the Y chromosome called the pseudo autosomal regions (i.e., PAR1, PAR2, and PAR3). The PAR1 and PAR2 are located at the ends of the short arms and long arms of the X and Y chromosomes, respectively. PAR3 defined in only 2% of the general population has over 98% of sequence homology with Xq21.3 which was identified in 2013.

Due to this lack of recombination, the mutational processes are the most responsible for diversity along the Y chromosomes of paternal lineages. These mutations will be passed to the next generations and not be repaired due to the lack of crossing over. Therefore, as the only known mechanism for variation between the males of a paternal lineage is a mutation, more mutations in a marker lead to further power of discrimination and greater use in forensic application and human identification. As a mosaic aneuploidy, the loss of chromosome Y (LOY) can be typically detected in the blood samples of males. There is a direct correlation between the incidence of LOY and chronological age in healthy males older than 16 years. Since buccal or white blood cells are sampled and routinely employed in forensic cases, the LOY may affect the results. It is necessary to carry out further studies on the effect of the LOY on the forensic analyses of these samples for better clarifications (Barros et al. 2020).

Markers on Y Chromosome and Polymorphisms

Y-chromosome variations and polymorphisms were revealed in searching for the 1000 Genomes Project data set. Different types of polymorphisms, including microsatellites, minisatellites, and single-nucleotide polymorphisms (SNPs), are located on the Y chromosome along the non-recombining region and make it a suitable target for polymerase chain reaction (PCR)-based analysis for forensic purposes. In a simple classification, the mostly employed forensically relevant Y-chromosome markers are divided into two main categories, namely, biallelic markers with only two possible alleles and multi-allelic markers.

Repeating microsatellites or Y short tandem repeats (Y-STRs) as the main multi-allelic forensic genetic markers and Y-chromosome single-nucleotide polymorphisms (Y-SNPs) in the subset of biallelic markers are widely used for human identification objectives. Another type of polymorphism defined as different internal sequence variations among alleles with a similar length is called SNPSTR. An example of this type of polymorphism on the Y chromosome revealed by base composition analysis is two “17” iso-alleles reported for

DYS635. Both of them appear the same alleles in capillary electrophoresis (CE) analysis, indicating different sequence variations in the next-generation sequencing (NGS) (Butler 2014).

Y-STRs

Genetic information based on the autosomal short tandem repeat (aSTR) markers are inherited half from the mother and half from the father; however, Y-STRs, as popular non-autosomal markers, are located on the human Y chromosome. A Y-STR haplotype consists of a combination of alleles noticed at all tested loci as a single entity. Therefore, Y-STRs are indicated as paternal lineage markers, and each of the loci has a single allele, rather than two alleles in aSTRs, and is considered a Y-haplotype and not a genotype.

Since there is one Y chromosome, instead of homo−/heterozygosity observed in aSTRs, there is hemizygosity in Y-STRs. Approximately 4,500 Y-STRs were identified and utilized in different applications and other multidisciplinary studies (He et al. 2019). The Human Genome Organization Gene Nomenclature Committee assigned names for Y-STR markers based on the standards for human gene nomenclature. Accordingly, a marker name consists of two parts, including the DYS that is the abbreviation for "deoxyribonucleic acid (DNA) Y-chromosome single-copy sequence" followed by a unique identification number reflecting the order in which the Y-STR is discovered. The DYS- and DYF-numbers are referred to as single-copy and multi-copy markers, respectively. The first forensically useful Y-STR locus applied for human identification was Y-27H39, currently better known as DYS19, which was described by Lutz Roewer et al. in 1992. Besides, the potential application of the Y-STRs in forensic analysis has been determined for almost two decades. In a short time following the characterization and assessment of the first Y-chromosomal STR polymorphism, its practicality in crime casework was shown when a mixed stain from a vaginal swab of a raped and murdered female victim was resolved using Y-STR analysis, and a falsely convicted male was excluded in 1992 (Gill et al. 2001; Roewer et al. 1992).

The first set of nine Y-STR loci includes DYS19, D YS389I/II, DYS390, DYS391, DYS385 a/b, DYS392, and DYS393selected as a standard for data transfer and population data management by the Scientific Working Group on DNA Analysis Methods (SWGDAM) and European Community and referred to as the minimal haplotype set in 1997. All minimal haplotype loci are included in commercially available Y-STR typing kits. Moreover, to date, tens of the new additional Y-STRs are included in the forensic grade Y-STR typing kits by different kit providers. Although most Y-STR loci included in Y-STR typing kits have tetranucleotide repeats (i.e., four-base pairs), the loci with five- or six-base pair repeats have been selected for forensic applications. In addition to a high degree of polymorphism, penta- and hexanucleotide repeat loci show a much lower degree of stutter product formation; as a result, they are considered valuable tools for mixture interpretations (Butler 2003).

According to the comprehensive determination of the physical map of human Y-STRs, it was shown that most of the Y-STR loci were located within the Yq11.221, Yq11.222, and Yq11.223 segments and Yp11.2 segment on the long and short arms of the chromosome, respectively. No Y-STR loci were located in the telomeric regions of the Y chromosome, Yp11.32, and Yq12. Furthermore, only three loci (i.e., DYS631, DYS716, and DYS707) were located in a centromeric region on Yp11.1 (Hanson and Ballantyne 2006).

Several markers in commercial kits always show more than one allele due to the identical copies of sequence on the Y chromosome (e.g., DYS385 and DYF387S1, each of which comprised of two loci) despite the fact that each Y-STR locus demonstrates a single allele. DYS464 is the most impressive polymorphic Y-STR discovered, with at least four copies in the Y chromosome as a duplicated locus (Butler 2003; Roewer et al. 2020). The DYS389 locus is complex consisting of two polymorphic regions. DYS389I is a polymorphic element within DYS389II as the whole locus. The forensic application of Y-STRs is the result of their high levels of polymorphism in human populations, small size in base pairs (range: 100–400 bp), and the ability for automation via a multiplex Y-STR PCR reaction (Hammer and Redd 2005).

Rapidly Mutating Y-STRs

The evaluation of mutation rates for Y-STRs indicated that most of them mutate at a rate similar to that of aSTRs. Although it might be different in different loci in various populations, it is estimated that Y-STRs mutate about 1–4 per thousand per generation. As previously discussed, the only known mechanisms for variation among Y-STRs are the mutational events of slipped strand during the replication of DNA. Even though all Y-STRs are located and linked on the same chromosome, the mutation of various Y-STR loci occurs independently from each other.

The structure of the locus and quantity of repeats (i.e., allele number) are for the most part associated with the mutation rate of Y-STRs, as more repeats lead to more DNA slippage during replication. Therefore, it is estimated that fathers passing greater Y-STR allele numbers have an increased chance of mutation to occur in comparison with fathers passing short Y-STR alleles at the same loci (Ralf et al. 2020). As a result, Y-STR mutation rates show inter-locus and intra-locus variations. The Y-STR loci, including compound repeats, tend more to mutate than simple repeat ones. They almost change to a single repeat gain, more common than multi-repeat changes, and the mutation rates differ based on the locus-specific and allele-specific manners.

In order to decrease the suspect pool via the separation of distant male relatives from close ones and further diversification of the population, forensic geneticists tend to apply an increasing number of Y-STRs, especially the application of rapidly mutating (RM) Y-STRs, to provide more chance of the detection of Y-STR mutations allowing reaching higher discrimination capacities. Out of the first 17 Y-STR loci in use, only the mutation rates of DYS458 and DYS439 are higher

than 0.5%. Nevertheless, a higher rate of mutation is shown for several more recently added loci, such as DYS570 (2.23%), DYS576 (1.91%), and DYS481 (0.955%) (Butler 2014).

The estimation of Y-STR locus mutation rates is based on the Y-haplotype comparisons from direct father-son pairs ((Ballantyne et al. 2010; Claerhout et al. 2019). The results of comprehensive Y-STR mutation rate assay is almost 2,000 confirmed father-son couples revealing that out of 200 investigated Y-STRs, 13 Y-STRs (including single- or multi-copy RM Y-STR markers of DYF403S1, DYF404S1, DYS449, DYS518, DYS526, DYF387S1, DYF399S1, DYS547, DYS570, DYS576, DYS612, DYS626, and DYS627) were identified with exceedingly high rates of mutation ($>10^{-2}$ per locus per generation) nominated for RM Y-STRs. The evaluation of the RM Y-STR set in the context of male differentiation and power of detecting hidden paternal relationships demonstrated that it could help to discriminate between two male relatives even in populations with high levels of endogamy and ethnic fragmentation with lower proportions of unique haplotypes of RM Y-STRs (Ballantyne et al. 2014; Adnan et al. 2016; Della Rocca et al. 2019; Almohammed and Hadi 2019).

Manfred Kayser et al. investigated the mutability of Y-STRs, mutation rates, features, molecular bases, and forensic implications. The results of the aforementioned study showed that Y-STR mutations occurred only after 11 or higher homogeneous adjacent repeats. In addition, two of the Y-STRs (i.e., DYS570 and DYS576) represented much higher rates of mutation in comparison with those included in the commercial kits (Kayser 2017). The goal of increasing the discrimination power of multiplex Y-STR haplotype leads to implement RM Y-STR markers in newly emerged Y-STR kits. In 2010, Manfred Kayser et al. indicated that differentiating a male from his close relatives may be possible by the application of a set of 13 RM Y-STRs. The development of RM Y-STRs provides a new perspective on the future of the forensic analysis of the Y chromosome.

The Y-STR markers are classified into four groups according to estimated rates of mutation, namely, slowly mutating (SM Y-STRs), moderately mutating (MM Y-STRs), fast mutating (FM Y-STRs), and RM Y-STRs. The mutation rates of SM Y-STRs are reported as $<10^{-3}$; however, MM Y-STRs are reported with the mutation rates within 5×10^{-3} to 1×10^{-2} per marker per meiosis. The calculated rates of mutation for MM Y-STRs are within 1×10^{-3} to 5×10^{-3}. Furthermore, the mean mutation rate of RM Y-STRs is reported as 2.6×10^{-2} ($\geq 10^{-2}$). The DYF399S1 is identified as the most mutable of all currently noted RM Y-STR markers, with an estimated mutation rate of 6.9×10^{-2} (Ralf et al. 2020).

The higher number of mutations investigated in RM Y-STRs leads to the greater differentiation among male relatives; therefore, applying RM Y-STRs could increase the power of discrimination between unrelated individuals and males of the same patrilineage providing important tools to compel several of the current limitations of Y chromosome analysis in forensic casework (Ballantyne et al. 2012). There is an ongoing investigation of new Y-STR markers to be added to the next generations of the Y-STR multiplex kits used in casework analysis. The incorporation of RM Y-STRs in modern Y-STR profiling kits (AmpFlSTR Yfiler Plus™; Applied

Biosystems, Foster City, CA, USA) and PowerPlex® Y23 System (Promega Corporation, Madison, WI, USA) increased the power of discrimination and weight of evidence for a Y-haplotype match.

The RM Y-STRs due to their high rates of mutation raised drawbacks to the calculation of paternity/kinship probabilities turning them not suitable for these applications (Claerhout et al. 2019). Although it is expected that more Y-STR markers, particularly RM Y-STRs, will be included in subsequent commercial Y-STR kits, different Y-STR panels/kits should be adopted for various forensic implementations of Y-STR haplotyping (i.e., paternal lineage differentiation, paternal/kinship determination, and familial searching). Accordingly, RM Y-STR-based kits are highly suitable for the differentiation of multi-suspect sexual assault with paternal relationships. Moreover, Y-STRs reported with low/medium rates of mutation are appropriate markers for the determination of patrilineality/kinship and familial searching.

Mini Y-STRs

The Y-STR analysis in the forensic context has not always been successful. Forensic biological evidence also consists of minimal samples, buried underground for extended periods, which are greatly degraded. The amplification of several Y-STR amplicons cannot be easily achieved in such compromised samples. Two well-known modifications of the PCR condition include an increase in the number of PCR cycles and the addition of further DNA polymerases to the reaction. Another efficient strategy to overcome degradation problems and achieve successful results is the incorporation of new primer sets (i.e., Y-mini plex primer sets). Designing primers was carried out after a mini Y-STR approach through the close placement of primers to the Y-STR markers to the greatest extent possible (Nunez et al. 2017). Therefore, in the case of archeological analysis of human DNA, it is particularly suggested to consider these coupled strategies as efficacious tools in future studies by molecular anthropologists for the conventional analysis of Y-STR.

Inheritance Pattern of Y-STR Markers

The Y-STR loci located on the NRY can be considered a single locus; accordingly, Y-STR data are linked together on the Y chromosome and exclusively transmitted through the patrilineal line as a "haplotype." Therefore, if individuals are expected to have an equal Y-STR haplotype, they are from a similar paternal lineage. Because the haplotypes are dealt with as the alleles of a single locus, Y-haplotypes could be assumed as a family fingerprint assisting in paternity testing (Barros et al. 2020).

A Y-haplotype will be shared with the patrilineal relatives (i.e., brothers, father, sons, and paternal uncles) of a specific male which should be taken into account in any evidential analyses. Additionally, the Y chromosome contains a unique record of human history due to its male-specific (i.e., holandric)inheritance. Moreover, in

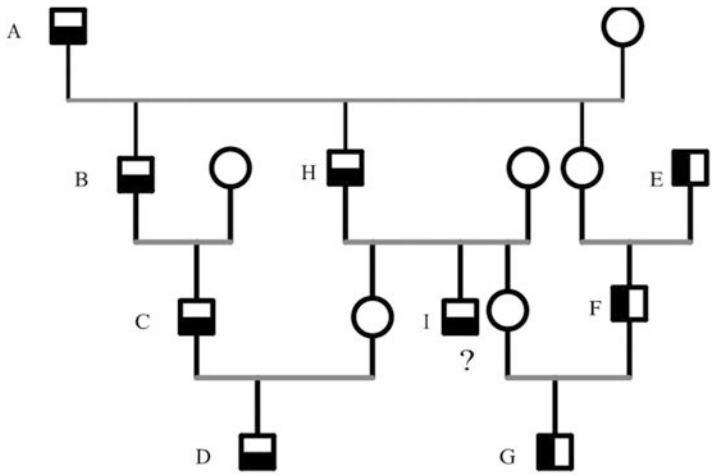


Fig. 1 A pedigree indicating patrilineal inheritance of Y-STR markers. It can assist to identify the person in question (I) in disaster victim identification (DVI) as well as familial searching in comparison with Y-STR data of the other male relatives (A, B, C, D, and H) and not (E, F, and G) as reference samples. More attention is needed in complicated pedigrees with consanguineous marriages

addition to the absence of crossover for most of the Y-chromosome lengths and their uniparental inheritance model, haploid markers are considered among the advantageous investigative tools available for forensic geneticists and genetic genealogists to study the historical figures' Y lineages as seen in Fig. 1.

Most notable studies have been carried out on the lineage of the third US President Thomas Jefferson, Napoléon, French House of Bourbon, the Russian Romanov family, Genghis Khan, and Emperor CáoCào of China. The results of a study conducted by Wen et al. on clans claiming to be the living descendants of Genghis Khan's son revealed that thousands of miles away from China in Kazakhstan, the Y lineage of the Tore clan is traced back to Jochi, the first son of Genghis Khan, as their common ancestor (Wen et al. 2019).

Y-SNPs

Biallelic markers on the Y chromosome are shown with Y-SNPs and insertions/deletions (i.e., Y-indels). Biallelic markers present a perspective view of a paternal lineage quite similar to a picture of a flower garden taken by a wide-angle lens instead of a high-resolution flower close-up taken by a camera equipped with a telephoto lens. An Alu insertion (i.e., DYS287) was the first biallelic marker observed on the Y chromosome abbreviated as YAP for Y-chromosome Alu polymorphism (Butler 2003). Two essential terms (i.e., haplotype and haplogroup) should be defined before discussing some differences between Y-STRs and

Y-SNPs. The haplotype is described as a combination of Y-STR alleles directly inherited from male ancestors. In addition, the haplogroup is defined as a group of similar haplotypes sharing a common ancestor with identical Y-SNPs (Qian et al. 2017).

Apart from Y-STR markers, Y-SNPs evolve more slowly. The typical estimated mutation rate of Y-SNPs is about 3×10^{-8} per nucleotide per generation. The low mutation rate (i.e., 100,000 times) is regarded as the superiority of the application of Y-SNPs over Y-STRs suggesting them as the DNA markers of choice for the assessment of distant relationships between male individuals, populations, and paternal biogeographic ancestry investigations (Ballantyne et al. 2014). The mutation occurring in the germline cells brings about groups of individuals with similar Y-chromosomal haplotypes, which can be classified into specific haplogroups (Diepenbroek et a. 2019).

A system of defining Y-DNA haplogroups has been established by the Y Chromosome Consortium by letters A to T, with further subdivisions through the application of numbers and lowercase letters defined by particular genetic variants occurring in the MSY (Barros et al. 2020). Historically, it is supposed that anatomically modern humans are attributed to a single recent common origin in Africa, Adam, which is attributed to a theoretical male as the most recent common patrilineal ancestor of all humans. It was estimated that for the first time modern humans left Africa about 100,000 years ago and arrived on different continents. This theory provides sufficient generation steps to allow continental variations at different Y-SNPs resulting from Y-chromosome mutations. The results of the aforementioned study also demonstrated that a minority of contemporary East Africans and Khoisans are thought to be the descendants of the most ancestral patrilineages of modern humans (Muhanned et al. 2015). Technical limitations in the multiplexing of Y-SNPs as for Y-STRs should be overcome to achieve total superiority required for biogeographic ancestry (Kayser 2017).

Y-STR Kits

In the field of forensic genetics, scientists should work with a limited body of evidence consisting of a variety of degradations and mixtures which should be analyzed within the minimum analysis time while expecting the highest standards of precision and accuracy. The application of Y-STR markers in forensic casework requires principal features which should be clarified in standard protocols. First of all, the analytical results of DNA analysis without considering the markers and legal aspects of the case should be reproducible in every accredited laboratory, and the raw data and analyzed genotypes or haplotypes could be transferable and applicable for further comparisons. Consequently, forensic genetic communities adopt the sets of Y-STR markers for their national/international forensic analysis, and kit providers select Y-STR markers for their forensic grade kits.

The most widely utilized Y-STR typing kits not only select minimal haplotype loci (as previously described) but also include many more Y-STRs with a different

number of repeats, different locations on the Y chromosome, and some extra features making them beneficial for solving compromised forensic casework. The potential for multiplexing is the key factor in the selection of Y-STRs allowing fast typing of many markers in the minimum time required for a single analysis. As the Olympic motto, the following three specifications encourage kit designers to do their best in the field of forensic analysis of Y-STRs:

The faster in terms of the implication of fast PCR reaction mix and DNA polymerase enzymes compatible with direct PCR amplification and saving the time required for the DNA extraction process

The higher in terms of the discrimination power and informativeness of the analytical results (Y-haplotypes) and tolerance to different PCR inhibitors and DNA degradation

The stronger in terms of the ability to produce results by the minimum yields of extracted male DNA in the presence of higher yields of the victim (female) DNA as a minor contributor to solve mixture casework

Continuous research on new markers, chemistries, and detection techniques accelerate the development of commercial amplification systems in the last decade. Regardless of commercial considerations and trademarks, every forensic geneticist should be aware of the specifications, strengths, and weakness points of the systems and proper applications for which kits are designed. Through the aforementioned factors, each kit turns into a powerful tool in the hands of the trained investigator to tackle complex forensic casework. Concerning all kit providers, four popular Y-STR typing kits are discussed currently employed in forensic genetic laboratories. These characteristics should be embedded in standard operating procedures to provide greater consistency and accuracy among laboratories worldwide (SWGDAM 2014).

AmpFlSTR Yfiler™ and AmpFlSTR™ Yfiler Plus™ (Applied Biosystems, Foster City, CA, USA), PowerPlex Y System, and PowerPlex® Y23 System (Promega Corporation, Madison, WI, USA) have been designed and provided for the amplification of human male-specific DNA in the last two decades. The PowerPlex® Y System, introduced in 2003, provides an analysis of simultaneous amplification of 12 polymorphic Y-STR loci, namely, DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439, including both the European minimal haplotype and SWGDAM. The PowerPlex® Y System was designed for the analysis by the ABI PRISM® Genetic Analyzers using four-color chemistry panels.

The primers of 12 Y-STR loci are labeled with fluorescein® (i.e., DYS389I/II, DYS391, and DYS439), JOE® (i.e., DYS19, DYS392, DYS437, and DYS438), and TMR® (DYS385a/b, DYS390, and DYS393). The fourth fluorescent color (CXR) was allocated for Internal Lane Standard 600. Product specifications include a high level of specificity for human DNA and sensitivity with as low as 250 pg of DNA extracts. The PowerPlex kit is optimized to amplify 0.5–1.0 ng of male DNA, and amplified products are limited to lower than 335 bp in length benefiting no reactivity with >100-fold excess of female DNA. The provided allelic ladder is designed

to have an average of 35 bp separating spaces between two adjacent loci for the reduction of the potential for locus overlaps and consists of 102 alleles (Promega 2003).

AmpFlSTR YfilerTM PCR Amplification Kit (released in 2004 by Life Technologies, Foster City, CA, USA) is a five-dye multiplex PCR kit with reagents for the multiplex amplification of 17 Y-STR loci (i.e., DYS19, DYS385 [counted as two loci], DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS456, DYS458, DYS635, DYS448, and Y GATA H4) enabling full DNA profiles based on 125 pg of male DNA in only 3 h (Yfiler® Plus PCR Amplification Kit 2016). The PowerPlex® Y23 System (PPY23, Promega Corporation, Madison, WI, USA), released in 2012, as a five-dye Y-STR multiplex was developed for Y-STR typing at 23 loci in only 72 min. For the construction of the kit, beneficial properties, such as the short length of the fragments and an uninterrupted repeat structure, were considered. Six new markers (i.e., DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643), two of which (i.e., DYS570 and DYS576) classified as RM Y-STRs, were attached to the Yfiler® panel (PowerPlex Y23 System Technical Manual 2012; Thompson et al. 2013; Purps et al. 2014). However, since even close relatives might present one or more mismatches, especially at DYS570 and DYS676, the application of the PPY23 kit in kinship analysis or familial searching will render these practices increasingly complex. For the aforementioned applications, compulsory utilization of likelihood-based techniques should be considered which will be discussed later.

The PowerPlex® Y23 System benefits from a high sensitivity in case there is a female DNA (<0.1 ng, 1:6000 male to female DNA ratio) and a high tolerance for PCR inhibitors. Therefore, it is expected that the analysis of aged or damaged DNA samples will be considerably enhanced using the PPY23 kit (Purps et al. 2014; Goodwin 2016). Functional assays indicated that the application of Powerplex® Y23 improved the quality of Y-STR profiles in comparison with the use of Yfiler® with the possibility of extending timing intervals for obtaining vaginal swab samples up to 48 h passed from digital-vaginal penetration and 6 days following penile-vaginal intercourse in semen-negative samples in the guidelines for sexual assault examination (Owers et al. 2018).

The Yfiler Plus kit® (Thermo Fisher Scientific, Waltham, MA, USA) as the successor to the Yfiler® kit was released in 2014. Fast PCR implemented and direct amplification system simultaneously target 27 Y-STR loci consisting of all 17 loci included in the Yfiler kit, 7 RM Y-STR loci, and 3 highly polymorphic Y-STR loci designed for the analysis of forensic casework in less than 95 min. Among most tetranucleotide Y-STRs included in Yfiler® Plus, trinucleotide repeats (i.e., DYS392 and DYS481), a pentanucleotide repeat (i.e., DYS438), and a hexanucleotide (DYS448) are selected (see Table 1).

The Yfiler® Plus kit® is more sensitive than the YFiler™ kit with full profiles obtained mostly with male DNA inputs down to 125 pg. Moreover, the Yfiler® Plus kit® has been validated to function well in case the male/female DNA ratio is 1:1000 and improved PCR inhibitor tolerance. A high first-pass success rate (86%) for the direct PCR of reference FTA® samples was obtained using the Yfiler® Plus kit®

Table 1 Y-STR markers included in the minimal haplotype and commercially available Y-STR typing kits with dye labels and repeated motifs for loci in different kits

Kit/Loci/Repeated Motif	Minimal Haplotype	PowerPlex® Y System	AmpFlSTR Yfiler™	PowerPlex® Y23 System	AmpFlSTRTM Yfiler Plus™
Tetra (TAGA)	**DYS19**	**DYS19**	**DYS19**	**DYS19**	**DYS19**
Tetra (TCTG)(TCTA)	**DYS389I**	**DYS389I**	**DYS389I**	**DYS389I**	**DYS389I**
Tetra (TCTG)(TCTA)	**DYS389II**	**DYS389II**	**DYS389II**	**DYS389II**	**DYS389II**
Tetra (TCTA)(TCTG)	**DYS390**	**DYS390**	**DYS390**	**DYS390**	**DYS390**
Tetra (TCTA)	**DYS391**	**DYS391**	**DYS391**	**DYS391**	**DYS391**
Tetra (GAAA)	**DYS385a/b**	**DYS385 a/b**	**DYS385 a/b**	**DYS385 a/b**	**DYS385 a/b**
Three (TAT)	**DYS392**	**DYS392**	**DYS392**	**DYS392**	**DYS392**
Tetra (AGAT)	**DYS393**	**DYS393**	**DYS393**	**DYS393**	**DYS393**
Tetra (TCTR)compound		**DYS437**	**DYS437**	**DYS437**	**DYS437**
Penta (TTTTC)		**DYS438**	**DYS438**	**DYS438**	**DYS438**
Tetra (AGAT)		**DYS439**	**DYS439**	**DYS439**	**DYS439**
Tetra (AGAT)			**DYS456**	**DYS456**	**DYS456**
Tetra (GAAA)			**DYS458**	**DYS458**	**DYS458**
Hexa (AGAGAT)			**DYS448**	**DYS448**	**DYS448**
Tetra (TSTA)compound			**DYS635**	**DYS635**	**DYS635**
Tetra (TAGA)			**Y-GATA-H4**	**Y-GATA-H4**	**Y-GATA-H4**
Three (CTT)				**DYS481**	**DYS481**
Tetra (ATCT)				**DYS533**	**DYS533**
Tetra (GATA)				**DYS549**	
RM YSTR-Tetra (TTTC)simple				**DYS570**	**DYS570**
RM YSTR-Tetra (AAAG)simple				**DYS576**	**DYS576**
Penta (CTTTT)				**DYS643**	
RM YSTR-Tetra (AAAG)complex					**DYS627**
RM YSTR-Tetra (AAAG)complex					**DYS518**
RM YSTR-Tetra (TTTC) complex					**DYS449**
RM YSTR-Tetra (RAAG) complex and multi copy					**DYF387S1a/b**
Tetra (ATAG)					**DYS460**

(Henry et al. 2015). The most noteworthy advantage is improved balance in male DNA samples combined with female DNA at the ratios of >1: 1000 and faster time to obtain results with a higher power of discrimination mainly due to the included

RM Y-STRs. Due to the additional loci, the Yfiler® Plus kit® is expected to have a higher discrimination power and better ability to determine the correct number of contributors to a mixed profile with an almost equal cost of analysis per sample than the Yfiler or PowerPlex Y-23 System (Ferreira-Silva et al. 2018; Dash et al. 2020b).

The amplification of degraded DNA samples by the Yfiler® Plus® is facilitated due to the enhanced polymerase and buffer systems together with improvements in a panel design with 11 mini Y-STR loci (below 220 bp in size) due to the inclusion of 6 dyes (i.e., FAM™, VIC™, NED™, TAZ™, SID™, and LIZ™). Mobility-modifying non-nucleotide linkers are implemented in some of the primers to facilitate inter-locus spacing (Alghafri et al. 2015).

There was no cross-reactivity of the Yfiler™ Plus Kit® with any of these nonhuman species except for the DNA samples of chimpanzees and gorillas producing partial profiles in the 100–330-base pair region (Yfiler® Plus PCR Amplification Kit 2016; Gopinath et al. 2016; Ambers et al. 2018).

As the more Y-STR implemented and enzymes, master mix, and primer design improved, the results of the 27 loci become much better and reliable. However, these strategies have drawbacks. As more and more loci were profiled, PCR products inevitably become large and could not effectively amplify, especially in longer alleles. Apart from seven RM-YSTs added to the panel for the improvement of discrimination power, the RM Y-STR databases should be established for the estimation of match probabilities (Ballantyne et al. 2012). As Saadi of Shiraz, a major Persian poet, said in his masterpiece The Rose Garden that "Either makes no friends with elephant-keepers or build a house suitable for elephants," before the utilization of such a sensitive Y-STR typing kit, every laboratory should adopt contamination prevention strategies, along with updated guidelines for reliable interpretation (Tootkaboni 2017).

Interpretation of Y-STR Results

The interpretation of the results and accurate wording of the final reports are the two most critical steps in the Y-STR DNA typing process. A forensic investigator should be aware, fully experienced, and prepared for different situations happening in each step of the whole process. The interpretation step, by itself, should be objectively and consistently performed by two genetic experts. The guidelines for interpreting and reporting the results of Y-STR DNA are provided by national authorities and international societies. The application of these international standard protocols can terminate the local setups and personal wordings which may lead to misunderstanding, confusion, and misinterpretation of the unique analytical results in national/international casework.

An updated guideline of the application of Y-STRs in forensic analysis was published by the DNA Commission of the International Society of Forensic Genetics (Gusmao et al. 2006). The pattern of Y-STR peaks is obtained after the electrophoresis of multiplex PCR products, and haplotypic data are analyzed by the application of GeneMapper® ID-X (GMID-X)® software (Life Technologies, USA). Prior to the

interpretation of the Y-STR typing results, some preliminary assessments should be applied. The process is initiated by confirming properly functioned positive and negative controls, correct fragment sizing and allele callings, considering artifacts (especially stutters), designating the consistency of profile with being from a single male or a mixture, and determining which loci are interpretable. As a well-characterized PCR artifact, stutter refers to the appearance of a minor peak one repeat unit smaller than the target Y-STR allele product (i.e., minus stutter) or less frequently one repeat larger (i.e., plus stutter) (Butler 2014).

The next following step is the evaluation of the quality of a Y-STR profile for the determination of the sufficiency of the Y-haplotype information for application in further comparisons. Therefore, a forensic laboratory should establish guidelines to determine the qualification of the Y-STR haplotype for comparisons (SWGDAM 2017; Roewer et al. 2020). For the interpretation of a 23 Y-STR haplotype, there should be a minimum of 6 loci with data above the analytical threshold. Single source and single major contributor haplotypic data with a minimum of ten loci should be subjected to statistical calculation (North Carolina 2016). All the questioned samples should be solitarily qualified, analyzed, and interpreted before being compared to known (i.e., suspect) profiles. After the standard checkup of the data and comparison of questioned Y-haplotypes to known reference haplotypes, there are four possible results:

No Interpretable Results

No signals (i.e., peaks) above baseline or signals above baseline but below the analytical threshold at one or more loci were detected on the electropherogram.

Inconclusive Results

One or more loci provide interpretable peaks; nevertheless, no conclusive result can be achieved based on these limited data due to insufficient quality and/or quantity not supporting rendering a conclusion to be included or excluded. Partial haplotype as a result of allele dropout, stochastic effects, locus dropout due to inhibition, degradation, or limited amounts of DNA may cause inconclusive results. In contrast to what was previously discussed (no interpretable results), in this section there are limited or partial data; however, the haplotype is of no comparative value (i.e., uninterpretable) and not suitable for further comparisons. According to the updated version of Federal Bureau of Investigation Y-STR Interpretation Guidelines, at least 14 Y-STR data in a given Y-STR haplotype should be included to report a cannot exclude statement. The Y-STR haplotypes, including lower sets of data, will be used only for exclusionary purposes. Comparisons, in which exclusion cannot be made, will be assumed as inconclusive (FBI 2020).

Exclusion

In addition to the abovementioned benefits of Y-STRs in the forensic analysis of DNA, it has some special limitations because out of all Y-STR data directly inherited from a male to his male lineages, only an exclusion result of Y-STR analysis can be helpful. Once the Y-STR typing results were evaluated and approved for further comparison and decision-making, the haplotypes of known (i.e., reference) and questioned samples should be compared. The side-to-side comparison of two known and unknown Y-STR haplotypes may demonstrate discrepancies between the allele(s) of one or more loci. Functionally, all mismatches can be categorized into two main groups.

The first class consists of incompatible Y-STR haplotypes due to contributors belonging to different male lineages, called factual mismatches/exclusions. No analytical errors, technical limitations, or scientific justifications could be attributed to this class of mismatch cases. The second group is composed of differences that could be explained via scientific reasons (e.g., dropout, drop-in, mutations, and duplications), called factitious mismatches/exclusions. As more Y-STR loci are included in modern Y-STR typing kits, a single locus inconsistency between a male and his biological father is relatively regularly expected, and cases with double paternal genetic inconsistencies have been more frequently reported. In some forensic/kinship cases, there could be a difference in the multi-locus haplotypes of several brothers at a single locus as a result of the mutation of the fathers' germ cells which may not be observed in all of his sons (Andersen and Balding 2017; Mertoglu et al. 2018). No effective binding of a primer might result from a point mutation in the 3′end of the primer binding leading to the absence of a detectable amount of PCR product bringing about a null (silent) allele (especially in DYF387S1 and DYS385). Y-chromosome microdeletions caused null alleles at one or more loci. Accordingly, it is worth mentioning that such structural/sequence alterations might be interpreted as factitious exclusions. Suppose that two Y-STR haplotypes are compared from a buccal swab of a suspect and an unknown blood spot, detected on probative crime scene evidence. If the Y-STR haplotypes of the evidence and suspect differ by at least one allele, the suspect will be ruled out as a contributor to the unknown Y-STR haplotype given the haplotype interpretation and interpretation assumptions, including factitious mismatches (exclusion). A verbal statement (e.g., someone other than the suspect is the source of the DNA) should be mentioned to support the results.

Forensic geneticists may have to deal with the structural/sequence changes of the Y chromosome. In case of such an occurrence, it is of great importance to know how to interpret the nature of these results and how to disclose them in final reports. A simple strategy for exclusion interpretation in paternity or kinship analysis should be based on a minimum of three or more differences between two male samples to avoid factitious exclusions due to mutational events (Kayser and Sajantila 2001). A more conservative and smart approach for the exclusion of paternity or other kinship questions would be according to a flexible dynamic model, as an alternative to the application of a fixed rule-based upon ad hoc cutoffs (i.e., the exclusion of a

minimum of three or more Y-STRs). Accordingly, four fundamental WH questions need to be answered before the assessment of exclusion:

1. How many Y-STR loci have been analyzed?
2. What is the estimation of locus-specific mutation rate in the relevant population?
3. How many repeat number differences were observed between mismatching alleles?
4. What is the difference between the generations of two individuals with incompatible Y-STR haplotypes?

Although the profiling of Y-STR is advantageous for exclusionary goals due to the unambiguity of the results without the necessity to provide a statistical weight, the aforementioned considerations should be taken into account for the prevention of false exclusions. According to the guidelines of the SWGDAM, it is recommended to apply Y-STR mixtures not presenting a clear major contributor only for the exclusion (Ballantyne et al. 2014; Hampikian et al. 2017).

Inclusion/Match

In a comparison of a reference haplotype to an unknown haplotype obtained from questioned evidence, they may possess similar alleles at each locus (i.e., identical length alleles in case of the application of the CE method or identical sequence variants in case of the use of NGS), or few non-match alleles were observed which could be explained via scientific or technical resonating. Moreover, the reference cannot be ruled out as a possible contributor to the evidence haplotype.

A full match between a known suspect and evidence of Y-STR haplotypes only indicates that he could be one (and not the only) of the suspects contributing to the crime scene evidence as his male lineages could. Due to this linked inheritance, the strength of the match could be evaluated by the calculation of the rarity of the Y-STR haplotype in the relevant population. The estimation for a random match with Y-STR haplotypes, as a single locus by magnitudes more variable than single aSTR loci, is carried out via the counting method instead of the common product rule applying for aSTR genotypes. As the SWGDAM stated, an appropriate and conservative statistical technique for the assessment of the probative value of a match is provided using the counting method incorporating the upper-bound estimate of the count proportion.

In the counting method, the profile should be searched in different databases for a possible match to determine the rarity of a Y-STR haplotype. Basically, it means that how frequently a Y-STR haplotype could be observed in the relevant database, and it can be calculated by dividing the number of Y-haplotype observation/s (the numerator) by the total number of haplotypes included in the utilized database as the denominator. Therefore, as the size of the database increases, the random match more rarely occurs. A more useful Y-STR database indicates better statistics in this regard. A match probability (MP) between the two Y-STR haplotypes of 0.001

shows that there is a 1 in 1,000 chance of random selection of a second individual with this haplotype provided that it has already been noticed once. Given the chance of 1 in a billion commonly provided in forensic reports for the profiling of autosomal markers, 1 in 1,000 for YSTRs may not be persuasive enough. Consequently, the MP results of Y-STR typing could not effectively differentiate between two male individuals as aSTRs could, and aSTR typing is always preferred to solve forensic casework.

Alternatively, the assessment of the weight of a Y-STR match is feasible using a likelihood ratio approach. An additional advantage of the likelihood ratio method is the possible management of population subdivisions with an increased potential for co-ancestry. Accordingly, in the case of two Y-STR matches, it is recommended to carry out a quantitative assessment of the value of the match using relevant population/metapopulation data according to the national guidelines. The evaluation of the findings via the likelihood ratio (LR) approach needs the formulation of alternative hypotheses. The alternative hypotheses may state as follows:

Hypothesis 1: The source of the Y-haplotype is regarded as the known suspect.
Hypothesis 2: The source of the Y-haplotype is regarded as a random male Y from the reference population, and Y is another man, not the suspect.

The Y-STR profile detected in the crime stain is LR times more likely to be observed under hypothesis 1 than those under hypothesis 2. The remaining issue is the possibility that a male individual, with no close relation to the person of interest, has an identical Y-STR profile. This possibility can be evaluated by the application of the discrete Laplace (DL) method, approximating the proportion of haplotypes in the population/metapopulation to which the suspect belongs, with the same Y-haplotype. The DL method is a parametric statistical model applied for the estimation of the frequencies of Y-STR haplotypes considering a reference database.

A report based on quantitative assessment via the LR approach could be that suspect (known) could have contributed to the male source of the detected DNA. In addition, all male relatives on the paternal line and approximately 1 in 1,000 (i.e., likelihood ratio) unrelated males cannot be excluded (Taupin 2017). It is recommended to apply RM Y-STRs to further analyze evidence and known samples in case of a match for possible exclusion and reduce the number of individuals sharing the haplotype of the suspect.

National reference Y-STR databases are often established according to a historical concept of ethnic affiliation in countries, such as the USA, Brazil, the UK, or China, with a strong population substructure. Haplotype diversity is considered a very applicable indicator for the evaluation of the effect of population substructure. The future expansion of national and global Y-STR haplotype databases of much more loci included in a typical haplotype could help to draw more informative conclusions regarding Y-STR analyses and interpretations. Needless to say, the guidelines should be developed by laboratories for the number of Y-STR loci applied in searching population databases. If there is limited or irrelevant population data for the assessment of a Y-STR match, a qualitative statement could be reported in case the issue of

male relatedness is properly addressed in the final report. A conservative qualitative statement for a Y-STR match report between a sample recovered from a crime scene and a known suspect could be made in the following format:

> The Y-STR haplotype of the crime scene sample matches at all the examined loci to the Y-STR haplotype of the known individual (suspect). As a result, the suspect cannot be ruled out (excluded) as the donor of the sample recovered from the crime scene. Furthermore, all patrilineal-related male relatives in addition to an unknown number of unrelated males cannot be excluded as donors of the crime scene sample (Taupin 2017; Roewer et al. 2020).

Mixture Interpretation Via Y-STR Analysis

In addition to all the troubles that should be addressed in Y-STR interpretations, particular complications occur during Y-STR mixture interpretation. Mixture interpretation is defined as strategies applied to determine and separate possible genotype combinations of the contributors. Peak height variation across the whole Y-STR haplotype, differential degradation of the contributors to a mixture, and locus−/allele-specific mutation rates should be considered in mixture haplotype assessment. Rarely, it may be two or more detected peaks on DYS19, DYS390, and DYS391.

Awaiting the phenomenon of duplication or rare triplication of these Y-STR loci which may pass across generations, for the exact interpretation of mixture, a conservative strategy based on testing additional Y-STR loci and consideration of the entire Y-STR haplotype are required while determining the sample as a mixture (Butler 2014; SWGDAM 2017). Additionally, a combination of Y-STR results and even partial aSTR data may improve the rarity of a match. The combination of autosomal and Y-STR profiling results needs a consistent method for the addressed statistical question. Based on the different natures of underlying population structures, it is concluded that a simple combination of the genetic data acquired from the Y-STRs with data derived from the aSTRs as a single likelihood ratio is debatable and necessary to be applied in a conservative way (Amorim 2008).

In contrast, separate reporting of the autosomal and Y-STR results seems to be a more appropriate method of avoiding misunderstanding at the court scene. A conservative way to estimate the joint probability, under the assumption of negligible dependencies, is achieved by multiplying the highest value of the group of autosomal match probabilities by the calculated matching probability for the Y-STRs. The MPS empowers forensic DNA laboratories to overcome the limitations related to the CE and current STR typing techniques. Several sequence variants with similar core sequence lengths of the Y-STR typing products were noticed using MPS which could not be determined via CE-based typing methods, thereby further empowering them for forensic identification. In addition, MPS also provides a powerful automation capability for integrating STR genotyping with the detection of some other associated forensic markers in the same reaction (Hammer and Redd 2005; Zhao et al. 2015).

Once more, it is worth mentioning that the multiplicity of forensic casework and associated potential complexities result in the need for professional judgment and expertise in terms of the Y-STR interpretation. A fixed regulation cannot or should not comply with every situation. Universal guidelines only help to plan a standard scientific framework supported by the analytical data and prevent laboratory report audiences (i.e., judiciary systems and clients) from any misconceptions and conclusions (SWGDAM 2014).

Forensic Use of Y-Chromosome Testing

Y-STR typing is used for the analysis of variations on the Y chromosome in nuclear DNA. The special feathers of the human Y chromosome and Y-STR specifications proposed them as master keys to handle several investigative problems and solve forensic genetic cold cases. The high degree of polymorphism declared in the Y-STRs analysis of human populations, their small-size PCR products (range: 100–400 bp), and their ability to multiplexing and automating are the most critical specifications turning them into being useful or sometimes markers of choice for forensic investigations.

Sexual Assault Investigation

One of the notable responsibilities in the life of a forensic genetic analyst is the examination of different types of evidence from crime scenes, especially sexual assault intimate samples consisting of one or more male perpetrators. Sexual assault is a ubiquitous crime occurring in every culture, in all levels of society, and every country around the world disproportionally influencing adolescents and young women. Instead of all other crime types with no statistically significant changes within 2017 to 2018, the frequency of rape or sexual assault in the USA doubled in this period (victimizations per 1,000 individuals with the age of 12 years or over). Based on the national US statistics, the general population consists of 49% males and 51% females. However, male offenders involve in 77% of violent incidents, and the frequency of violent incidents involving female offenders is only reported as 18%. This rate was 1.6 times higher than the frequency of males represented in the general population (49%) (Morgan et al. 2019).

The conventional forensic serological testing of sexual assault samples consists of body fluid identification by colorimetric or antigen-based protocols and/or microscopic examination for sperm. Most of the guidelines described labor-intensive and time-consuming methods yielding false-positive/false-negative results. Therefore, many laboratories looked for new robust and reliable methods to streamline the examination of sexual assault cases. These alternative methods could act as the gatekeepers for which the samples proceed to STR analysis. The SWGDAM proposed a direct-to-DNA Y-screening case approach as the next-generation body fluid identification technique.

Direct-to-DNA Y-screening is a direct-to-DNA workflow using quantitative PCR to quantify male DNA, applying the male DNA quantity rather than conventional serology to direct downstream DNA STR testing more sensitively and accurately in shorter turnaround times. The most critical aspect is to choose a kit that is sensitive enough to detect low amounts of male DNA in the presence of high levels of female DNA. The results of Y-screening may indicate that these types of samples might not be suitable for obtaining male aSTR profiles; however, they would be good candidates for Y-STR analysis (SWGDAM 2020).

Quantitative real-time PCR kits provided the capabilities of the calculation of the quantities of both male and female DNA, along with degradation and inhibition indices generally accepted for quantification pre-PCR step. It assisted the proper selection of an amplification strategy and maximizing the potential for the determination of the male component(s) of a mixture. A comparison of detection limits between the Quantifiler Trio® Human Male DNA Quantification Kit® (Thermo Fisher Scientific, USA) and Yfiler Plus® indicated the low likelihood of obtaining a Yfiler Plus profile appropriate for comparison objectives in case of no detectable male DNA with Quantifiler Trio. This may be a proper indicator for triaging samples to a further step using Yfiler Plus (SWGDAM 2014; Henry and Scandrett 2019).

Biological evidence resulting from sexual assault demonstrates multiple challenges the most important of which is the presence of small quantities of male (suspect) DNA occurring with a relatively high quantity of the victim's (mostly female) DNA. Due to the nature of the different samples, differential DNA extraction strategy may not be efficient in all cases. In addition, the general assessment of an aSTR mixture profile may unintentionally mask minor male alleles not immediately discernible from the alleles of the female complainant (i.e., the major contributor). This is more pronounced if the victim and suspect have been closely related.

Therefore, the development of male-specific markers targeting the male fraction of a mixed DNA sample to prohibit female DNA competition for reagents during the preferential PCR amplification or masking minor male alleles may be useful for saving the trace levels of male DNA and deconvolving the mixture DNA profiles, including a minimum of two individuals and at least one male. This is also used in cases with high amounts of female DNA and limited yields of the male DNA, such as azoospermia/vasectomized perpetrators, or in case of alleged digital penetration in which the epithelial cells of the perpetrator should be analyzed instead of sperm cells (Aramayo et al. 2007; McDonald et al. 2015).

In the vaginal cavity, there is a decrease in the number of sperm cells with time and mimic sperm cell negative samples in which no sperm cells can be detected by microscopic screening via Baecchi's staining or immunofluorescence method of Sperm Hy-Liter. It is affected by factors, such as the personal hygiene of the victim, natural drainage, menstruation, damage to the sperm cell membranes, and extended interval postcoital sampling. According to the results of a study carried out by Albani et al. on the background yields of male DNA in the vaginal cavity, it was indicated that even up to 6 days following intercourse and up to 44 h after a semen negative offense, full Y-STR haplotype could be successfully obtained. Nevertheless, individuals differed in time frames which may also rely on the cycle stage of the uterine

(Miller et al. 2011; Albani et al. 2018; Henry and Scandrett 2019). The analysis of recovered male DNA under scratching female fingernails following a violent attack showed that although the exogenous debris quickly decays over time, Y-STR typing could provide highly useful data to single out the male perpetrator(s) up to 5 h after scratching. The Y-STRs also provide male-specific genetic data in amylase positive (licking, biting, or kissing samples) determined on the victim's body and screened via saliva presumptive tests (Taupin 2017; Iuvaro et al. 2018). The interpretation of mixture aSTR profiles is an overwhelming duty of the forensic genetic expert, and sometimes it is needed to work with mixture analysis software. Different from the conventional approaches, Y-STRs may be preferred to aSTRs to determine whether a familial relationship between the victim and perpetrator shares many autosomal alleles. The Y-STR analysis works by the removal of an overriding female profile in case of multiple perpetrator rapes and helps to simplify the interpretation of male-male cases.

The application of Y-STR markers would be beneficial to the reliable exclusion of male suspects from involvement in a crime through non-matching Y-STR haplotypes. Furthermore, they are helpful for the identification of the paternal lineage to which a trace donor belongs in such a highly complicated forensic casework. Therefore, guidelines defining flowcharts, under which forensic samples are subjected to Y-STR typing, should be established by every laboratory. While expecting a combination of male and female DNA (e.g., vaginal swabs) or a sperm-negative sample, it is suggested that these samples should be conserved for further Y-STR typing (Taupin 2017; Roewer et al. 2020).

As a general flowchart, for the generation of an eligible profile of the male DNA from either a single-source DNA or mixtures, aSTR typing should always be selected as the priority. In the case of the inconclusiveness of autosomal findings for the male component, it is necessary to analyze the sample via standard Y-STR typing. Whenever the Y-STR profile of the sample was informative for an individual matching the questioned haplotype, sufficient amounts of remaining extracts would be subjected to the supplementary panel of RM Y-STR marker set for the third step to exclude available close relatives. Finally, a quantitative assessment using either the DL method or count estimates should be provided to support matching results (Roewer et al. 2020).

Sex Determination

Sex determination based on molecular techniques has been demonstrated to be preferable to any other present anthropological methods which are mostly based on the visual examination of skeletal morphology and required intact bone(s). There are molecular tools applied for sex determination among which sex determination tests on the basis of the amelogenin gene as a part of the majority of commercial PCR multiplex reaction kits have been extensively applied in forensic casework, DNA database, medical/archeological specimens, preimplantation, and prenatal diagnoses (Rezaei et al. 2017).

The amelogenin locus contributing to the tooth enamel matrix formation is encoded by two single-copy genes (i.e., *AMELX* and *AMELY*) located on the short arm of the X (Xp22.1–22.3) and Y (Yp11.2) chromosomes, respectively. The most frequently utilized primer set of amelogenin delimits a 6-bp deletion in the intron 1 of AMELX, producing 106 bp amplicon for X and 112 bp amplicons for the Y chromosome. Although amelogenin is an effective technique for sex-typing of the biological samples in most cases, in rare cases of chromosomal microdeletions or variations in primer binding sites, relying only on the use of amelogenin as the sole sex marker will not provide foolproof results due to the dropout phenomenon (i.e., null alleles).

There have been detected cases of amelogenin-negative males worldwide reported to be especially high in individuals of Indian origin. Among several well-known genetic mechanisms underlying *AMELY* dropout, including the Y-chromosome microdeletions of various sizes covering AMELY locus, there are more mutations in the primer binding site of *AMELY* allele shown to a lesser extent. From the medical perspective, *AMELY* expression was detected at only 10% of the amount of *AMELX*, and the medical investigation of two cases of AMELY deletion showed normal teeth, suggesting minimal or no effect of deletion on enamel formation (Jobling et al. 2007). In sexual assault cases and missing individual investigations, it is also the first and not foremost finding that the questioned sample belongs to a male or a female source. The false-negative results of gender determination for contributors of a biological sample and mixed interpretation of undergarments of victims in rape may be misinterpreted as a negative case or mistakenly identified as a female (Butler 2014; Borovko et al. 2015). The closest Y-STR locus to the amelogenin-Y (i.e., DYS458 and rarely DYS456) may also be missed from the Y-STR haplotype due to the deletion events on the short arm of the Y chromosome at the pericentromeric bands. A total of 5 different deletion classes among 45 *AMELY*-null males from 12 populations were reported by Jobling et al. in 2007, among whom several other Y-STR loci presenting within or close to the amelogenin (i.e., DYS570, DYS576, DYS458, DYS481, DYS449, DYS627, and DYS391) may include in the deleted part (Chen et al. 2014; Cheng et al. 2019; Roewer et al. 2020; Dash et al. 2020a). Therefore, it can be helpful to determine the relative position of loci on the Y chromosome.

There have been reports of the simultaneous AMELY and DYS458 null alleles due to the Yp11.2 deletion in various populations, with a much higher frequency in Sri Lankan population (8.3%) following Nepalese and Indian populations (6.5% and 0.23–3.2%, respectively). As experienced in the South Asian tsunami victim identification of 2004, higher rates of null amelogenin observed in the affected populations raised the need for additional sex determination techniques, such as the presence/absence of SRY (Jobling et al. 2007). One suggestion regarding Y-null amelogenin cases is the real-time PCR quantitation assay of Y chromosome-specific targets, as they are used to confirm the presence of male DNA in the questioned sample (Kumar et al. 2019).

The other known solution for incorrect sexual identification is provided by testing additional markers, such as SRY- or Y-specific markers, for accurate gender

determination. Therefore, they should be applied in multiplex PCR genotyping kits. Three Y-STRs (i.e., DYS391, DYS570, and DYS576) and one Y indel loci were chosen as alternative sex-determining markers by the new generations of autosomal multiplex kits, such as GlobalFiler™ (Thermo Fisher Scientific, USA), PowerPlex® Fusion 6C (Promega Corporation, USA), and Investigator 24plex GO! Kit (QIAGEN, Germany). As a result, in addition to the amelogenin loci, one Y-STR (i.e., DYS391) and one Y-indel locus (i.e., rs2032678) were recently incorporated in AmpflSTR Globalfiler™ (Thermo Fisher Scientific, USA) to reduce the risk of misidentification of male samples.

Y-indel marker (originally known as M175) is located on Yq11.221 with two different conditions. A repeat of five nucleotides (TTCTC) deleted or mostly inserted in this position (referred to one or two alleles, respectively) illustrates male samples. The application of these additional markers assists male gender confirmation in cases of null amelogenin (Dash et al. 2020a). The results of fetal gender determination during the third trimester of pregnancy using the Y-STR analysis of maternal plasma indicated the feasible application of Y-STR typing as an accurate gender determining test in order to identify the fetus gender (Aal-Hamdan et al. 2015).

Paternity Testing

The Y-STR haplotyping is also relevant to the paternity dispute investigations of male offspring, several types of patrilineal kinship allegations, and familial searching. In paternity testing, Y-STR haplotyping is especially appropriate in deficiency cases in which the putative father of a male child is not at hand for aSTR typing. In different scenarios of ambiguous paternity, in the case of a female with multiple male partners, it is desired to carry out the noninvasively prenatal paternity testing (as recently discussed for gender determination) through analyzing the circulating cell-free fetal DNA, retrieved from maternal plasma. The Y-STR haplotype investigation of DNA extracted from maternal plasma can effectively be used as an alternative for exclusion purposes. It can only be applied to mothers carrying a male fetus. The main limitation of this application is that the meaning of MP for two similar Y-haplotypes is observed for the alleged father and fetus. It could only be concluded that they belong to the same paternal lineage, and this alternative strategy should not be used in a population reported for a high rate of endogamy (Barra et al. 2015).

Disaster Victim Identification

The Y-STR markers can be beneficial in disaster victim identification. Routinely, the results of DNA typing from remained bodies should be compared to personal objects (e.g., hair brass, toothbrushes, and razor blades), biological father, and reference samples from the alleged male lineages. In these situations, Y-STR typing can increase the number of alternative reference samples, simplify the sorting of the

tissue samples that remained in mass disasters, and help the forensic expert to select possible distant male relatives as informative subsidiary reference samples (Fig. 1). However, as previously discussed, Y-STR kits are not appropriate enough for the identification of recovered male victims due to the fact that all the male relatives possess the identical YSTR haplotype (Kayser 2017; Ambers et al. 2018).

Historical Investigations

Although the main benefits of Y-STR typing in forensic casework are shown in case of extremely low amounts of male DNA in mixture (male(s)/female) samples recovered from crime scenes, Y-chromosome testing is also applicable to investigate various anthropological issues, such as human identification, familial relationships, historical events, and male migration and mating patterns in different geographical regions. For this reason, researchers should apply many different Y-chromosome markers, including Y-STRs, Y-SNPs, and Y-indels, to pursue the male lineages living in or migrating across their geographical region. Nevertheless, in order to enhance haplotype diversity and lineage resolution among populations, several Y-STR markers appear more appropriate than others.

By far, a global analysis of Y-chromosomal haplotype diversity for 23 STR loci was published in 2014 as the most extensive collection of Y-chromosomal STR haplotypes worldwide for one of the biggest studies on the Y-STR data. A total of 19,630 unrelated male Y-STR haplotypes obtained from 129 populations in 51 countries worldwide were collected from 84 participating laboratories. According to the results, DYS481 and DYS570 markers demonstrated the largest numbers of different alleles, out of the six Y-STR markers distinguishing PPY23 from Yfiler. The DYS643 marker was observed to be more variable among Africans but less variable in Native Americans of Latin America than that of other continental groups. Moreover, the highest number of null alleles was shown for the DYS448 locus. A large deletion was detected at Yp11.2 covering the AMELY region with DYS570, DYS576, DYS458, and DYS481 in nine samples (with Asian ancestry, from Singapore, Tamils, and British Asians mostly originating from or living in India (Purps et al. 2014).

Yuxiang Zhou et al. investigated the extent of genetic differentiation between populations (i.e., genetic distance) in over 20,000 haplotypes of Yfiler Plus set gathered from 41 global populations in 2020. They defined the mean of allele frequency difference (mAFD) as a reliable value for the estimation of the variance of genetic markers in populations and evaluation of population differentiation. The results of their study indicated that DYS392 and YGATAH4, among the studied Y-STRs, were reported with the largest (0.3802) and smallest (0.1845) mAFD values, respectively (Barra et al. 2015).

In another study carried out by Grugni et al., 10 Y-STR loci and 88 Y-chromosome biallelic markers were investigated in 15 ethnic groups of the Iranian population. The results showed that 65 different Y-chromosome lineages belong to 15 main haplogroups in which J (31.4%) and R (29.1%) are frequently observed in the

northwest, and G (11.8%) and E (9.2%) are most frequently observed due to different migratory events in the history. A clear African component (i.e., haplogroup E-M2) is observed in Hormozgan, Iran, indicating the presence of sub-Saharan Africans. Another Y-STR haplotype analysis of the Eastern Iranian population in comparison with haplotypes from the neighboring countries (Afghanistan and Pakistan) shows small genetic discrimination between neighbor populations living in this close region. Such historical-based population genetic studies applied reliable Y-chromosomal variation portrait to shed light on the genetic structures and expansion patterns to help reconstruct ancient migration routes (Grugni et al. 2012; Tabrizi et al. 2015).

Y-STR Haplotype Databases

The global utilization of Y-STR typing in forensic genetic laboratories raised the need for national/international Y-STR databases. As previously discussed, the quantitative evaluation of the informativeness of a match between two haplotypes requires the assessment of the haplotype rarity based on the frequency estimates of the Y-STR haplotype in a reliable, robust, and relevant database. Standing on the first step in 1997, the SWGDAM and European Scientific Committee announced a recommended minimal haplotype as a cornerstone for the establishment of large global databases. After several years, Y Chromosome Haplotype Reference Database (YHRD; available at http://www.yhrd.org/) and US Y-STR Database (available at http://www.usystrdatabase.org) are founded as the two most important Y-STR-based databases.

Lutz Roewer et al. created the largest and most extensively utilized Y-STR database in 1999, currently announcing the 63rd release. Over 321,000 minimal haplotypes of anonymous male individuals have been imported into YHRD 3.0 since 2000 up to November 2020. Complementary to the Y-STR haplotypes, YHRD has been enlarged with Y-SNP data and at present includes 25,000 Y-SNP data. The US Y-STR Database has been founded under the management of the US National Center for Forensic Science and financial supports of the National Institute of Justice since 2007. The US Y-STR Database consists of 35,658 Y-STR haplotypes generated by 30 different forensic, academic, and commercial laboratories since 2018 (release 4.2). It has been permanently transferred to YHRD for the continuation of usage and finally was decommissioned after June 30, 2019 (SWGDAM 2014).

In addition to forensic Y-STR databases, several Y-STR-based genealogy databases were created to aid forensic investigators to link between a particular Y-STR haplotype and family surnames in most human societies, where the male surnames are paternally co-inherited. A surname is a part of a personal name demonstrating the family, tribe, or community the individual belongs to. Despite the observation of strong co-ancestry for several rare surnames, common surnames showed low or no Y-chromosome correlations restricting the predictive value of Y-STR haplotyping for common surnames. King et al. in their study examined unrelated males accounting for 150 different British surnames. In this study, the prediction of a correct surname was observed in only 19% of the cases; nevertheless, the sensitivity of

prediction increased to 34% for less common surnames (Kayser 2017; Claerhout et al. 2020).

Another population genetic analysis conducted on the Y-STR haplotypes of the five most common surnames in South Korea indicated limited genetic variations among five surnames due to the genetic heterogeneity of each surname. In addition, various kinds of non-surname-related haplotypes were distributed among five surnames. Therefore, the use of surname-based prediction of Y-STR haplotype data for narrowing down the suspect searching domain has not been adequately effective (Kim et al. 2009).

For simplicity, to increase the advantages of the Y-STR database, forensic genetic society should follow the common motivating quote of the bigger the better.

Despite the importance of increasing the size of reference databases that are used to estimate the Y-STR match probabilities, it is important to note some population highlights. As the diversity of Y-STR haplotypes in small, isolated, and migrant populations decreases, estimated discriminatory power reduces. Simultaneously, the presence of subpopulations in the database with distinct geographical distribution patterns, different from those of the full population profiles, may result in biased frequency estimations for the Y-STR haplotypes. It was more obvious while the assessment of partial Y-STR haplotypes as evidence in forensic casework (Amorim and Budowle 2016).

One solution for the effective application of Y-STR typing and Y-based database construction in populations with the aforementioned specifications (e.g., the Finnish population) might be the inclusion of more Y-STR loci based on national standard protocols (Palo et al. 2008). Another issue may arise when a Y-STR database containing different Y-STR sets of loci applied for Y-haplotype frequency discrimination is a paradoxical phenomenon. For clarity, if highly discriminating Yfiler Plus data were searched against even in a large database with 9000 minimal haplotypes and 1000 records of 27 loci haplotypes, a full match means 10 times less discrimination estimated (1/1000) for a more discriminative 27 loci set than a minimal set (1/10000) in terms of the frequency of matching haplotypes per number of compared profiles.

The simple explanation is fewer database records of highly discriminating Y-haplotypes. Therefore, in addition to great efforts by the forensic genetics community that are currently underway to extend sample sizes for populations through including extended Y-STR haplotype records, it is necessary to provide guidelines for the number of Y-STR loci used for searching population databases. The application of reduced locus set criteria is of great importance when a partial Y-STR haplotype resulted from the evidence of a crime scene (SWGDAM 2014).

Familial Searching

The Y-STR data inherited in a patrilineal way result in the appropriateness of Y-STR haplotyping for familial searching in forensic cases when a newly submitted autosomal DNA profile does not have a full match. It helps to discriminate among male

offender(s) via searching for his male relatives which is called forensic familial searching. When multiple Y-STR haplotypes are consistent with the Y-STR haplotype of a lineage, Y-STR profiling could help in narrowing the extent of searching and accelerating the investigation. One critical point in successful familial searching is the appropriate selection of different Y-STR panels in developing and filtering the search of Y-STR haplotype databases. A filtered panel of slow/moderate mutation rate Y-STRs is a selective searching strategy to boost the chance of finding distant relatives; however, RM Y-STRs are more appropriate to discriminate among close relatives (Claerhout et al. 2019).

Apart from the great extra advantages of familial searching in sexual offense investigations to solve cold cases, some ethical, social, and legal issues are raised and needed to be legislated before familial searching and even before establishing a database plan. In this regard, the extra Y-STR testing of the forensic samples stored in the national DNA database and then expanded voluntary DNA mass screenings require much more budget. As the nature of Y-chromosome data sharing among paternal relatives, they belong not only to the suspect but also his innocent relatives (Fig. 1). On the other hand, accumulating much more genetic data essentially means further requirements for strict security protocols. In countries with the regular inclusion of Y-haplotypes in their national DNA databases (e.g., Austria), with serious crimes leading to long imprisonment, a minimum of 27 Y-STRs are incorporated in the commercially available Yfiler® Plus kit® which could be extended to 46 Y-STR markers used to provide a lower number of false-negative surname matches and prevent false-positive errors (Claerhout et al. 2020).

Haplogroup Prediction

The Y-haplogroups are determined by the pattern of SNPs which can also be analyzed by direct testing of SNPs. The outcomes of haplogroup determination are beneficial for the assessment of genealogical relationships between two or more males. However, the process of haplogroup determination can sometimes be expensive and long. Therefore, there is considerable interest in the prediction of the Y-chromosome haplogroup, as a group or family of Y-chromosomes related by descent from a set of Y-STR markers. The Y-haplogroup predictors are considered tools extensively applied to have access to haplogroups based on Y-STR values. Needless to say, the prediction of haplogroups using predictors requires slower mutating markers and knowledge of the allele frequencies for each haplogroup (Athey 2005).

Several approaches implemented in Y-haplogroup predictors differ in strengths and limitations. The prediction algorithm applied by Family Tree DNA (FTDNA) in collaboration with the University of Arizona (UAZ), USA, is on the basis of the genetic distance of the Y-haplotype in question to other Y-haplotypes with a previously determined haplogroup. Although the FTDNA/UAZ approach claimed to be successful for 80% of customers, the main limitation is indicated in case the absence of haplogroup-confirmed haplotype is observed in the database within a threshold distance; accordingly, there is no estimate of a haplogroup.

Another approach adopted on a website since 2004 is based on the allele frequency estimations for each haplogroup and the score in which a test haplotype fits the pattern of alleles in each haplogroup. The investigator could enter any numbers of the FTDNA set (out of including 37 markers), and the program estimates goodness of fit scores for 10 haplogroups (i.e., E3a, E3b, G, I1a, I1b, I1c, J2, N3, R1a, and R1b). Whit Athey's predictor was developed by Whit Athey as the first publicly available haplogroup prediction tool.

Nevgen Y-DNA haplogroup predictor applies the Bayesian allele frequency approach to estimate the intended haplogroup of a Y-STR haplotype, utilized by Athey's haplogroup predictor. Predictions based on this tool for the abovementioned markers are usually elaborated by two values, including probability (initial 100% probability of an individual belonging to specific haplogroup (as a numerator) divided into all possible Y-haplogroups) and fitness score which should be assumed different for various haplogroups (Szargut et al. 2019). Diepenbroek et al. investigated 23 markers of the human Y chromosome in remains exhumed from historical Nazi-occupied regions located in Białystok (i.e., Eastern Poland). The Y-DNA haplogroup prediction by Nevgen revealed that over 80% of the studied samples were suggested with European origin and typical haplogroup for the Polish population other than one ethnic minority (Diepenbroek et al. 2019).

Bouakaze et al. developed PredYMaLe as a versatile machine learning program for the prediction of Y-haplogroups based on 32 Y-STR multiplex which is mutationally balanced in 2020 (Bouakaze et al. 2020). Emmerova et al. compared several Y-chromosomal haplogroup predictors. The final ranking results of software packages based on the number of non-concordant estimates (the lower the better in only 19 STRs) is presented as Whit Athey's HAPEST, Nevgen, YPredictor (by Vadim Urasin), Jim Cullen's World Haplogroup, and Haplo-I Subclade Predictor. In the aforementioned study, it was concluded that minimally a Y-haplotype should consist of more than 12 Y-STRs; otherwise, it should not be applied for the accurate prediction of the haplogroup. It is suggested that the integration of further predictors reduces the risk of false haplogroup assignment. For more details, Y-haplogroup predictors are available at the following sites:

Whit Athey: Haplogroup predictor-HAPEST available (www.hprg.com/hapest5)
Jim Cullen: World Haplogroup and Haplo-'I' Subclade Predictor (members.bex.net/jtcullen515/haplotest. htm)
Nevgen Y-DNA haplogroup predictor (www.nevgen.org)
Vadim Urasin: YPredictor (predictor.yDNA.ru) (www.y-str.org/2013/06/y-haplogroup-predictor.hTMl) (Emmerova et al. 2017)

Conclusion

Forensic genetic experts benefited from the Y-chromosome analysis in different forensic scenarios of sexual assault investigation and disaster victim identification to paternity dispute analysis and historical investigations. Apart from the advantages

of Y-chromosome marker analysis, several technical limitations shall be weighed and addressed in standard interpretations, reporting, and application guidelines.

References

Aal-Hamdan HMH, Refaat AM, Babu SR, Choudhry AR (2015) Fetal gender determination through Y-STR analysis of maternal plasma during the third trimester of pregnancy. Egypt J Med Hum Genet 16(1):47–53. https://doi.org/10.1016/j.ejmhg.2014.08.005. (Aal-Hamdan 2015)

Adnan A, Ralf A, Rakha A, Kousouri N, Kayser M (2016) Improving empirical evidence on differentiating closely related men with RM Y-STRs: a comprehensive pedigree study from Pakistan. Forensic Sci Int Genet 25:45–51. https://doi.org/10.1016/j.fsigen.2016.07.005

Albani PP, Patel J, Fleming RI (2018) Background levels of male DNA in the vaginal cavity. Forensic Sci Int Genet 33:110–116. https://doi.org/10.1016/j.fsigen.2017.12.003

Alghafri R, Goodwin W, Ralf A, Kayser M, Hadi S (2015) A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. Forensic Sci Int Genet 17:91–98. https://doi.org/10.1016/j.fsigen.2015.04.004

Almohammed E, Hadi S (2019) A study of genetic analysis using novel rapidly mutating Y-STR multiplex for Qatari population. Forensic Sci Int 7(1):875–877. https://doi.org/10.1016/j.fsigss.2019.11.008

Ambers A, Votrubova J, Vanek D, Sajantila A, Budowle B (2018) Improved Y-STR typing for disaster victim identification, missing persons investigations, and historical human skeletal remains. Int J Legal Med 132(6):1545–1553. https://doi.org/10.1007/s00414-018-1794-8

Amorim A (2008) A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. Forensic Sci Int Genet 2(4):376–378. https://doi.org/10.1016/j.fsigen.2008.04.001

Amorim A, Budowle B (2016) Handbook of forensic genetics: biodiversity and heredity in civil and criminal investigation. Imperial College Press

Andersen MM, Balding DJ (2017) How convincing is a matching Y-chromosome profile? PLoS Genet 13(11):e1007028. https://doi.org/10.1371/journal.pgen.1007028

Aramayo R, Hanson EK, Ballantyne J (2007) An ultra-high discrimination Y chromosome Short tandem repeat multiplex DNA typing system. PLoS One 2(8). https://doi.org/10.1371/journal.pone.0000688

Athey TW (2005) Haplogroup prediction from Y-STR values using an allele frequency approach. J Genet Geneal 1:1–7

Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Brauer S (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 87(3):341–353. https://doi.org/10.1016/j.ajhg.2010.08.006

Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A, Vermeulen M, de Knijff P, Kayser M (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6(2):208–218. https://doi.org/10.1016/j.fsigen.2011.04.017

Ballantyne KN, Ralf A, Aboukhalid R, Achakzai NM, Anjos MJ, Ayub Q, Balažic J, Ballantyne J, Ballard DJ, Berger B, Bobillo C, Bouabdellah M, Burri H, Capal T, Caratti S, Cárdenas J, Cartault F, Carvalho EF, Carvalho M, Cheng B, Coble MD, Comas D, Corach D, D'Amato ME, Davison S, de Knijff P, De Ungria MC, Decorte R, Dobosz T, Dupuy BM, Elmrghni S, Gliwiński M, Gomes SC, Grol L, Haas C, Hanson E, Henke J, Henke L, Herrera-Rodríguez F, Hill CR, Holmlund G, Honda K, Immel UD, Inokuchi S, Jobling MA, Kaddura M, Kim JS, Kim SH, Kim W, King TE, Klausriegler E, Kling D, Kovačević L, Kovatsi L, Krajewski P, Kravchenko S, Larmuseau MH, Lee EY, Lessig R, Livshits LA, Marjanović D, Minarik M, Mizuno N, Moreira H, Morling N, Mukherjee M, Munier P, Nagaraju J, Neuhuber F, Nie S, Nilasitsataporn P, Nishi T, Oh HH, Olofsson J, Onofri V, Palo JU, Pamjav H, Parson W, Petlach M, Phillips C, Ploski R, Prasad SP, Primorac D, Purnomo GA,

Purps J, Rangel-Villalobos H, Rębała K, Rerkamnuaychoke B, Gonzalez DR, Robino C, Roewer L, Rosa A, Sajantila A, Sala A, Salvador JM, Sanz P, Schmitt C, Sharma AK, Silva DA, Shin KJ, Sijen T, Sirker M, Siváková D, Skaro V, Solano-Matamoros C, Souto L, Stenzl V, Sudoyo H, Syndercombe-Court D, Tagliabracci A, Taylor D, Tillmar A, Tsybovsky IS, Tyler-Smith C, van der Gaag KJ, Vanek D, Völgyi A, Ward D, Willemse P, Yap EP, Yong RY, Pajnič IZ, Kayser M (2014) Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. Hum Mutat 35(8):1021–1032. https://doi.org/10.1002/humu.22599

Barra GB, Santa Rita TH, Chianca CF, Velasco LF, de Sousa CF, Nery LF, Costa SS (2015) Fetal male lineage determination by analysis of Y-chromosome STR haplotype in maternal plasma. Forensic Sci Int Genet 15:105–110. https://doi.org/10.1016/j.fsigen.2014.11.006

Barros B, Morais M, Teixeira AL, Medeiros R (2020) Loss of chromosome Y and its potential applications as biomarker in health and forensic sciences. Cytogenet Genome Res 160(5):225–237. https://doi.org/10.1159/000508564

Borovko S, Shyla A, Korban V, Borovko A (2015) Amelogenin test abnormalities revealed in Belarusian population during forensic DNA analysis. Forensic Sci Int Genet 15:98–104. https://doi.org/10.1016/j.fsigen.2014.10.014

Bouakaze C, Delehelle F, Saenz-Oyhéréguy N, Moreira A, Schiavinato S, Croze M, Delon S, Fortes-Lima C, Gibert M, Bujan L, Huyghe E, Bellis G, Calderon R, Hernández CL, Avendaño-Tamayo E, Bedoya G, Salas A, Mazières S, Charioni J, Migot-Nabias F, Ruiz-Linares A, Dugoujon J-M, Thèves C, Mollereau-Manaute C, Noûs C, Poulet N, King T, D'Amato ME, Balaresque P (2020) Predicting haplogroups using a versatile machine learning program (PredYMaLe) on a new mutationally balanced 32 Y-STR multiplex (CombYplex): unlocking the full potential of the human STR mutation rate spectrum to estimate forensic parameters. Forensic Sci Int Genet 48. https://doi.org/10.1016/j.fsigen.2020.102342

Butler JM (2003) Recent developments in Y-Short tandem repeat and Y-single nucleotide polymorphism analysis. Forensic Sci Rev 15(2):91–111

Butler JM (2014) Advanced topics in forensic DNA typing: interpretation. Elsevier Science. ISBN: 9780124058545

Chen W, Wu W, Cheng J, Zhang Y, Chen Y, Sun H (2014) Detection of the deletion on Yp11.2 in a Chinese population. Forensic Sci Int Genet 8(1):73–79. https://doi.org/10.1016/j.fsigen.2013.07.003

Cheng JB, Liu Q, Long F, Huang DX, Yi SH (2019) Analysis of the Yp11.2 deletion region of phenotypically Normal males with an AMELY-null allele in the Chinese Han population. Genet Test Mol Biomarkers 23(5):359–362. https://doi.org/10.1089/gtmb.2018.0231

Claerhout S, Van der Haegen M, Vangeel L, Larmuseau MHD, Decorte R (2019) A game of hide and seq: identification of parallel Y-STR evolution in deep-rooting pedigrees. Eur J Hum Genet 27(4):637–646. https://doi.org/10.1038/s41431-018-0312-2

Claerhout S, Roelens J, Van der Haegen M, Verstraete P, Larmuseau MHD, Decorte R (2020) Ysurnames? The patrilineal Y-chromosome and surname correlation for DNA kinship research. Forensic Sci Int Genet 44:102204. https://doi.org/10.1016/j.fsigen.2019.102204

Dash HR, Rawat N, Das S (2020a) Alternatives to amelogenin markers for sex determination in humans and their forensic relevance. Mol Biol Rep 47(3):2347–2360. https://doi.org/10.1007/s11033-020-05268-y (a)

Dash HR, Shrivastava P, Das S (2020b) Principles and practices of DNA analysis: a laboratory manual for forensic DNA typing, Springer protocols handbook. Springer. ISBN: 107160273X, 9781071602737 (b)

Della Rocca C, Alladio E, Barni F, Cannone F, D'Atanasio E, Trombetta B, Berti A, Cruciani F (2019) Low discrimination power of the YFILER™ Plus PCR amplification kit in African populations. Do we need more RM Y-STRs? Forensic Sci Int 7(1):671–673. https://doi.org/10.1016/j.fsigss.2019.10.133

Diepenbroek M, Cytacka S, Szargut M, Arciszewska J, Zielinska G, Ossowski A (2019) Analysis of male specific region of the human Y chromosome sheds light on historical events in Nazi

occupied eastern Poland. Int J Legal Med 133(2):395–409. https://doi.org/10.1007/s00414-018-1943-0

Emmerova B, Ehler E, Comas D, Votrubova J, Vanek D (2017) Comparison of Y-chromosomal haplogroup predictors. Forensic Sci Int 6:e145–e147. https://doi.org/10.1016/j.fsigss.2017.09.025

Federal Bureau of Investigation Forensic Biology Manager (2020) FB Y-STR Interpretation Guidelines. Available at: https://www.pbso.org/qualtrax

Ferreira-Silva B, Fonseca-Cardoso M, Porto MJ, Magalhaes T, Caine L (2018) A comparison among three multiplex Y-STR profiling kits for sexual assault cases. J Forensic Sci 63(6):1836–1840. https://doi.org/10.1111/1556-4029.13757

Gill P et al (2001) DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. Int J Legal Med 114:305–309

Goodwin W (2016) Forensic DNA typing protocols. Springer, New York. ISBN: 1493935976, 9781493935970

Gopinath S, Zhong C, Nguyen V, Ge J, Lagace RE, Short ML, Mulero JJ (2016) Developmental validation of the Yfiler((R)) Plus PCR amplification kit: an enhanced Y-STR multiplex for casework and database applications. Forensic Sci Int Genet 24:164–175. https://doi.org/10.1016/j.fsigen.2016.07.006

Grugni V, Battaglia V, Hooshiar Kashani B, Parolo S, Al-Zahery N, Achilli A, Olivieri A, Gandini F, Houshmand M, Sanati MH, Torroni A, Semino O (2012) Ancient migratory events in the Middle East: new clues from the Y-chromosome variation of modern Iranians. PLoS One 7(7):e41252. https://doi.org/10.1371/journal.pone.0041252

Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM, International Society of Forensic G (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int J Legal Med 120(4):191–200. https://doi.org/10.1007/s00414-005-0026-1

Hammer M, Redd AJ (2005) Forensic applications of Y Chromosome STRs and SNPs. Available at http://www.ncjrs.gov/App/publications/abstract.aspx?ID=233445

Hampikian G, Peri G, Lo SS, Chin MH, Liu KL (2017) Case report: coincidental inclusion in a 17-locus Y-STR mixture, wrongful conviction and exoneration. Forensic Sci Int Genet 31:1–4. https://doi.org/10.1016/j.fsigen.2017.08.004

Hanson EK, Ballantyne J (2006) Comprehensive annotated STR physical map of the human Y chromosome: forensic implications. Leg Med (Tokyo) 8(2):110–120. https://doi.org/10.1016/j.legalmed.2005.10.001

He G, Wang Z, Su Y, Zou X, Wang M, Chen X, Gao B, Liu J, Wang S, Hou Y (2019) Genetic structure and forensic characteristics of Tibeto-Burman-speaking U-Tsang and Kham Tibetan highlanders revealed by 27 Y-chromosomal STRs. Sci Rep 9(1):7739. https://doi.org/10.1038/s41598-019-44230-2

Henry J, Scandrett L (2019) Assessment of the Yfiler(R) Plus PCR amplification kit for the detection of male DNA in semen-negative sexual assault cases. Sci Justice 59(5):480–485. https://doi.org/10.1016/j.scijus.2019.05.001

Henry J, Simon C, Linacre A (2015) The benefits and limitations of expanded Y-chromosome short tandem repeat (Y-STR) loci. Forensic Sci Int 5:e28–e30. https://doi.org/10.1016/j.fsigss.2015.09.012

Iuvaro A, Bini C, Dilloo S, Sarno S, Pelotti S (2018) Male DNA under female fingernails after scratching: transfer and persistence evaluation by RT-PCR analysis and Y-STR typing. Int J Legal Med 132(6):1603–1609. https://doi.org/10.1007/s00414-018-1839-z

Jobling MA, Lo IC, Turner DJ, Bowden GR, Lee AC, Xue Y, Carvalho-Silva D, Hurles ME, Adams SM, Chang YM, Kraaijenbrink T, Henke J, Guanti G, McKeown B, van Oorschot RA, Mitchell RJ, de Knijff P, Tyler-Smith C, Parkin EJ (2007) Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Hum Mol Genet 16(3):307–316. https://doi.org/10.1093/hmg/ddl465

Kayser M (2017) Forensic use of Y-chromosome DNA: a general overview. Hum Genet 136(5): 621–635. https://doi.org/10.1007/s00439-017-1776-9

Kayser M, Sajantila A (2001) Mutations at Y-STR loci: implications for paternity testing and forensic analysis. Forensic Sci Int 118(2–3):116–121. https://doi.org/10.1016/s0379-0738(00)00480-1

Kim SH, Jung J, Cho NS, Hwang JH, Yoo SY, Seong KM, Han MS, Kim W (2009) Y-STR genetic structure of the most common surnames in Korea. Genes Genomics 31(3):243–249. https://doi.org/10.1007/BF03191196

Kumar N, Chauhan A, Guptha R, Maitray A, Shukla SK (2019) Deletion of Y-allele in autosomal STR and role of Y-STR. J Forensic Sci Digit Investig 2019:50–54. https://doi.org/10.29199/2637-7136/FSDI-201020

McDonald A, Jones E, Lewis J, O'Rourke P (2015) Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa. Forensic Sci Int Genet 15:84–89. https://doi.org/10.1016/j.fsigen.2014.10.015

Mertoglu E, Filoglu G, Zorlu T, Bulbul O (2018) Estimation of the Y-chromosomal short tandem repeat (Y-STR) mutation rates in Turkey. Turk J Biochem 43(2):142–150. https://doi.org/10.1515/tjb-2017-0240

Miller KW, Old J, Fischer BR, Schweers B, Stipinaite S, Reich K (2011) Developmental validation of the SPERM HY-LITER kit for the identification of human spermatozoa in forensic samples. J Forensic Sci 56(4):853–865. https://doi.org/10.1111/j.1556-4029.2011.01796.x

Morgan R, Oudekerk BA, BJS Statisticians (2019) Criminal Victimization, 2018. NCJ 253043. Available at https://www.bjs.gov/index.cfm?ty=pbdetail&iid=6686

Muhanned AK, Ameera OH, Imad HH (2015) Y-chromosome short tandem repeat, typing technology, locus information and allele frequency in different population: a review. Afr J Biotechnol 14(27):2175–2178. https://doi.org/10.5897/ajb2015.14457

North Carolina Office of Indigent Defense Services Forensic Biology Forensic Scientist Manager and DNA Technical Leader. (2016) Procedure for Y23 interpretation, forensic biology section. Available at: http://www.ncids.com/forensic/labs/Biology/Technical/DNACaseworK

Nunez C, Baeta M, Ibarbia N, Ortueta U, Jimenez-Moreno S, Blazquez-Caeiro JL, Builes JJ, Herrera RJ, Martinez-Jarreta B, de Pancorbo MM (2017) 17 to 23: a novel complementary mini Y-STR panel to extend the Y-STR databases from 17 to 23 markers for forensic purposes. Electrophoresis 38(7):1016–1021. https://doi.org/10.1002/elps.201600313

Owers R, McDonald A, Montgomerie H, Morse C (2018) A casework study comparing success rates and expectations of detecting male DNA using two different Y-STR multiplexes on vaginal swabs in sexual assault investigations where no semen has been detected. Forensic Sci Int Genet 37:1–5. https://doi.org/10.1016/j.fsigen.2018.07.016

Palo JU, Pirttimaa M, Bengs A, Johnsson V, Ulmanen I, Lukka M, Udd B, Sajantila A (2008) The effect of number of loci on geographical structuring and forensic applicability of Y-STR data in Finland. Int J Legal Med 122(6):449–456. https://doi.org/10.1007/s00414-008-0250-6

PowerPlex Y23 System Technical Manual (2012). Available at: http://www.promega.com/resources/protocols/technical-manuals/101/powerplex-y23-system-protocol/. Accessed 10 Ded 2012

Promega Corporation: PowerPlex® Y system technical manual. (2003)

Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, Angustia SM, Santos LH, Anslinger K, Bayer B, Ayub Q, Wei W, Xue Y, Tyler-Smith C, Bafalluy MB, Martinez-Jarreta B, Egyed B, Balitzki B, Tschumi S, Ballard D, Court DS, Barrantes X, Bassler G, Wiest T, Berger B, Niederstatter H, Parson W, Davis C, Budowle B, Burri H, Borer U, Koller C, Carvalho EF, Domingues PM, Chamoun WT, Coble MD, Hill CR, Corach D, Caputo M, D'Amato ME, Davison S, Decorte R, Larmuseau MH, Ottoni C, Rickards O, Lu D, Jiang C, Dobosz T, Jonkisz A, Frank WE, Furac I, Gehrig C, Castella V, Grskovic B, Haas C, Wobst J, Hadzic G, Drobnic K, Honda K, Hou Y, Zhou D, Li Y, Hu S, Chen S, Immel UD, Lessig R, Jakovski Z, Ilievska T, Klann AE, Garcia CC, de Knijff P, Kraaijenbrink T, Kondili A, Miniati P,

Vouropoulou M, Kovacevic L, Marjanovic D, Lindner I, Mansour I, Al-Azem M, Andari AE, Marino M, Furfuro S, Locarno L, Martin P, Luque GM, Alonso A, Miranda LS, Moreira H, Mizuno N, Iwashima Y, Neto RS, Nogueira TL, Silva R, Nastainczyk-Wulf M, Edelmann J, Kohl M, Nie S, Wang X, Cheng B, Nunez C, Pancorbo MM, Olofsson JK, Morling N, Onofri V, Tagliabracci A, Pamjav H, Volgyi A, Barany G, Pawlowski R, Maciejewska A, Pelotti S, Pepinski W, Abreu-Glowacka M, Phillips C, Cardenas J, Rey-Gonzalez D, Salas A, Brisighelli F, Capelli C, Toscanini U, Piccinini A, Piglionica M, Baldassarra SL, Ploski R, Konarzewska M, Jastrzebska E, Robino C, Sajantila A, Palo JU, Guevara E, Salvador J, Ungria MC, Rodriguez JJ, Schmidt U, Schlauderer N, Saukko P, Schneider PM, Sirker M, Shin KJ, Oh YN, Skitsa I, Ampati A, Smith TG, Calvit LS, Stenzl V, Capal T, Tillmar A, Nilsson H, Turrina S, De Leo D, Verzeletti A, Cortellini V, Wetton JH, Gwynne GM, Jobling MA, Whittle MR, Sumita DR, Wolanska-Nowak P, Yong RY, Krawczak M, Nothnagel M, Roewer L (2014) A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet 12:12–23. https://doi.org/10.1016/j.fsigen.2014.04.008

Qian X, Hou J, Wang Z, Ye Y, Lang M, Gao T, Liu J, Hou Y (2017) Next generation sequencing plus (NGS+) with Y-chromosomal markers for forensic pedigree searches. Sci Rep 7(1):11324. https://doi.org/10.1038/s41598-017-11955-x

Ralf A, Lubach D, Kousouri N, Winkler C, Schulz I, Roewer L, Purps J, Lessig R, Krajewski P, Ploski R, Dobosz T, Henke L, Henke J, Larmuseau MHD, Kayser M (2020) Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers. Hum Mutat 41(9):1680–1696. https://doi.org/10.1002/humu.24068

Rezaei E, Beiraghi-Toosi A, Ahmadabadi A, Tavousi SH, Alipour Tabrizi A, Fotuhi K, Jabbari Nooghabi M, Manafi A, Ahmadi Moghadam S (2017) Can skin allograft occasionally act as a permanent coverage in deep burns? A pilot study. World J Plast Surg 6(1):94–99

Roewer L, Amemann J, Spurr NK, Grzeschik KH, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89(4):389–394. https://doi.org/10.1007/BF00194309

Roewer L, Andersen MM, Ballantyne J, Butler JM, Caliebe A, Corach D, D'Amato ME, Gusmao L, Hou Y, de Knijff P, Parson W, Prinz M, Schneider PM, Taylor D, Vennemann M, Willuweit S (2020) DNA commission of the International Society of Forensic Genetics (ISFG): recommendations on the interpretation of Y-STR results in forensic analysis. Forensic Sci Int Genet 48: 102308. https://doi.org/10.1016/j.fsigen.2020.102308

SWGDAM (2014) SWGDAM Interpretation Guidelines for Y-Chromosome STR Typing by Forensic DNA Laboratories. Available at https://www.swgdam.org/publications

SWGDAM (2017) SWGDAM Interpretation Guidelines for Autosomal STR typing by forensic DNA testing laboratories. Available at https://www.swgdam.org/publications

SWGDAM (2020) Scientific Working Group on DNA analysis methods next generation body fluid identification working group report on Y-screening of Sexual Assault Evidence Kits (SAEKs). Available at https://www.swgdam.org/publications

Szargut M, Diepenbroek M, Zielinska G, Cytacka S, Arciszewska J, Jalowinska K, Piatek J, Ossowski A (2019) Is MPS always the answer? Use of two PCR-based methods for Y-chromosomal haplotyping in highly and moderately degraded bone material. Forensic Sci Int Genet 42:181–189. https://doi.org/10.1016/j.fsigen.2019.07.016

Tabrizi AA, Hedjazi A, Kerachian MA, Honarvar Z, Dadgarmoghaddam M, Raoofian R (2015) Genetic profile of 17 Y-chromosome STR haplotypes in East of Iran. Forensic Sci Int Genet 14: e6–e7. https://doi.org/10.1016/j.fsigen.2014.10.010

Taupin JM (2017) Introduction to forensic DNA evidence for criminal justice professionals. Taylor & Francis. ISBN: 9781439899106

Thompson JM, Ewing MM, Frank WE, Pogemiller JJ, Nolde CA, Koehler DJ, Shaffer AM, Rabbach DR, Fulmer PM, Sprecher CJ, Storts DR (2013) Developmental validation of the PowerPlex® Y23 system: a single multiplex Y-STR analysis system for casework and database samples. Forensic Sci Int Genet 7(2):240–250. https://doi.org/10.1016/j.fsigen.2012.10.013

Tootkaboni GHM (2017) The rose garden: a new translation of Gulistan from the Persian. Tilde Publishing and Distribution. ISBN: 073467001X, 9780734670014

Wen SQ, Yao HB, Du PX, Wei LH, Tong XZ, Wang LX, Wang CC, Zhou BY, Shi MS, Zhabagin M, Wang J, Xu D, Jin L, Li H (2019) Molecular genealogy of Tusi Lu's family reveals their paternal relationship with Jochi, Genghis Khan's eldest son. J Hum Genet 64(8):815–820. https://doi.org/10.1038/s10038-019-0618-0

Yfiler® Plus PCR Amplification Kit, Rev 2016. Available at https://tools.lifetechnologies.com/content/sfs/manuals/4485610_YfilerPlus_UG.pdf

Zhao X, Ma K, Li H, Cao Y, Liu W, Zhou H, Ping Y (2015) Multiplex Y-STRs analysis using the ion torrent personal genome machine (PGM). Forensic Sci Int Genet 19:192–196. https://doi.org/10.1016/j.fsigen.2015.06.012

Haplodiploid Markers and Their Forensic Relevance

9

Antonio Amorim and Nadia Pinto

Contents

Abstract

The interpretation of a DNA profile requires the knowledge of the mode of transmission of the involved genetic markers. Living beings show a remarkable variety of transmission fashions, and restricting our analysis to eukaryotic sexual organisms, we can formalize the main ones into three categories: uniparental (normally haploid, as happens with plastids or mitochondria),

A. Amorim (✉)
Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

Institute for Research and Innovation in Health (i3S), University of Porto, Porto, Portugal

Faculty of Sciences of the University of Porto (FCUP), Porto, Portugal
e-mail: aamorim@ipatimup.pt

N. Pinto
Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal

Institute for Research and Innovation in Health (i3S), University of Porto, Porto, Portugal

Centre of Mathematics of the University of Porto, Porto, Portugal
e-mail: npinto@ipatimup.pt

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_7

biparental homogametic (usually diploid), and heterogametic or hybrid (uni- or biparental according to parental sex, haplodiploid). The haplodiploid mode of transmission, albeit widespread, has been overlooked, and the theoretical developments required for its analysis lagged. Worse still, haplodiploid markers are many times misused and analyzed under classical diploidy assumption. Here, we provide a synthesis of the theoretical framework required for the analysis of this type of markers, enabling a fast guide for their use in DNA profiling at individual, familial, and population levels, with a special emphasis on forensic applications.

Keywords

Haplodiploidy · Haplodiploid markers · Genetic transmission · X chromosome

Introduction

The production of a DNA profile from a biological sample simply requires the knowledge of the technique(s) used and the appropriate protocols. The interpretation of the results is however more demanding as it implies the theory behind the assumed mode of transmission of the involved genetic markers.

Realizing that current life on the planet relies upon nucleic acids (and sensu stricto on DNA) for storage of genetic information, the variety of transmission modes observed is remarkable. Even restricting our analysis to eukaryotic sexual organisms (Ashman et al. 2014), we can formalize the main ones (and widespread in animals) into three categories: uniparental (normally haploid, as for plastids or mitochondria), biparental homogametic (nuclear, usually diploid), and hybrid or heterogametic (uni- or biparental according to parental sex, from now on referred to as haplodiploid). It is noteworthy also that, in all eukaryotic free-living organisms, at least two of these modes of transmission coexist (Yahalomi et al. 2020), corresponding to distinct genomic sections.

The haplodiploid mode of transmission occurs in many forms and depths. In most mammals, for instance, with very few exceptions (Matveevsky et al. 2017), sex is determined by a karyotypic determination system, in which males XY are haploid for one chromosome, while females are diploid, XX. More radically, haplodiploidy determines the sex in all members of some entire insect orders (Hymenoptera, Thysanoptera): males develop from unfertilized eggs and are haploid, while females are diploid (see the consequences for genome assembly in [Yahav and Privman 2019]). A serious consequence of this gender genetic asymmetry is that the haplodiploid sex is affected by deleterious alleles at the chromosome(s) involved, even if they are recessive in the diploid sex, as the case of hemophilia in humans (Franchini and Mannucci 2012).

Albeit widespread, haplodiploidy has been overlooked, and the theoretical developments required for its analyses lagged. Worse still, haplodiploid markers are many times misused and analyzed under classical diploidy framework. This is unfortunately the case in forensic applications (Ferragut et al. 2019), where, instead of being a problem, haplodiploidy can be a solution to various common casework situations (Pinto et al. 2011).

With this motivation we provide here a synthesis of the theoretical framework required for the analysis of this type of markers, enabling a fast guide to their use in DNA profiling at individual, familial, and population levels, with a special emphasis on forensic applications.

For clarity sake, we will begin a single, per marker, approach and then proceed to the simultaneous analyses, enabling to a global comprehension of the complexities brought by asymmetrical recombination and gametic association (linkage disequilibrium, LD). The case of mammalian XX/XY type of haplodiploidy will be specially used, not only for anthropocentric reasons and applications, but also because of the existence in both X and Y chromosomes of genomic sections (two in humans, one in the rest of known mammals), with a particular transmission behavior, which is neither strictly biparental nor haplodiploid, the so-called pseudoautosomal regions (PARs) (Otto et al. 2011).

As introduced above, the haplodiploid mode of transmission is associated with sexual reproduction and, therefore, with meiosis. So we will briefly and schematically present an overview of the process.

Sexually reproducing eukaryotic organisms do not transmit their genetic blueprint in bulk to offspring. Instead, through a special cellular division (meiosis), they halve (most of) their (diploid) genome, producing transitory gametes, which, if successfully fused into an egg, create a new being, with reconstructed double amount of information. This is not true not only for cytoplasmic genomes but also, as in many animals with genetic determination of sex, for some information carried by nuclear chromosomes. Figure 1 illustrates this process using the mammalian type as example and makes visible the very distinct pattern of recombination in different parts of the nuclear genome: while in female meiosis all chromosomes behave homogeneously, in males crossing-over is absent from the sex chromosomes with the exception of PAR(s). This heterogeneous distribution of recombination has a major impact on the distinctive transmission properties of haplodiploid markers that will be explored below.

Theoretical and Statistical Analyses

Single Locus

At the familial level, the haplodiploid formal model for a single locus is indeed very simple. Using again Fig. 1, we can easily design a predictive matrix in the form of a Mendelian chessboard, in which the extension to population proportions assuming Hardy-Weinberg equilibrium (HWE) is shown, associating the frequency of the observed alleles in the gene pool (p is the frequency of A1, and 1-p the frequency of A2, which can be either an allele or the result of pooling all non-A1 alleles) – see Fig. 2.

Note that, besides HWE assumptions for homogametic systems (infinite population size, random mating, and absence of selection, migration, and mutation), we must also *assume equal allele frequencies in males and females*. There is no need to

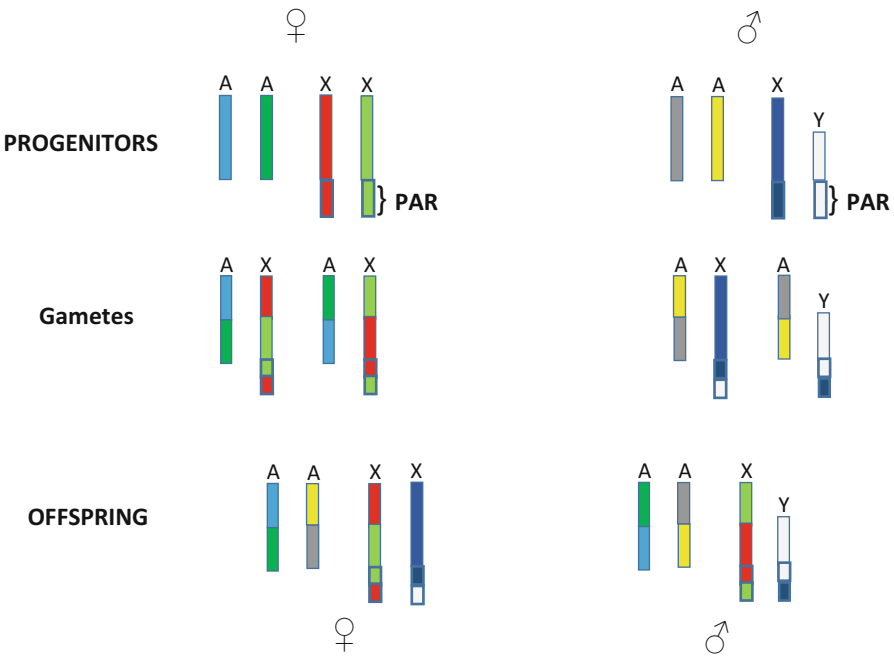


Fig. 1 Chromosomal basis of haplodiploidy. Schematic representation of the auto- and heterosomal structure and mode of transmission and recombination, in the mammalian type of sex determination. Coloring of autosomes (A) and heterosomes (X and Y) identifies the parental origin of the chromosomes and allows to trace the occurrence of recombination. Note that recombination can occur between X and Y but only at the pseudoautosomal region (PAR) and that haplodiploid markers do not recombine in male meiosis

		X	X	Y
		A1 p	**A2** (1-p)	**–**
X	**A1** p	♀ **A1A1** p^2	♀ **A1A2** p(1-p)	♂ **A1–** p
X	**A2** (1-p)	♀ **A1A2** p(1-p)	♀ **A2A2** $(1-p)^2$	♂ **A2–** (1-p)

Fig. 2 Predictive matrix of genotypic distribution, considering haplodiploid transmission and Hardy-Weinberg equilibrium. Allele frequencies: A1 – p; A2 – (1-p)

assume equal absolute numbers of males and females nor equal proportions of X- and Y-bearing gametes in male gametogenesis, and real populations do not observe these proportions.

Conformity of observed genotype distribution to HWE expectations is not as simple as in homogametic systems. In fact, the females' sample can indeed be tested in the usual way (*caveat: using allele frequencies estimated in this sex!*), but we have an extra verification to make: are the allele frequencies in the two sexes the same? To test this, various statistical solutions are possible, but the simplest is to treat male and female gene pools as two populations and compare the two observed allele distributions (a contingency test 2 × n alleles table). The computation of a Kolmogorov-Smirnov test could be a standard approach for this purpose.

At this point we can already extract some conclusions relevant to practical DNA profiling and forensic application of a haplodiploid marker:

1. *No individual male can show more than one allele (no recessiveness)*, and so, in males, genotype and allele frequencies are the same; observed discrepancies to this rule must be investigated (they can be due to previously undetected duplications, which would lead also to triple allelic patterns in females) or to primer unspecificity (Gusmão et al. 2009; Diegoli et al. 2014).
2. *Any significant deviation from HWE compromises the use of the marker* at least on the population under scrutiny, as standard statistical approaches assume such condition for the calculation of expected frequencies. Possible causes deserve a dedicated research as their origins are multiple (and some can be inferred in silico): technical problems, silent alleles (heterozygote excess), selection (genotypes' deficiency – particularly detectable in males if an allele is severely deleterious), migration (sex biased or not), nonrandom mating, drift (caused by small effective male and/or female population sizes), or even mutation. This remark holds against the careless and misplaced use of pseudo-conservative approaches, artificially lowering the p-values of significance thresholds (Ye et al. 2020).

Haplotype

Due to space limitations, we will analyze the simplest case (a pair of biallelic markers) with greater detail, the generalization being straightforward. Haplotype is here therefore defined as the pairwise combination of two alleles from different loci. In Table 1 we show, in a similar way to the one presented for a single locus in Fig. 2, haplotype distributions and intergenerational evolution. Needless to say, *this analysis can only proceed if both loci involved have shown to be in HWE*.

The simultaneous consideration of two makers introduces a new degree of complexity: the association between them. This association has also two components: lack of randomness at gamete formation (linkage) and at the population level. In any case, the haplotype frequency results being different from the product of the frequencies of the non-alleles involved.

Table 1 Simultaneous analysis of two loci: haplotype distributions and intergenerational evolution. Each locus is assumed in HWE. Allele frequencies: locus A – p, q (=1-p); locus B – u, v (=1-p); haplotype frequencies: A1B1 (h_1; in HWE = p × u), A1B2 (h_2; in HWE = p × v), A2B1 (h_3; in HWE = q × u), and A2B2 (h_4; in HWE = q × v), considered to be equal in males and females; r – recombination rate

	Female	$\frac{A1B1}{A1B1}$ h_1^2	$\frac{A1B1}{A1B2}$ $2h_1h_2$		$\frac{A1B1}{A2B1}$ $2h_1h_3$		$\frac{A1B1}{A2B2}$ $2h_1h_4$				$\frac{A1B2}{A1B2}$ h_2^2	$\frac{A1B2}{A2B1}$ $2h_2h_3$				$\frac{A1B2}{A2B2}$ $2h_2h_4$		$\frac{A2B1}{A2B1}$ h_3^2	$\frac{A2B1}{A2B2}$ $2h_3h_4$		$\frac{A2B2}{A2B2}$ h_4^2
	Gametic output	A1B1	½ A1B1	½ A1B2	½ A1B1	½ A2B1	½ (1−r) A1B1	½ (1−r) A2B2	½ (r) A1B2	½ (r) A2B1	A1B2	½ (1−r) A1B2	½ (1−r) A2B1	½ (r) A1B1	½ (r) A2B2	½ A1B2	½ A2B2	A2B1	½ A2B1	½ A2B2	A2B2
Male gametic output	A1B1 h1	$\frac{A1B1}{A1B1}$ h_1^3	$\frac{A1B1}{A1B1}$ $h_1^2h_2$	$\frac{A1B1}{A1B2}$ $h_1^2h_2$	$\frac{A1B1}{A1B1}$ $h_1^2h_3$	$\frac{A1B1}{A2B1}$ $h_1^2h_3$	$\frac{A1B1}{A1B1}$ (1−r) $h_1^2h_4$	$\frac{A1B1}{A2B2}$ (1−r) $h_1^2h_4$	$\frac{A1B1}{A1B2}$ (r) $h_1^2h_4$	$\frac{A1B1}{A2B1}$ (r) $h_1^2h_4$	$\frac{A1B1}{A1B2}$ $h_1h_2^2$	$\frac{A1B1}{A1B2}$ (1-r) $h_1h_2h_3$	$\frac{A1B1}{A2B1}$ (1-r) $h_1h_2h_3$	$\frac{A1B1}{A1B1}$ (r) $h_1h_2h_3$	$\frac{A1B1}{A2B2}$ (r) $h_1h_2h_3$	$\frac{A1B1}{A1B2}$ $h_1h_2h_4$	$\frac{A1B1}{A2B2}$ $h_1h_2h_4$	$\frac{A1B1}{A2B1}$ $h_1h_3^2$	$\frac{A1B1}{A2B1}$ $h_1h_3h_4$	$\frac{A1B1}{A2B2}$ $h_1h_3h_4$	$\frac{A1B1}{A2B2}$ $h_1h_4^2$
	A1B2 h2	$\frac{A1B1}{A1B2}$ $h_1^2h_2$	$\frac{A1B1}{A1B2}$ $h_1h_2^2$	$\frac{A1B2}{A1B2}$ $h_1h_2^2$	$\frac{A1B1}{A1B2}$ $h_1h_2h_3$	$\frac{A1B2}{A2B1}$ $h_1h_2h_3$	$\frac{A1B1}{A1B2}$ (1−r) $h_1h_2h_4$	$\frac{A1B2}{A2B2}$ (1−r) $h_1h_2h_4$	$\frac{A1B2}{A1B2}$ (r) $h_1h_2h_4$	$\frac{A1B2}{A2B1}$ (r) $h_1h_2h_4$	$\frac{A1B2}{A1B2}$ h_2^3	$\frac{A1B2}{A1B2}$ (1−r) $h_2^2h_3$	$\frac{A2B1}{A1B2}$ (1−r) $h_2^2h_3$	$\frac{A1B1}{A1B2}$ (r) $h_2^2h_3$	$\frac{A1B2}{A2B2}$ (r) $h_2^2h_3$	$\frac{A1B2}{A1B2}$ $h_2^2h_4$	$\frac{A1B2}{A2B2}$ $h_2^2h_4$	$\frac{A1B2}{A2B1}$ $h_2h_3^2$	$\frac{A1B2}{A2B1}$ $h_2h_3h_4$	$\frac{A1B2}{A2B2}$ $h_2h_3h_4$	$\frac{A1B2}{A2B2}$ $h_2h_4^3$
	A2B1 h3	$\frac{A1B1}{A2B1}$ $h_1^2h_3$	$\frac{A1B1}{A2B1}$ h_1h_2 h_3	$\frac{A1B2}{A2B1}$ h_1h_2 h_3	$\frac{A1B1}{A2B1}$ $h_1h_3^2$	$\frac{A2B1}{A2B1}$ $h_1h_3^2$	$\frac{A1B1}{A2B1}$ (1−r) $h_1h_3h_4$	$\frac{A2B1}{A2B2}$ (1−r) $h_1h_3h_4$	$\frac{A1B2}{A2B1}$ (r) $h_1h_3h_4$	$\frac{A2B1}{A2B1}$ (r) $h_1h_3h_4$	$\frac{A1B2}{A2B1}$ $h_2^2h_3$	$\frac{A1B2}{A2B1}$ (1−r) $h_2h_3^2$	$\frac{A2B1}{A2B1}$ (1−r) $h_2h_3^2$	$\frac{A1B1}{A2B1}$ (r) $h_2h_3^2$	$\frac{A2B1}{A2B2}$ (r) $h_2h_3^2$	$\frac{A1B2}{A2B1}$ $h_2h_3h_4$	$\frac{A2B1}{A2B2}$ $h_2h_3h_4$	$\frac{A2B1}{A2B1}$ h_3^3	$\frac{A2B1}{A2B1}$ $h_3^2h_4$	$\frac{A2B1}{A2B2}$ $h_3^2h_4$	$\frac{A2B1}{A2B2}$ $h_3h_4^2$
	A2B2 / h4	$\frac{A1B1}{A2B2}$ $h_1^2h_4$	$\frac{A1B1}{A2B2}$ h_1h_2 h_4	$\frac{A1B2}{A2B2}$ h_1h_2 h_4	$\frac{A1B1}{A2B2}$ $h_1h_3h_4$	$\frac{A2B1}{A2B2}$ $h_1h_3h_4$	$\frac{A1B1}{A2B2}$ (1−r) $h_1h_4^2$	$\frac{A2B2}{A2B2}$ (1−r) $h_1h_4^2$	$\frac{A1B2}{A2B2}$ (r) $h_1h_4^2$	$\frac{A2B1}{A2B2}$ (r) $h_1h_4^2$	$\frac{A1B2}{A2B2}$ $h_2^2h_4$	$\frac{A1B2}{A2B2}$ (1−r) $h_2h_3h_4$	$\frac{A1B2}{A2B2}$ (1−r) $h_2h_3h_4$	$\frac{A1B1}{A2B2}$ (r) $h_2h_3h_4$	$\frac{A2B2}{A2B2}$ (r) $h_2h_3h_4$	$\frac{A1B2}{A2B2}$ $h_2h_4^2$	$\frac{A2B2}{A2B2}$ $h_2h_4^2$	$\frac{A2B1}{A2B2}$ $h_3^2h_4$	$\frac{A2B1}{A2B2}$ $h_3h_4^2$	$\frac{A2B2}{A2B2}$ $h_3h_4^2$	$\frac{A1B2}{A2B2}$ h_4^3
	Y	$\frac{A1B1}{Y}$ h_1^2	$\frac{A1B1}{Y}$ h_1h_2	$\frac{A1B2}{Y}$ h_1h_2	$\frac{A1B1}{Y}$ h_1h_3	$\frac{A2B1}{Y}$ h_1h_3	$\frac{A1B1}{Y}$ (1−r) h_1h_4	$\frac{A2B2}{Y}$ (1−r) h_1h_4	$\frac{A1B2}{Y}$ (r) h_1h_4	$\frac{A2B1}{Y}$ (r) h_1h_4	$\frac{A1B2}{Y}$ h_2^2	$\frac{A1B2}{Y}$ (1−r) h_2h_3	$\frac{A1B1}{Y}$ (1−r) h_1h_3	$\frac{A1B1}{Y}$ (r) h_1h_3	$\frac{A1B1}{Y}$ (r) h_1h_3	$\frac{A1B2}{Y}$ h_2h_4	$\frac{A2B2}{Y}$ h_2h_4	$\frac{A2B1}{Y}$ h_3^2	$\frac{A2B1}{Y}$ h_3h_4	$\frac{A2B2}{Y}$ h_3h_4	$\frac{A2B2}{Y}$ h_4^2

Association at Gametogenesis Level: Linkage

By definition, haplodiploid loci are always linked, as they sit at a chromosome that does not recombine in one sex, but even in females, recombination rate (r) varies ($0 \leq r \leq 0.5$) with the physical distance between loci. Several mapping functions were designed to convert genetic distances into recombination rates, such as in (Haldane 1919) or (Kosambi 1944) functions for humans.

This means that the general prediction for the production of gametes in doubly heterozygous females A1B1/A2B2 is ½(1−r) A1B1 + ½(1−r) A2B2 + ½(r) A1B2 + ½(r) A2B1, instead of the expectation under independence, ¼ for each type. Note also that double heterozygotes can harbor another, genotypically indistinguishable, haplotype phase, A1B2/A2B1, and then gamete distribution will be, symmetrically:

$$\S(r)\ \mathrm{A1B1} + \S(r)\ \mathrm{A2B2} + \S(1-r)\ \mathrm{A1B2} + \S(1-r)\ \mathrm{A2B1}.$$

The consequences for general profiling, forensics, and other applications, such as genetic counseling, are:

1. *Simultaneous analysis of various haplodiploid markers requires haplotypic phasing,* which can be difficult as it requires either familial inference (sometimes unavailable) or special technical approaches (long read sequencing).
2. *Robust estimates of recombination rates between haplodiploid markers are always needed* whenever their analysis exceeds purely descriptive statistics.

Association at the Population Level

Association between markers at the population level is defined as the situation of a population in which haplotype frequencies differ from the expected ones, so that, in the absence of association, the frequency of haplotype A1B1 should be the product of the frequencies of the non-alleles involved, $a1 \times b1$. Unfortunately (and misleadingly, as unlinked markers can be associated), the situation is frequently referred in the literature as *linkage disequilibrium* (or simply LD) despite many attempts of replacement with better wording for the concept (Hedrick 1987).

This simple definition is however often misunderstood, and the verification if LD is present in a specific population faces difficulties, both of theoretical and of statistical nature. One of the common misunderstandings is rooted on the fact that the LD definition addresses the association between a pair of alleles at two loci, implying that (with the exception of strictly biallelic systems) *at these two loci can coexist pairs of alleles which show association and others that do not*. In the case of haplodiploid markers, we must add an extra complexity: *LD can be different in the two sexes*.

On the practical application, first of all, we face a sampling problem, driven by the number of (independent) parameters required. Instead of the estimates required for allele frequencies at each locus in both sexes (in codominance by simple gene

counting), we now need haplotype frequencies. This estimation requires large datasets, as the number of possible haplotypes is the product of allele numbers at each locus (Amorim and Pinto 2018). Moreover, in the homogametic sex, haplotypes are normally not accessible to direct observation (remember the problem of double heterozygote phasing; note also that simple haplotype counting can be performed in males, but it does not inform us on the frequencies at the other sex). Realizing that, it is clear that when performing a HWE test on the female genotype distribution in a population sample, we will be testing simultaneously (*i*) the random association of alleles from different loci to form gametic haplotypes (gametic association or linkage disequilibrium) and (*ii*) the random association of gametes into genotypes (assortative mating). It is possible to disentangle the two effects, separately evaluating the latter through a test, in which expected values are calculated using haplotype instead of allele frequencies. However, those are themselves not directly observed and countable, so they must be estimated in statistically much less efficient methods that anyway cannot inadvertently assume random association.

Unfortunately, a consensual framework for the theoretical foundation of an approach to the association between haplodiploid markers has not yet been reached (Sved and Hill 2018; Thomson and Single 2014; Gorelick and Laubichler 2004; Sabatti and Risch 2002) nor are publicly available software or reviews usable by non-aficionados. A useful guide to forensic applications can be found in (Tillmar et al. 2017), and we can nonetheless formulate the following cautionary conclusions:

1. *All haplodiploid markers in a species with XX/XY-like sex determination system are linked,* as they are located in the same chromosome, which does not recombine in the heterogametic sex; *however, some behave as unlinked like if they were in different chromosomes, provided their physical location allows the recombination rate of independent transmission (0.5)*.
2. *Linked markers can show absence of association in a specific population, and, conversely, unlinked markers can exhibit high LDs,* depending on the evolutionary history of the population.
3. *Association between haplodiploid markers is difficult to prove or disprove,* as large sample sizes are required, not only for test robustness but also for estimation of parameters involved in the calculation of expected values, *which may be different in the two sexes*.

Applications

Individual Profiling

Individual profiling of haplodiploid markers is technically similar to homogametic loci and indeed simpler in the case of heterogametic sex, since genotyping equates to haplotyping. The situation in the homogametic sex is exactly the same as for autosomal markers, with identical problem of phasing. Establishment of haplotypes in a homogametic individual is however of little interest if the profiling purpose is

merely descriptive, intending to characterize an individual. The need for haplotypes is, in contrast, essential if the purpose of profiling is comparative, since two individuals with the same genotype may have distinct haplotype configurations. They may, therefore, produce significantly different gametogenesis (if recombination rate is $< ½$) and expected frequencies of occurrence in the population (if $LD \neq 0$). The use of haplodiploid markers in profile comparisons, namely, in the inference of kinship, will be analyzed in the next section.

Genetic Relatedness

The definition of genetic relatedness between two individuals is grounded in the concept of identity by descent. Two alleles are said to be identical by descent (or IBD) if they have descended from the same ancestral allele. Mutation is considered to break IBD, and therefore, IBD alleles are identical by state.

Generally, two individuals are defined to be related if they share IBD alleles due to familial transmission (Jacquard 1970; Weir et al. 2006). In the case of haploid or haplodiploid transmission, this definition must be revisited as (autosomally) related individuals may not share Y- or X-chromosomal IBD alleles. Indeed, for a specific relationship, different degrees of genetic relatedness may be expected for different modes of genetic transmission (Pinto et al. 2011, 2012). For example, a pair of male mammals sharing the same father but with unrelated mothers, i.e., paternal half-siblings, are related as identical twins from Y-chromosomal point of view and as unrelated from X-chromosomal one – see Table 2. This has practical implications leading, for example, that the mating between a pair of paternal half brother-sister does not carry additional risk for X-chromosomal related conditions than the one associated with the mating of two unrelated individuals – see Fig. 3a and c and Table 2. Notwithstanding, this is not the case for the mating between a pair of maternal half brother-sister, as the probability of individuals sharing a pair of IBD X-chromosomal alleles equates ½, as for autosomes – see Fig. 3b and Table 2 (Pinto et al. 2011, 2012). Indeed, the degree of haploid and haplodiploid genetic relatedness is sex dependent, contrarily to what occurs for autosomes.

Considering one marker and a specific mode of genetic transmission, a set of IBD partitions can be associated with any pair of individuals, depending solely on their relationship. These IBD partitions, also known as Jacquard's coefficients in the case of autosomes (Jacquard 1970), represent all the possibilities of identity by descent between and within individuals' alleles.

Considering autosomal transmission, any relationship between a pair of individuals can be genetically characterized through the probabilities associated with nine IBD partitions. When excluding inbreeding, the number of partitions reduces to three, representing the possibility of the pair of individuals sharing two, one, or none pairs of IBD alleles (Jacquard 1970; Weir et al. 2006). This is also the case for X-chromosomal transmission between two females, reducing the number of IBD partitions to four for a female-male pair (two if excluding inbred females). Finally, when a pair of males is at stake, two trivial IBD patterns are possible: either the pair

Table 2 Probabilities of two individuals sharing n pairs of IBD alleles, depending on the mode of genetic transmission and considering the pedigrees mentioned in the text (NA = not applicable)

Sex of the analyzed individuals	Pedigrees	Modes of genetic transmission									
		Diploid			Haplodiploid			Haploid			
								Y-chr.		mtDNA	
		Probability of sharing n pairs of IBD alleles									
		$n = 2$	$n = 1$	$n = 0$	$n = 2$	$n = 1$	$n = 0$	$n = 1$	$n = 0$	$n = 1$	$n = 0$
Male-male	Father-son	0	1	0	NA	0	1	1	0	0	1
	Full brothers	¼	½	¼		½	½	1	0	1	0
	Paternal half brothers	0	1/2	1/2		0	1	1	0	0	1
	Paternal grandfather-granddaughter										
	Paternal uncle-nephew										
	Unrelated	0	0	1		0	1	0	1	0	1
Male-female	Full brother-sister	¼	½	¼	NA	½	½	NA	NA	1	0
	Paternal half brother-sister	0	1/2	1/2		0	1			0	1
	Maternal half brother-sister					½	½			1	0
	Unrelated	0	0	1		0	1			0	1
Female-female	Mother-daughter	0	1	0	0	1	0	NA	NA	1	0
	Full sisters	¼	½	¼	½	½	0			1	0
	Paternal grandmother-granddaughter	0	½	½	0	1	0			0	1
	Paternal half sisters									0	1
	Unrelated	0	0	1	0	0	1			0	1

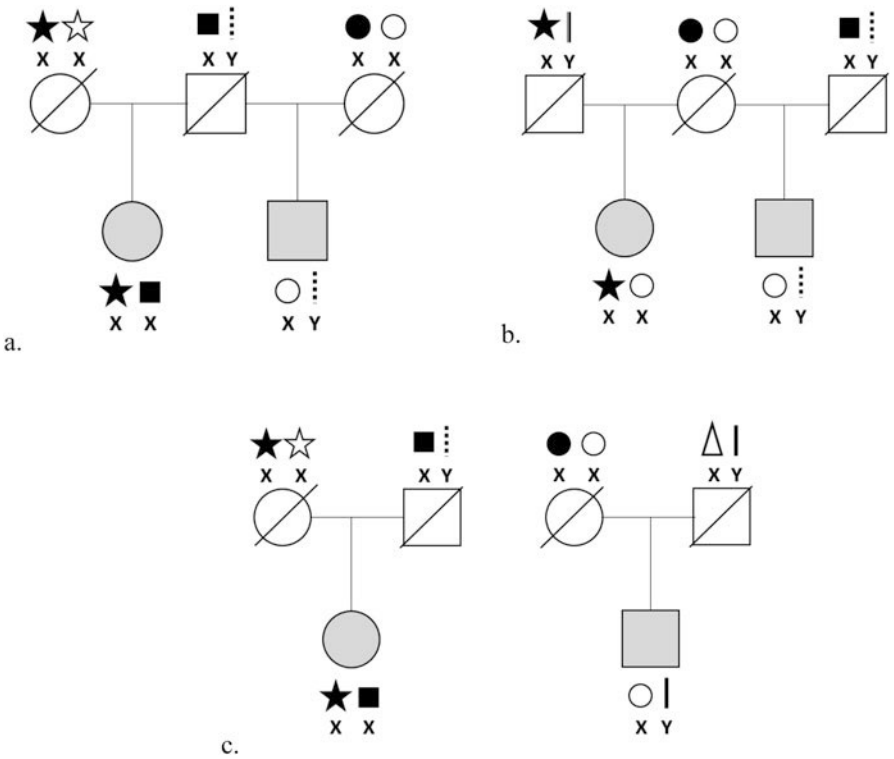


Fig. 3 Examples of genetic transmission, considering one haplodiploid locus and a pair of available individuals related as paternal (**a**) or maternal half brother-sister (**b**) and as unrelated (**c**)

of X-chromosomal alleles is IBD or not. For both diploid and haplodiploid modes of genetic transmission, simple counting rules to determine IBD coefficients for relationships linking two non-inbred individuals were already developed (Pinto et al. 2010b, 2012).

For any mode of genetic transmission, IBD coefficients are crucial for the quantification of the evidence, which is grounded in the Bayes' theorem and in the computation of a likelihood ratio, comparing the probabilities of the observations (genotypes) given the alternative hypotheses (relationships) (Weir et al. 2006; Pinto et al. 2011). These probabilities depend on both the IBD probabilities of the relationships at stake and on the population frequency of the alleles.

At this point it is noteworthy to highlight that (*i*) different kinships may have associated the same set of IBD probabilities and that (*ii*) a specific kinship may have associated different IBD probabilities when considering different modes of genetic transmission.

Proposition (*i*) implies that different kinships may be theoretically indistinguishable when analyzing independent markers. The most commonly cited example is the case of avuncular, half-siblings, and grandparent-grandchild when a pair of

individuals is analyzed for autosomal transmission. Assuming haplodiploid independent transmission, the same occurs for diverse sets of pedigrees, as is the case of father-son and unrelated males, or for paternal half sisters and mother-daughter (Pinto et al. 2011). In Table 2 we present three pedigrees, relating two males: paternal half brothers, paternal grandfather-grandson, and paternal uncle-nephew, theoretically indistinguishable by the analysis of independent markers considering diploid, haploid, and haplodiploid genetic transmission.

On the other hand, proposition (*ii*) highlights that different proportions of IBD alleles may be expected for the same pedigree when different modes of transmission are considered. For example, a pair of full siblings does not necessarily share IBD autosomal alleles, contrarily to what occurs when considering X-chromosomal ones – see Table 2. This leads that some modes of genetic transmission are expected to provide stronger results than others, depending on the hypotheses under analysis. The most commonly cited example in forensics to highlight that haplodiploid markers can be a major added value comparing to autosomes is related with complex kinship problems, where the alleged father of a daughter is not directly available for testing, contrarily to his undoubted mother or daughter. In these cases, the main hypothesis considers females related as full or paternal half sisters, or as paternal grandmother-granddaughter, and in both cases the sharing of IBD X-chromosomal alleles is mandatory, unless mutation or a silent allele occur – see Table 2. This leads that Mendelian incompatibilities can be identified, which represents a major improvement in the power of statistical analyses (Pinto et al. 2013).

Conclusions

Before concluding, let us clarify that since the molecular biology and technical aspects of DNA profiling were addressed in various chapters of this book, here we have approached haplodiploidy in the formal, statistical, and theoretical perspective. Readers particularly interested in the forensic typing of X-chromosome markers can find useful information in other chapters as well as in (Diegoli 2015; Gomes et al. 2020). Only when technical issues were important to the correct modeling of the approach – as in the case of haplotype phasing – were these questions lightly discussed.

Also, despite the enormous importance of haplodiploidy in various branches of life sciences – from genetics proper to physiology and evolution – and the diversity of applications, namely, of medical nature, we have put emphasis on the use of haplodiploid markers in a forensic context.

In this area, not only for historical reasons but also due to practical, legal, and ethical questions, autosomal markers have been – and will continue to be – the first (and in many times the only ones) to use. However, in many kinship cases, the current standard autosomal profiling (which is satisfactorily powerful for first-degree relationships, as maternity or paternity) fails in producing convincing evidence. In many of these, such as paternal sisters or paternal grandmother-granddaughter, the use of haplodiploid markers can provide a powerful contribution. Moreover,

(unlinked) autosomal markers are unable to distinguish between alternative genealogies (Pinto et al. 2010a), when they belong to the same kinship class (as between avuncular and half-siblings). Here again the use of haplodiploid markers enables the distinction between alternative pedigrees (as between paternal and maternal aunt-niece, or maternal half brothers and maternal uncle-nephew). Finally, haplodiploid markers can also be used in sexing a sample of unknown or disputed sex, not only in humans but also in any species with haplodiploidy (as in birds, for which sometimes sexual dimorphism is scant [Henderson et al. 2013]).

At this point it may be interesting to recall that all commercial (autosomal) kits for generating STR (short tandem repeat) profiles to be deposited in forensic genetic databases (as CODIS) include a sexing marker, AMEL (amelogenin), a gene that occurs on the sex-specific region of both the X and Y chromosomes (Issan et al. 2019). This marker is prone to many amplification failures and therefore risks sex misidentification (Thangaraj et al. 2002; Alves et al. 2006). We believe some of these problems could have been solved if haplodiploid markers were included in these kits, as recently shown by (Zajac et al. 2020).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments IPATIMUP integrates the i3S research unit, which is partially supported by the Portuguese Foundation for Science and Technology. This work was partially financed by FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the COMPETE 2020 Operational Program for Competitiveness and Internationalization (POCI), Portugal 2020, and by Portuguese funds through FCT (Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação) in the framework of the project "Institute for Research and Innovation in Health Sciences" (POCI-01-0145-FEDER-007274).

References

Alves C, Coelho M, Rocha J, Amorim A (2006) The amelogenin locus displays a high frequency of X homologue failures in São Tomé Island (West Africa). Int Congr Ser 1288:271–273. https://doi.org/10.1016/j.ics.2005.10.036

Amorim A, Pinto N (2018) Big data in forensic genetics. Forensic Sci Int Genet 37:102–105. https://doi.org/10.1016/j.fsigen.2018.08.001

Ashman T et al (2014) Tree of sex: a database of sexual systems. Sci Data 1:140015. https://doi.org/10.1038/sdata.2014.15

Diegoli TM (2015) Forensic typing of short tandem repeat markers on the X and Y chromosomes. Forensic Sci Int Genet 18:140–151. https://doi.org/10.1016/j.fsigen.2015.03.013. Epub 2015 Mar 25. PMID: 25934544

Diegoli TM, Linacre A, Coble MD (2014) Population genetic data for 15 X chromosomal short tandem repeat markers in three U.S. populations. Forensic Sci Int Genet 8(1):64–67. https://doi.org/10.1016/j.fsigen.2013.07.008

Ferragut JF, Pinto N, Amorim A, Picornell A (2019) Improving publication quality and the importance of Post Publication Peer Review: the illustrating example of X chromosome analysis and calculation of forensic parameters. Forensic Sci Int Genet 38:e5–e7. https://doi.org/10.1016/j.fsigen.2018.11.006

Franchini M, Mannucci PM (2012) Past, present and future of hemophilia: a narrative review. Orphanet J Rare Dis 7:24. https://doi.org/10.1186/1750-1172-7-24

Gomes I, Pinto N, Antão-Sousa S, Gomes V, Gusmão L, Amorim A (2020) Twenty years later: a comprehensive review of the X chromosome use in Forensic Genetics. Front Genet 11:926. https://doi.org/10.3389/fgene.2020.00926

Gorelick R, Laubichler MD (2004) Decomposing multilocus linkage disequilibrium. Genetics 166 (3):1581–1583. https://doi.org/10.1534/genetics.166.3.1581

Gusmão L, Sánchez-Diz P, Alves C et al (2009) A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. Int J Legal Med 123:227–234. https://doi.org/10.1007/s00414-008-0309-4

Haldane JBS (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors. J Genet 8:299–309

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117(2): 331–341

Henderson A, Lee CM, Mistry V, Thomas MD, Iyengar A (2013) Reliable and robust molecular sexing of the hen harrier (*Circus cyaneus*) using PCR-RFLP analysis of the CHD 1 gene. J Forensic Sci 58(2):491–494. https://doi.org/10.1111/1556-4029.12030

Issan Y, Avlas O, Daniel O (2019) Investigation of different ways in which the CODIS 7.0 may be used in mass disaster identification. J Forensic Sci 64(3):852–856. https://doi.org/10.1111/1556-4029.13915

Jacquard A (1970) Structures Génétiques des Populations. Masson & Cie, Paris. English translation: genetics of human populations, Springer, New York, 1974

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics 12: 172–175

Matveevsky S, Kolomiets O, Bogdanov A, Hakhverdyan M, Bakloushinskaya I (2017) Chromosomal evolution in Mole Voles Ellobius (Cricetidae, Rodentia): Bizarre sex chromosomes, variable autosomes and meiosis. Genes (Basel) 8(11):306. https://doi.org/10.3390/genes8110306

Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF (2011) About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet 27(9):358–367. https://doi.org/10.1016/j.tig.2011.05.001

Pinto N, Gusmão L, Amorim A (2010a) Likelihood ratios in kinship analysis: contrasting kinship classes, not genealogies. Forensic Sci Int Genet 4(3):218–219. https://doi.org/10.1016/j.fsigen.2009.06.010

Pinto N, Silva PV, Amorim A (2010b) General derivation of the sets of pedigrees with the same kinship coefficients. Hum Hered 70(3):194–204. https://doi.org/10.1159/000316390

Pinto N, Gusmão L, Amorim A (2011) X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial. Forensic Sci Int Genet 5(1):27–32

Pinto N, Silva PV, Amorim A (2012) A general method to assess the utility of the X-chromosomal markers in kinship testing. Forensic Sci Int Genet 6:198–207

Pinto N, Gusmão L, Egeland T, Amorim A (2013) Paternity exclusion power: comparative behaviour of autosomal and X-chromosomal markers in standard and deficient cases with inbreeding. Forensic Sci Int Genet 7:290–295

Sabatti C, Risch N (2002) Homozygosity and linkage disequilibrium. Genetics 160(4):1707–1719

Sved JA, Hill WG (2018) One hundred years of linkage disequilibrium. Genetics 209(3):629–636. https://doi.org/10.1534/genetics.118.300642

Thangaraj K, Reddy AG, Singh L (2002) Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? Int J Legal Med 116(2):121–123

Thomson G, Single RM (2014) Conditional asymmetric linkage disequilibrium (ALD): extending the biallelic r2 measure. Genetics 198(1):321–331. https://doi.org/10.1534/genetics.114.165266

Tillmar AO, Kling D, Butler JM, Parson W, Prinz M, Schneider PM, Egeland T, Gusmão L (2017) DNA Commission of the International Society for Forensic Genetics (ISFG): guidelines on the use of X-STRs in kinship analysis. Forensic Sci Int Genet 29:269–275. https://doi.org/10.1016/j.fsigen.2017.05.005

Weir B, Anderson A, Hepler A (2006) Genetic relatedness analysis: modern data and new challenges. Nat Rev Genet 7:771–780. https://doi.org/10.1038/nrg1960

Yahalomi D, Atkinson SD, Neuhof M et al (2020) A cnidarian parasite of salmon (Myxozoa: Henneguya) lacks a mitochondrial genome. Proc Natl Acad Sci U S A 117(10):5358–5363. https://doi.org/10.1073/pnas.1909907117

Yahav T, Privman E (2019) A comparative analysis of methods for de novo assembly of hymenopteran genomes using either haploid or diploid samples. Sci Rep 9:6480. https://doi.org/10.1038/s41598-019-42795-6

Ye Z, Wang Z, Hou Y (2020) Does Bonferroni correction "rescue" the deviation from Hardy-Weinberg equilibrium? Forensic Sci Int Genet 46:102254. https://doi.org/10.1016/j.fsigen.2020.102254

Zajac BK, Scheiper S, Zehner R et al (2020) Brother-brother or father-son? How a dropout of AmelX may facilitate the elucidation of a familial relationship. Int J Legal Med 134:1305–1310. https://doi.org/10.1007/s00414-019-02084-3

Single-Nucleotide Polymorphism

10

A Forensic Perspective

Anubha Gang and Vivek Kumar Shrivastav

Contents

Abstract

In forensic analysis, the conventional short tandem repeat (STR) markers are routinely used for examination of the biological samples. But, in the challenging casework studies such as mass disaster or natural calamities, in which DNA samples are either highly degraded or are present in very minute quantity, single-nucleotide polymorphism (SNPs) serve as a potential marker of choice over STRs. Various techniques of SNP analysis such as molecular beacons,

A. Gang (✉) · V. K. Shrivastav
Regional Forensic Science Laboratory, Indore, MP, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_8

SNaPshot, DNA microarray, flow cytometry, and mass spectrometry opened the channels of analyzing forensic samples. Though SNPs pose certain limitations like low discrimination power, less number of alleles per loci, still, SNPs play a fundamental role in human identification, kinship analysis of genetically related individuals, complex paternity disputes, identification of suspect ethnicity, establishing biogeographical ancestry, as well as phenotypic information of missing suspect. Various databases are available for collection, integration, and analysis of SNP for their application in forensic science. As per forensic perspective, the method of standardization and validation of SNP markers, in consensus with legislators and the scientific community, need to be established.

Keywords

Single-nucleotide polymorphism (SNPs) · Short tandem repeat (STR) · SNaPshot · Kinship analysis · Molecular beacons · Paternity disputes · Mass disaster

Introduction

In forensic DNA typing, the use of genetic markers for the characterization of biological samples serves as a hallmark in human identification. First approach made for SNP-based typing for forensic analysis was with HLA-DQA1 locus (formerly known as HLADQα) polymorphism. Another polymarker studied was Ampli®TypePM which is an expansion of HLA-DQA1 analysis. It contained five different markers, viz., LDLR, GYPA, HBGG, D7S8, and GC with high power of discrimination. The method posed limitation for analysis of DNA samples with more than one contributor, which was overcome by development of conventional STR typing (Butler 2011). Although the use of short tandem repeat (STR) loci as a predominant genetic marker for human identification is well established, yet certain additional markers such as single-nucleotide polymorphism (SNPs) have been explored for their potential to be used by a genetic analyst (Budowle and Van Daal 2008; Canturk et al. 2014). The chapter provides insight on SNPs giving highlights about their advantages over other markers as per forensic perspective, categories of SNP markers, techniques involved in the analysis of SNP, and their applications in forensic science. The chapter also quotes the limitations of using SNP as a marker and certain information regarding various databases available for SNP.

SNPs and Their Attributes

The human genome consists of about 3.2 billion base pairs with the number of repeated DNA sequences harboring an equal amount of difference in its genome. The variation in these repeat sequences arises largely due to size, number, and the length of core repeat units (Roewer 2013; Canturk et al. 2014). The variation in a

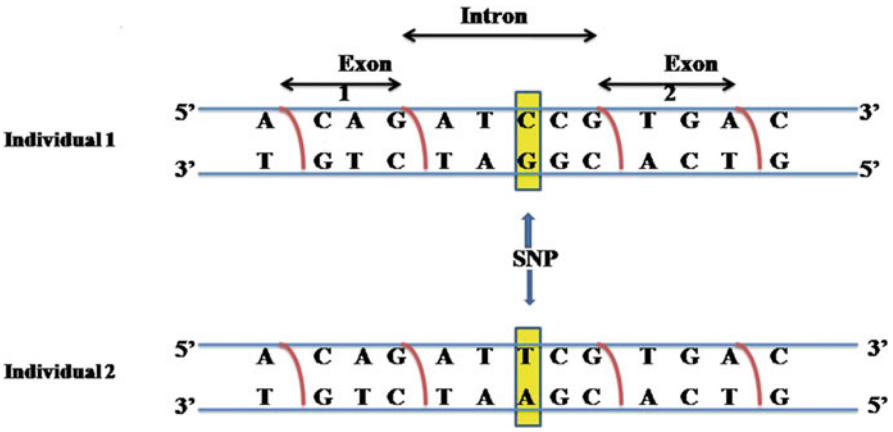


Fig. 1 Single nucleotide polymorphism

sequence called polymorphism is an important characteristic of the human genome. Single-nucleotide polymorphism (SNP) refers to the sequence variants occurring in a genome with a single base pair change (Fig. 1). This change in sequence results due to base substitution, insertion, or deletion at a single site (Li et al. 2015; Brookes 1999). Every individual has millions of SNPs which add to provide a powerful tool in the forensic community for identification. There are nearly 10 million SNPs in the human genome, among these about 1.4 million SNPs have been identified. They comprise mainly the noncoding regions and sometimes coding regions of the genome (Li et al. 2015). Most of the SNPs are biallelic, but tri- and tetra-allelic SNPs have also been reported (Hüebner et al. 2007; Phillips et al. 2015).

Advantages of SNP

SNPs are abundant within the human genome and can be used as markers for forensic applications. Although STR markers are currently been used in genome base analysis for human identification and DNA database preparation, the use of SNP as promising markers needs to be addressed. The most challenging capability of SNP markers is their potential to retrieve information from a highly degraded forensic sample. The devastating natural events or calamities often results in degraded DNA samples or the samples in very minute (i.e., $<$100 bp DNA fragment amplicon size and less than 250 pg DNA sample) quantity (Cho et al. 2014; Canturk et al. 2014). SNPs need a very small target region to be analyzed, i.e., a single nucleotide, and, thus, work well with degraded DNA in contrast to STRs which need a range of 20–60 nucleotides (Butler et al. 2007).

In a study by Cho et al., 169 SNP markers were analyzed using Affymetrix Resequencing Array Technology, and degraded forensic samples were tested. Various markers which could not be shown in STR profiles due to their massive

Table 1 Comparative account of features of STR and SNP

Features	STR	SNP	References
Allele per locus	Usually 5–20 Many alleles	2 (rarely 3) Biallelic	Ballantyne (1997)
Abundance in genome	1 in 15,000 bases	1 in 1000 bases	Kwok and Chen (2003)
Rate of mutation	High	Low	Amorim and Pereira (2005)
Size of amplicon	75–400 bp	<100 bp	Butler (2011)
Discrimination power	High Same achieved by 12–13 STR	Low About 50 SNPs or more required	Gill (2001)
Success with degraded DNA samples	Average	High (DNA samples smaller than 100 bp and quantity less than 250 pg)	Cho et al. (2014)
Phenotypic identification	Cannot be done	Can be done	Butler et al. (2007)
Forensic application	Paternity, identity, sexual assault, and murder cases	High rate of success in cases related to mass disasters, paternity or kinship analysis cases	Budowle and Van Daal (2008)
Analysis methods	Capillary electrophoresis	SNaP shot method, microarray	Khlestkina and Salina (2006)
Stability over generation	Low	High	Butler et al. (2007)
Allele calling (stutters and artifacts)	Present	Not present	Canturk et al. (2014)
Limit of detection of suspects ethnicity	Low	High	Butler (2011)
Evolutionary stability	Low	High	Jehan and Lakhanpaul (2006)

degradation can be detected by Accu ID software using SNP marker. Highly damaged bone samples could even be processed with SNP analysis. On an average, about 130 SNPs were successfully typed from 0.01 ng of DNA (Cho et al. 2014).

Another important feature of SNP is their low mutation rate (10^{-8}) compared to that of STR (10^{-3}). The lower the mutation rate, the more stable is the pattern of inheritance, thereby helps in solving the kinship-related cases involving paternity analysis (Amorim and Pereira 2005; Butler et al. 2007). Bosting et al. used SNPforID 52plex assay in paternity testing and described it to be a useful alternative to the currently used STR typing method and emphasized better results in SNPforID 52 plex system due to low mutation rate in SNPs (Børsting et al. 2008) (Table 1).

SNP amplicons during amplification form small PCR products and reduce the challenge faced by strong PCR inhibitors, leading to higher recovery of DNA from highly degraded material, in which STR loci fail to be amplified by PCR or may result in the generation of partial profile. Thus, smaller PCR products show greater recovery of information from poorly damaged samples. Moreover, interpretation of results using SNP markers is easier due to the absence of stutter products or microvariants (Butler et al. 2007).

SNPs are biallelic, and hence more informative loci are not required for the analysis of forensic samples. In contrast, STRs show multiple allelic states (Butler et al. 2007). It is well established that loci representing low FST values (SNP Frame 1) and moderate heterozygosity (>40%) could be best applied for forensic purposes (Butler et al. 2007; Kidd et al. 2006). SNPs also add in providing phenotypic information such as hair, skin color, and eye color. Various multiplex systems are also available for analysis of SNPs such as microchip, arrays, etc.

In the present scenario, most of the forensic samples collected from the crime scene are either highly degraded or are present in very little amount. The analyses of such exhibits pose a serious problem. The methods need to be employed for analysis that has proper validation and standardization and have also qualified quality assurance tests. The use of SNP typing thus proves to be an adjunct in forensic technology (Canturk et al. 2014; Budowle 2004).

Categories of SNP Markers

Several studies have shown that DNA rather than proteins are stable forensic samples and the degree of variation shown at the DNA level can be explored for forensic studies (Budowle and van Daal 2008). In 2007, the International Society for Forensic Genetics held in Copenhagen, Denmark, emphasized the importance of SNP with a detailed discussion on SNP markers, applications, and technologies that can be explored for forensic casework studies (Butler et al. 2007; Butler 2011).

SNPs can be classified according to the position of occurrence in the genome. These are mainly of four basic types:

- Noncoding SNPs (ncSNPs) which is present in the introns.
- Synonymous type (syn SNP) which is present in exon and change the amino acid composition of the encoded polypeptide.
- Non-synonymous type (ns SNPs) which changes the encoded amino acid.
- Promoter SNPs (pSNPs) are present in the promoter region of the genome (Li et al. 2015).

Different classes of SNPs applicable to forensic analysis can be categorized as:

(i) Identity-testing SNPs
(ii) Lineage-informative SNPs

(iii) Ancestry-informative SNPs
(iv) Phenotype informative SNPs

Identity-Testing SNPs

Identity-testing SNPs may prove useful in excluding individuals which are not related genetically. It shows a similar function as the most common STR loci. It has the sound basis of differentiating people and excludes those who do not link themselves as family members. Some features mark the importance of identity-testing SNPs which include their high level of heterozygosity, i.e., 50% for biallelic and low FST values. This reduces the number of SNPs required for determining high levels of discrimination power, and even statistical assessment in different cases will require low input of datasets of the reference population.

European Forensic DNA community in 2003 developed SNPforID project which developed SNP assay techniques for forensic DNA analysis (Amigo et al. 2008). Several multiplex assays were developed which were clubbed with population data in order to determine the allelic frequency. The website for SNPforID, i.e., http://www.snpforid.org, contains links to population data (Butler 2011). Various studies have shown the development of different panels for SNP identification (Sanchez et al. 2006; Kidd et al. 2006; Pakstis et al. 2010; Yousefi et al. 2018).

Lineage SNPs

The set of highly linked SNPs showing very low mutation rate and no recombination thus are useful for evolutionary studies. It plays a demarcating role to resolve cases related to kinship analysis in which there is a gap of generations in the reference and evidence samples. It becomes an important asset to identify missing individuals from those families having limited members for comparison. Lineage SNPs present on the mitochondrial and the Y chromosome DNA genome have also been identified, but they show limited power of discrimination. Autosomal SNPs, identified as haplotypes, also serve as lineage-based markers. Lineage SNPs are commonly employed for the detection of missing person cases or the identification of mass disaster victims (Tishkoff and Verrelli 2003; Daly et al. 2001).

Ancestry Informative Markers

Those SNPs occurring at different frequencies in a varied population of the world are referred to as ancestral informative markers (AMIs). Due to the high level of allele-sharing among individuals in a population, STR loci serve as poor presenters of ancestral-based information. The SNPs, however, are present throughout the human genome, varying in their frequency of occurrence and focuses on ancestral origin among groups of unrelated individuals rather than describing physical characteristics (Frudakis et al. 2003). The ancestry of an individual may help to resolve certain

cases by indirectly providing limited information about the appearance of a person. The assessment of genetic variation among populations leads to the formation of various databases which in turn helps in quantifying ancestry informative markers. AIMs required low heterozygosity and high FST values (Budowle and Van Daal 2008).

Phenotype Informative SNPs

Phonotypic informative SNP markers represent a high probability of an individual showing particular phenotypic characters, viz., skin, hair, and eye color. The single-nucleotide polymorphism (SNP) was regarded as ideal for studying the phenotypic variation as phenotypic traits provide a strong genetic prediction of the appearance of an individual. The genetic difference among individuals with varying phenotypic characters has been explored in various forensic studies. Barreiro et al. in their study showed that natural selection brings about phenotypic variation in individuals. These variations include skin color, eye color, height, weight, etc. which are partially controlled by genes and are also governed by environmental effect. Their study showed FST value of 0.11 which shows a moderate level of differentiation (Barreiro et al. 2008).

Techniques of SNP Analysis

With the exploration of the human genome sequence, several million SNPs have been discovered, and various technologies had developed which analyzed nearly one million SNPs (Kumar et al. 2012). A large number of methods are accessible for the detection of SNP. Most of the methods include PCR amplification of the SNP region followed by post-PCR analysis by either probe hybridization, primer ligation, or extension of the primer. Several techniques have been developed for the identification of SNP and to create novel SNPs. The techniques involved in genotyping SNPs with high throughput and speed mainly fall into three categories, namely, arrays, mass spectrometry, and flow cytometry. A large number of techniques have been used from time to time for the detection and analysis of SNP (Table 2). Some of the techniques are described in detail.

Molecular Beacons (MB)

Molecular beacons (MB) are short oligonucleotide probes, having a hairpin structure with a known sequence. The probe is labeled with a fluorescence dye usually 6-FAM (6-carboxyfluoescein) or TET (tetrachorofluorescein) at the 5′end called as reported dye (R), as well as a non-fluorescence quencher dye (Q) such as TAMARA (tetramethylrhodamine) at 3′end. When the probe is intact, the quencher greatly reduces the emission of fluorescence by reporter dye due to FRET (fluorescence resonance energy transfer). While, when the probe binds with complementary target DNA, the

Table 2 Techniques for analysis of SNP

Technique	Features	Limitations	References
Direct sequencing	Direct way of SNP identification. Employs locus specific PCR amplification. Sequencing of PCR products and analysis to reveal novel SNPs	Time-consuming, costly, high rate of sequencing error (1 base per 100)	Gupta et al. (2001); Kirk et al. (2002); Jehan and Lakhanpaul (2006)
Direct mining from EST databases	Use partial cDNA clones of an organism to detect SNPs in coding region	High rate of sequence error	Buetow et al. (1999)
Electrophoretic assay	Show electrophoretic migration of DNA samples. Denaturing gradient gel electrophoresis (DGGE). Difference in single base causes three-dimensional conformational change thereby affecting migration rate in electrophoresis, leading to SNP detection	Time-consuming method	Orita et al. (1989)
Zn^{2+}-cyclen binding assay	Modified gel electrophoresis method. Binding of Zn^{2+}-cyclen to thymine base changes overall charge from 0 to +1 (thymine deprotonated). Simple, low-cost, and sensitive method.	Single mutations on DNA fragments with length of more than 300 bp were not detected	Kinoshita-Kikuta et al. (2002)
Peptide nucleic acid (PNA) directed PCR clamping assay	Synthetic analogue of DNA having N-2 amino-ethyl glycine backbone. Mimics DNA and binds to highly thermostable complementary DNA sequence	Requires hybridization of probe in different reaction tubes varying with conditions	Ørum et al. (1993); Giesen et al. (1998)
Temperature-modulated heteroduplex assay (TMHA)	TMHA also called as denaturing high performance liquid chromatography (DHPLC). Distinguishes homo- and heteroduplexes mismatches formed from annealing wild-type and mutant DNA by different chromatographic patterns	Not useful for highly polymorphic DNA. Time-consuming. Requires optimal assay temperature determination	Kota et al. (2001)

(continued)

Table 2 (continued)

Technique	Features	Limitations	References
Fluorescent resonance energy transfer-based method or Taqman assay	Based on real-time PCR. It works on the 5′→3′ exonuclease activity of Taq polymerase. Taqman TM probe is an amalgamation of a fluorescence reporter molecule at the 5′end and a quencher at the 3′end and a probe extension blocking group. Presence of SNP prevents duplex formation and thereby, no degradation and so no fluorescence	Low multiplexing capacity Discrimination solely by hybridization	Kirk et al. (2002)
Molecular beacons	PCR-dependent assay in which single-base mismatch leads to absence of hybridization with complementary sequence	Limited number of probes can be used. Expensive method	Tyagi and Kramer (1996)
Allele-specific hybridization assay	Assay involves use of donor and acceptor beads which produces chemiluminescent signal on hybridization of probe and PCR products for SNP detection. Show multiplexing	Requires expensive instrumentation	Beaudet et al. (2001)
Capillary array electrophoresis (CAE)	Use of 96-channel CAE microplate. The fluorescent-labeled DNA fragments are separated on microplate and target DNA (allele-specific PCR product with labeled primer) is detected by laser exited confocal fluorescent scanner. Sensitive method, high speed	Costly FRET probe required for each SNP	Kheterpal and Mathies (1999)
Electrochemical detection of mismatches in nucleic acids (EDEMNA)	Enzyme immunosorbent assay method coupled with electrochemical sensors in which hybridization with two single-stranded DNA probes conjugated with biotin on streptavidin-	Detection of single mismatch in long probe difficult	Huang et al. (2002)

(continued)

Table 2 (continued)

Technique	Features	Limitations	References
	coated electrochemical sensor surface generates signal. Mismatch of target DNA with either probes results in loss of signal		
Genetic bit analysis (GBA)	Solid phase, primer-based extension method using dd NTPs, biotin, and fluorescein determined by enzyme linked colorimetric method	High-cost method	Nikiforov et al. (1994)
Padlock probes	Oligonucleotide probes having end sequence complementarity to target sequence involving ligation of probes	Forms difficult secondary structures	Nilsson et al. (1994)
Pyrosequencing	Enzyme-coupled sequencing method following synthesis of short 20 nucleotide sequence and primers are designed for sequence flanking SNP sites using charge-coupled device (CCD) in 96 well microtiter base pyrosequencer	Requires PCR step. Costly method, needs single-stranded sequencing template	Ahmadian et al. (2000)
Invader assay	Sequence specific oligonucleotides probes synthesized involving exonuclease activity of flap endonuclease cleaving structure specific sites	Uniplex method, i.e., only one genotype can be performed per assay	Lyamichev et al. (1999)
SNaPshot/ minisequencing	A fluorescence-based primer extension method. Based on dideoxy single-base extension of unlabeled oligonucleotide primer binding to a complementary template	Not a quantitative methodology	Budowle (2004)
Mass spectrometry	Based on property of mass to charge ratio involving ionization,	Loss of signal intensity with increase in size of DNA	Griffin and Smith (2000)

(continued)

Table 2 (continued)

Technique	Features	Limitations	References
	size-based separation and detection		
Flow cytometry	Single biotinylated oligonucleotide annealed adjacent to SNP site, following extension by DNA polymerase and fluorescence ddNTPs. Primers are captured on streptavidin-coated microsphere, and fluorescence is measured	Primer heterodimer formation and false priming	Cai et al. (2000)

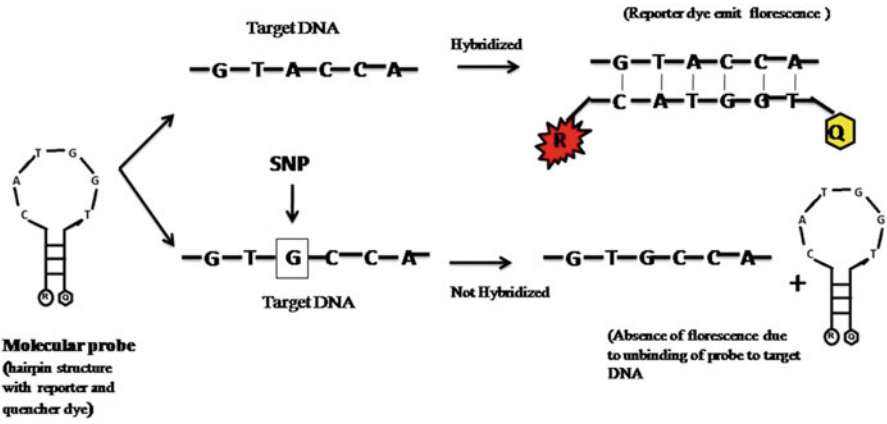


Fig. 2 SNP genotyping through Molecular Beacon assay. (Adapted from Tyagi et al. 1998)

reporter dye is no longer close to the quencher, the FRET is disturbed, and the reporter dye shows emission (Fig. 2). The presence of SNP mutation in the target DNA does not lead to hybridization with the probe; hence no fluorescence was detected. The length of the loop and stem sequence should be appropriate so that it dissociates at a temperature 7–10 °C higher than PCR annealing temperature. It is a gel-free assay, and immobilization could be carried out on a solid surface (Tyagi and Kramer 1996).

Microarrays/DNA Chips

DNA microarray technology has set unrivaled levels of multiplexing, throughput, and parallel analysis. This method is being used for SNP analysis. Microarrays constitute oligonucleotides which are immobilized on a solid surface. This technique

proved useful for SNP analysis subject to automation. This technique is based on sequencing by hybridization (SBH). Tiling strategy is applied for SNP genotyping. The labeled target DNA segments are poured on to immobilized oligonucleotide probe containing chips. Hybridization occurs by complementary base pairing. Each column on microarray has one set of four probes which differ only by a single base. Several columns are being used which differ from one another by substitution position. Hybridization with target wild-type DNA perfectly matches with one of the four probes, while there is a mismatch with the other three probes. SNP in the target DNA hybridized with one of the probes resulting in a perfect match, but in the adjacent set, hybridization results in a large number of mismatches as compared to wild type (Chee et al. 1996).

The semiconductor silicon microchip-based technique of SNP analysis was described by Gilles et al. using the electronic dot blot method. The method is rapid, i.e., each testing site can be controlled individually, showing the ability of multiplexing and liable to automation. The technique also pose some limitations showing a high rate of false postive results and is a high cost assay method (Gilles et al. 1999).

SNaPshot Multiplexing Method

SNaPshot Multiplex system also called minisequencing is based on the ability to utilize unlabeled primers. It uses a single-reaction system in which unlabeled oligonucleotide primer binds to the complementary template sequence in presence of DNA polymerase and fluorescent-labeled dideoxynucleotides (ddNTPs). The amplification of the region around each SNP is carried out using PCR. The primer is extended by one nucleotide in presence of polymerase, but the addition of a single ddNTP terminates the sequence at 3′end, giving fluorescence of respective base added (Fondevila et al. 2017). Post-PCR reaction remaining dNTPs and primers are destroyed by the addition of enzyme exonuclease (Exo), which digests the primer, and enzyme shrimp alkaline phosphatase (SAP) destroys dNTPs. Both these enzymes are together referred to as ExoSAP. This method can analyze up to 10 SNPs in a single reaction, wherever they are placed on the chromosome. It uses automated analysis using GeneMapper Software data analysis system. This is a robust, cost-effective, and reproducible method of SNP analysis. SNaPshot Multiplexing Method is used for developing multiplexed SNP assays for degraded DNA templates, determination of quantitative allele frequency helpful in gene expression, to assess mitochondrial DNA investigation, and for validation of next-generation sequencing (NGS) data (Budowle 2004; Butler 2011).

Mass Spectrometry

It is one of the most prominent techniques for SNP genotyping. It employs the detection of ions and their measurement based on mass to charge ratio (m/z). The separation of ions under vacuum makes this method extremely rapid. It is a more

accurate technique as it involves the measurement of the actual mass of DNA. Mass spectrometry method based on gas phase analysis has different ionization techniques, i.e., matrix-assisted laser desorption-ionization (MALDI) or electrospray ionization (ESI). Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry is more accurate than array methods as it takes milliseconds to complete ionization, size-based separation, and detection. With this technique, known as well as unknown SNPs can be detected. It has a size limitation of <100 nucleotides per sequencing reaction. The loss of signal intensity with an increase in the size of DNA limits this technique to some extent.

An automated electrospray ionization mass spectrometry method is a quantitative method that can detect clustered SNPs within an amplicon and does not require prior knowledge regarding its position. It provides accurate measurement of mass, i.e., 1–25 ppm, thereby helps in obtaining the base composition of PCR amplicon. The presence of SNP is represented by the difference in base composition between the two samples. The size limitation extends to about 250 bp in this technique. SNP typing capability enables mass spectrometry to play a functional role in forensic DNA analysis (Butler 2011; Budowle 2004; Jehan and Lakhanpaul 2006).

Kiesler and Vallone typed 40 autosomal SNP loci in a set of 194 US African American, Caucasian, and Hispanic populations for human identification using electrospray ionization mass spectrometry. The random match probability ranged from 10^{-16} to 10^{-21} for the 40 SNPs (Kiesler and Vallone 2013). NGS platform provides a large volume of data through computer and complex algorithms leading to the extension of bioinformatics.

Forensic Application of SNP Profiling

SNPs are an untapped DNA resource in the forensic arena. Currently, SNPs are used in forensic laboratories, primarily for human identification (HID) from skeletal remains, viz., bone, hair, and teeth, to resolve the disputed cases of paternity, kinship analysis of close relatives, identification of human ethnicity or racism, as well as phenotypic information about missing suspects.

Human Identification (HID) from Skeletal Remains

Forensic samples usually exist in very small amounts as well as in degraded conditions at most of the crime scene (SNP Frame 2). SNP analysis plays a paramount role over conventional STR profiling in such cases. SNP as a potent marker for obtaining reliable data interpretation from heavily damaged DNA samples was described in a study by Cho et al. They showed that out of 169 SNP (selected from 1000 genome data base), more than 120 were typed with just 10 pg of DNA from an artificially degraded sample (femur bone) and samples of hair, buccal swab, and cigarette butt (Cho et al. 2014). Improved DNA typing method for analysis of DNA extraction from degraded skeletal material

(charred femur bone) using mini STRs and SNPs to confirm identity was shown in a study by Fondevila et al. (2008).

Several SNP systems have been developed for their role in forensic science such as SNPforID 52plex, SNPforID consortium which comprises 52 bi-allelic SNP for human identification. Sanchez et al. developed a multiplex human identification genotyping assay with 52 unlinked autosomal single-nucleotide polymorphisms in the European, Asian, and African populations with 700 individuals. DNA typing was carried out, and allele frequencies were determined. Although partial STR profiles were recovered from the degraded samples, 52 SNP loci were efficiently amplified with a mean match probability of 5×10^{-19} in the population studied, showing that SNPs serve as a better marker when studying and analyzing degraded DNA samples (Sanchez et al. 2006).

Kidd et al. studied population genetics describing match probability by developing SNP panel in forensic identification of individuals from 40 populations showing genetic variation in the world's major geographical regions with a match probability of 10^{-15} (Kidd et al. 2006). Tomas et al. in their study described 49 plex autosomal SNP GenPlexTm assay system developed by Applied Biosystem individualization and relation testing. They included 49 out of 52 SNPforID SNPs (Tomas et al. 2013). With the growing array of SNP markers and multiplex systems, SNPforID consortium developed 52 SNP PCR multiplex for human identification using 23Plex and 29Plex single-base extension reaction of autosomal SNPs using AB 310 Genetic Analyzer (Bulbul et al. 2009).

Paternity Testing and Kinship Analysis

SNPs provide information on paternity disputed cases and kinship analysis of genetically related individuals in which STRs markers show limitation. In one of the study, 49 SNPs from 52 plex test were assessed for paternity determination in six cases, where STR analysis was found to be insufficient. SNP could provide a clear representation of the alleged father in one of the father-child duo case, confirming the paternity of the alleged father. In one of the trio case, the paternity of the alleged father was confirmed, while in another trio, study of paternity establishment, the grandfather was identified as the real biological father, and the alleged father was excluded by the SNP analysis. These and various other studies depict the importance of SNPs in addition to STR markers for clarifying problematic paternity and kinship cases (Borsting and Morling 2011).

STR analysis does not always prove useful for solving forensic cases. Sometimes, it may lead to ambiguous results, such as involvement of SE33 locus, which is more prone to mutagenesis and produces erroneous results, and, thus, difficult paternity cases cannot be resolved (Canturk et al. 2014). However, several studies revealed that SNP markers can solve the disputed cases which were not solved by conventional STR analysis (Canturk et al. 2014; Butler et al. 2007).

In one of the studies, Jin et al. explored the potential of micro-haplotypes in forensic research. They identified 22 micro-haplotypes and 7 compound markers on

29 novel genetic markers on SNP and insertion/deletion (InDel) polymorphic loci in Kazak and Mongolian groups in China for paternity testing and human identification. Their results showed a high level of polymorphism in all groups, and the selected 29 loci are very informative markers in Kazak and Mongolian groups in China (Jin et al. 2020).

In another study, doValle-Silva emphasized the applicability of SNPforID 52plex system in the Brazilian population for human identification, and evaluation of ancestry lineage was performed using next-generation sequencing (Haloplex and MiSeq platform). 51 out of 52 SNPforID markers were analyzed in the study, and allele frequencies were monitored. The study of admixture showed a high contribution of European (70%) whereas very low (<10%) contribution of Amerindian in the population. Thus, the study using 52plex kit gives a comparable result obtained using 16 STR panel (do Valle-Silva et al. 2019).

Phenotypic Information of a Missing Suspect

From the evidence present at crime scenes, the DNA profile obtained (from biological samples) is usually matched with that of the suspect. DNA database provides additional information in cases where no match with the suspect is found. Even then, if no match is found, the case remains unsolved. For acquiring more relevant information, genome-wide association studies (GWAS) use SNP genetic variation associated with phenotypic characteristics.

SNP act as a key determinant of variation in human eye color, hair color, and skin pigmentation (Kayser et al. 2008; Liu et al. 2009; Branicki et al. 2011). Association of SNP with human eye color variation in 6168 Dutch Europeans was studied by Liu et al., demonstrating genetic prediction of human phenotypes. They analyzed 37 SNPs from 8 genes including OCA2, HERC2, SLC24A4, SLC45A2, IRF4, and TYR genetic variants. Study of the population sample revealed blue eyes in 67.6% and brown eyes in 22.8%, and 9.6% showed intermediate eye color. The study was statistically significant. These DNA markers may serve as reliable biological evidence in forensic cases of unknown suspects of European descent with eye color prediction test (Liu et al. 2009). Discrimination of blue (SLC24A4 and SLC45A2) and brown (OCA2 and IRF4) eye colors was determined by homozygous alleles of HERC2 gene using six SNPs in eye color-associated genes.

Human skin color along with eye and hair color serve as relevant evidence in forensic casework studies. Walsh et al. (2013) investigated skin color from 77 SNPs from 37 genetic loci associated with human pigmentation. The study was carried out in 2025 individuals from 31 global populations. Numerous highly associated SNPs for human hair color were identified (Branicki et al. 2011; Sulem et al. 2007). In one of the studies by Branicki et al., 45 SNPs in 12 hair color-associated genes were studied. They predicted red, brown, blond, or black hair color in 13 markers present in 11 genes with 90% precision (Branicki et al. 2011). Han et al. (2008) conducted genome-wide association studies of natural hair color and skin

pigmentation in over 10,000 individuals of European Ancestry from the USA and Australia and found IRF4 and SLC24A4 loci highly associated with hair color.

Constraints of SNP

Considering SNP markers as a replacement for STR loci, it needs to be inscribed that SNP poses certain disadvantages which restrict its choice as an ideal marker for forensic studies. SNP loci are less polymorphic as compared to STR loci; thus a large number of SNPs (about 50–60 SNP loci) are required to reach a similar level of random match probability and equivalent power of discrimination as reached with 13–15 STR loci (Butler et al. 2007). In a single multiplex amplification reaction system using 500 pg of DNA template, 15 STRs can be routinely amplified, while the examination of a large number of genetic loci will be highly expensive and complex and will even pose difficulty in data interpretation. The higher the loci have been analyzed, the higher will be the peak signals, and the more will be the artifacts (Li et al. 2015). The use of SNP loci in the analysis of routine paternity cases may also result in a high frequency of inconclusive cases (Amorim and Pereira 2005). When a large number of loci are involved, the quality and integrity of DNA template become challenging, particularly when the DNA material is in a very low amount. Sometimes, recovery of DNA from a limited amount of sample does not allow repetition in the testing of samples to recover the lost loci (Butler et al. 2007). Till SNP markers gain a concord in the forensic community to be employed for DNA typing and database preparation, SNP typing cannot gain a foothold over the ascendancy of widespread STR typing. The limited number of alleles per locus in SNP makes interpretation of results in deciphering mixtures even difficult.

SNP Information Databases

Researchers successfully characterized around 18 million single-nucleotide polymorphism (SNP) loci in the human genome. Some of these SNPs are found either on noncoding or in coding regions of the genome that do not alter the protein sequence and thus have no functional consequences (Jehan and Lakhanpaul 2006; Phillips 2009). Therefore, extraction of useful SNPs from the databases is very important to find out genetic variability, correlation with the disease, as well as phenotypic characterization of an individual. Some examples of major online available SNP databases with their URL are as follows.

dbSNP (http://www.ncbi.nlm.nih.gov/sites/entrez.dbsnp)

In August 1999, National Human Genome Research Institute (NHGRI) and the National Centre for Biotechnology Information (NCBI) jointly constitute a database, i.e., dbSNP. The database includes common and rare single-base nucleotide variation (SNV), short (=<50 bp) deletion/insertion polymorphisms, and other classes of small genetic variations. It is available in a public domain

containing comprehensive data of human SNP. dbSNP can be accessed by Entrez, NCBI, or direct SNP summary page. SNP contains a unique identifier, i.e., rs number. Various genotyping centers continuously collect and merge submitted SNP sequence, which are used as reference SNP (refSNP); the submitted SNP and clustered or reference can be represented as ss and rs SNP, respectively (Smigielski et al. 2000).

HapMap (http://www.hapmap.org/cgi-perl/gbrows/)

HapMap project is an international consortium seeded with human genetic variation. It represents the best graphical overview of a chromosome. It is a database, which aims to determine the haplotype and to study intragenic variation present within the human genome. Thus it can be used as a graphical genome browser for the study of SNP variability (Gibbs et al. 2003).

Ensemble (http://www.ensembl.org/Homo_sapiens/info/index)

This database furnishes the alternative to NCBI, for accessing similar types of SNPs from the genomic data. This database specializes to analyze the genome features and sequences for the identification and annotation of the gene in a vast range of species studied. It also closely integrates with Swiss-Prot/UniProt and provides high-quality protein sequence data for various phenotypic analyses (Phillips 2009).

Santa Cruz (http://genome.ucsc.edu/cgi-bin/hgGateway)

Santa Cruz contains several features that provide information of SNP analysis in easier way than other databases, for the collection and analysis of SNP.

Nowadays, owing to the research and commercial use of SNPs, many private organizations are showing interest in SNP sequence collection, annotation, and analysis like the Celera SNP database.

Conclusion

Single-nucleotide polymorphisms (SNP) are potential markers for human identity testing from degraded DNA samples, identification of the missing person, kinship analysis, and phenotypic characterization. SNPs have the ability to recover the information from very minute quantity of biological material by combining with multiplex assays. However, it is improbable to replace the STR markers with SNPs due to extensive studies being done on STR markers and database prepared on its basis; still, the role of SNPs for human identification cannot be ignored. Moreover, additional studies should be made for SNP genotyping describing the mechanism of allelic discrimination, genotyping methods, and detection modalities. SNPs are likely to become a major tool for future forensic DNA testing by cataloging SNP loci of required interest and combining the features of SNP marker in a common format. Although, many SNPs which are present in the noncoding region do not have any functional consequences, still, it should be of prime focus to extract the useful SNPs from the database and correlate them with functional aspects, viz., diagnosis of disease, phenotypic traits, and the establishment of ancestral and evolutionary lineage.

SNP Frame 1

- **F_{ST}**

  It refers to the fixation index, i.e., measure of degree of population differentiation analyzed from genomic data. Fixation index can be calculated as:

$$F_{ST} = H_T - H_S/H_T,$$

where H_T represents heterozygosity of the total population and H_S represents heterozygosity of the sub-population.

Thus, F_{ST} describes about the genetic variation according among the sub-populations. Theoretically, value of F_{ST} ranges from 0.0 to 1.0, i.e., from no differentiation (i.e., 100% similarity) to complete differentiation.

High F_{ST} values reflect low level of shared alleles between individuals in a population reflecting different ancestral background, whereas low F_{ST} values show allele similarity among the population. However, this allelic variation may change over time, even though F_{ST} value can relate an individual demonstrating its ethnicity and can relate the person with the crime scene (Norrgard and Schultz 2008).

- **Random match probability (RMP)**

  It represents measure of probability of an unrelated person and matching its genotype derived from the evidence in population genetics. The genotype corresponds to a number of alleles, and each allele represents certain frequency in a population.

SNP Frame 2 9/11 World Trade Center Attack

Forensic DNA samples are typically highly degraded and/or present in minute quantities in cases of mass causality (plane crashes) or natural calamity. World Trade Center attack (September 11, 2001) was one of the largest forensic cases to mention (Biesecker et al. 2005). Before this attack, DNA profiling of samples, less than 500 individuals, were taken up for analysis, but this attack brought a breakthrough and identification of victims (>1000) scaled up. National Institute of Justice (NIJ) constituted a group of scientific and medical experts (Kinship and Data analysis Panel) KADAP to advise them on DNA identification. The challenging human remains collected from tiny, charred bone fragments, and pile of vestiges (over million tons) were the source samples for DNA identification. The primary DNA identification took more than 3 years to complete. The efforts led DNA identification project to generate more than 52,000 STR, 44,000 mtDNA, and 17,000 SNP profile from 19,913 putative victim tissue fragments. About ten individuals were

(continued)

identified using SNP analysis alone, and ten more identifications were possible by supplementing partial STR profiling data with SNP genotypes (Biesecker et al. 2005; Butler 2011). Thus, SNPs provide a supportive framework for forensic DNA analysis.

References

Ahmadian A, Gharizadeh B, Gustafsson AC, Sterky F, Nyrén P, Uhlén M, Lundeberg J (2000) Single-nucleotide polymorphism analysis by pyrosequencing. Anal Biochem 280(1):103–110

Amigo J, Phillips C, Lareu M, Carracedo Á (2008) The SNP for ID browser: an online tool for query and display of frequency data from the SNP for ID project. Int J Legal Med 122(5):435–440

Amorim A, Pereira L (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Sci Int 150(1):17–21

Ballantyne J (1997) Mass disaster genetics. Nat Genet 15(4):329–331

Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. Nat Genet 40(3):340

Beaudet L, Bédard J, Breton B, Mercuri RJ, Budarf ML (2001) Homogeneous assays for single-nucleotide polymorphism typing using AlphaScreen. Genome Res 11(4):600–608

Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Conneally PM, Duceman B (2005) DNA identifications after the 9/11 World Trade Center attack. Science 310(5751):1122–1123

Børsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, Morling N (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing. Forensic Sci Int Genet 2(4):292–300

Borsting C, Morling N (2011) Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. Forensic Sci Int Genet 5:236–241

Branicki W, Liu F, van Duijn K, Draus-Barini J, Pośpiech E, Walsh S, Kupiec T, Wojas-Pelc A, Kayser M (2011) Model-based prediction of human hair color using DNA variants. Hum Genet 129(4):443–454

Brookes AJ (1999) The essence of SNPs. Gene 234(2):177–186

Budowle B (2004) SNP typing strategies. Forensic Sci Int 146:S139–S142

Budowle B, Van Daal A (2008) Forensically relevant SNP classes. BioTechniques 44(5):603–610

Buetow KH, Edmonson MN, Cassidy AB (1999) Reliable identification of large numbers of candidate SNPs from public EST data. Nat Genet 21(3):323–325

Bulbul O, Phillips C, Argac D, Shahzad MS, Fondevilla M, Acar E, Aradas AF, Filoglu G, Altuncul H (2009) Internal validation of 29 autosomal SNP multiplex using a ABI 310 genetic analyser. Forensic Sci Int Genet Suppl Ser 2(1):129–130

Butler JM (2011) Advanced topics in forensic DNA typing: methodology. Academic Press, pp 347–362

Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Sci Med Pathol 3(3):200–205

Cai H, White PS, Torney D, Deshpande A, Wang Z, Marrone B, Nolan JP (2000) Flow cytometry-based minisequencing: a new platform for high-throughput single-nucleotide polymorphism scoring. Genomics 66(2):135–143

Canturk KM, Emre R, Kınoglu K, Başpınar B, Sahin F, Ozen M (2014) Current status of the use of single-nucleotide polymorphisms in forensic practices. Genet Test Mol Biomarkers 18(7):455–460

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SP (1996) Accessing genetic information with high-density DNA arrays. Science 274(5287):610–614

Cho S, Yu HJ, Han J, Kim Y, Lee J, Lee SD (2014) Forensic application of SNP-based resequencing array for individual identification. Forensic Sci Int Genet 13:45–52

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29(2):229–232

do Valle-Silva G, de Souza FDN, Marcorin L, Pereira ALE, Carratto TMT, Debortoli G, de Oliveira MLG, de Aguiar Fracasso NC, de Andrade ES, Donadi EA, Norton HL (2019) Applicability of the SNPforID 52-plex panel for human identification and ancestry evaluation in a Brazilian population sample by next-generation sequencing. Forensic Sci Int Genet 40:201–209

Fondevila M, Phillips C, Naveran N, Fernandez L, Cerezo M, Salas A, Carracedo A, Lareu MV (2008) Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. Forensic Sci Int Genet 2(3):212–218

Fondevila M, Børsting C, Phillips C, De La Puente M, Carracedo A, Morling N, Lareu MV, EN Consortium (2017) Forensic SNP genotyping with SNaPshot: technical considerations for the development and optimization of multiplexed SNP assays. Forensic Sci Rev 29(1):57–76

Frudakis T, Venkateswarlu K, Thomas M, Gaskin Z, Ginjupalli S, Gunturi S, Ponnuswamy V, Natarajan S, Nachimuthu PK (2003) A classifier for the SNP-based inference of ancestry. J Forensic Sci 48(4):771–782

Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu FL, Yang HM, Ch'ang LY, Huang W, Liu B, Shen Y, Tam PKH, Tsui LC, Waye MMY, Wong JTF, Zeng CQ (2003) The international HapMap project. Nature 426(6968):789–796

Giesen U, Kleider W, Berding C, Geiger A, Ørum H, Nielsen PE (1998) A formula for thermal stability (T m) prediction of PNA/DNA duplexes. Nucleic Acids Res 26(21):5004–5006

Gill P (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int J Legal Med 114(4):204–210

Gilles PN, Wu DJ, Foster CB, Dillon PJ, Chanock SJ (1999) Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. Nat Biotechnol 17(4):365–370

Griffin TJ, Smith LM (2000) Single-nucleotide polymorphism analysis by MALDI–TOF mass spectrometry. Trends Biotechnol 18(2):77–84

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr Sci 80(4):524–535

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4(5):e1000074

Huang TJ, Liu M, Knight LD, Grody WW, Miller JF, Ho CM (2002) An electrochemical detection scheme for identification of single nucleotide polymorphisms using hairpin-forming probes. Nucleic Acids Res 30(12):e55–e55

Hüebner C, Petermann I, Browning BL, Shelling AN, Ferguson LR (2007) Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example. Cancer Epidemiol Prev Biomarkers 16(6):1185–1192

Jehan T, Lakhanpaul S (2006) Single nucleotide polymorphism (SNP)–methods and applications in plant genetics: a review. IJBT 05(4):435–359

Jin X, Zhang X, Shen C, Liu Y, Cui W, Chen C, Guo Y, Zhu B (2020) A highly polymorphic panel consisting of microhaplotypes and compound markers with the NGS and its forensic efficiency evaluations in Chinese two groups. Gene 11(9):1027

Kayser M, Liu F, Janssens ACJ, Rivadeneira F, Lao O, van Duijn K, Vermeulen M, Arp P, Jhamai MM, van IJcken, W.F. and den Dunnen, J.T. (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet 82(2):411–423

Kheterpal I, Mathies RA (1999) Peer reviewed: capillary Array electrophoresis DNA sequencing. Anal Chem 71(1):31A–37A

Khlestkina EK, Salina EA (2006) SNP markers: methods of analysis, ways of development, and comparison on an example of common wheat. Russ J Genet 42(6):585–594

Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, Kungulilo S, Kim JJ, Lu RB, Odunsi A, Okonofua F (2006) Developing a SNP panel for forensic identification of individuals. Forensic Sci Int 164(1):20–32

Kiesler KM, Vallone PM (2013) Allele frequencies for 40 autosomal SNP loci typed for US population samples using electrospray ionization mass spectrometry. Croat Med J 54(3): 225–231

Kinoshita-Kikuta E, Kinoshita E, Koike T (2002) A novel procedure for simple and efficient genotyping of single nucleotide polymorphisms by using the Zn^{2+}-cyclen complex (vol 30, art no e126, 2002). Nucleic Acids Res 30(24):5593–5593

Kirk BW, Feinsod M, Favis R, Kliman RM, Barany F (2002) Single nucleotide polymorphism seeking long term association with complex disease. Nucleic Acids Res 30(15):3295–3311

Kota R, Wolf M, Michalek W, Graner A (2001) Application of denaturing high-performance liquid chromatography for mapping of single nucleotide polymorphisms in barley (Hordeum vulgare L.). Genome 44(4):523–528

Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics 2012

Kwok PY, Chen X (2003) Detection of single nucleotide polymorphisms. Curr Issues Mol Biol 5: 43–60

Li R, Norman S, Schober J (2015) Forensic biology. CRC Press, pp 333–343

Liu F, van Duijn K, Vingerling JR, Hofman A, Uitterlinden AG, Janssens ACJ, Kayser M (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19(5): R192–R193

Lyamichev V, Mast AL, Hall JG, Prudent JR, Kaiser MW, Takova T, Kwiatkowski RW, Sander TJ, de Arruda M, Arco DA, Neri BP (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. Nat Biotechnol 17(3): 292–296

Nikiforov TT, Rendie RB, Goelet P, Rogers YH, Kotewicz ML, Anderson S, Trainor GL, Knapp MR (1994) Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. Nucleic Acids Res 22(20):4167–4175

Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U (1994) Padlock probes: circularizing oligonucleotides for localized DNA detection. Science 265(5181): 2085–2088

Norrgard K, Schultz J (2008) Using SNP data to examine human phenotypic differences. Nature Education 1(1):85

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5(4):874–879

Ørum H, Nielsen PE, Egholm M, Berg RH, Buchardt O, Stanley C (1993) Single base pair mutation analysis by PNA directed PCR clamping. Nucleic Acids Res 21(23):5332–5336

Pakstis AJ, Speed WC, Fang R, Hyland FC, Furtado MR, Kidd JR, Kidd KK (2010) SNPs for a universal individual identification panel. Hum Genet 127(3):315–324

Phillips C (2009) SNP databases. Methods Mol Biol 578:43–71. https://doi.org/10.1007/978-1-60327-411-1_3

Phillips C, Amigo J, Carracedo Á, Lareu MV (2015) Tetra-allelic SNPs: informative forensic markers compiled from public whole-genome sequence data. Forensic Sci Int Genet 19: 100–106

Roewer L (2013) DNA fingerprinting in forensics: past, present, future. Investig Genet 4(1):1–10

Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27(9):1713–1724

Smigielski EM, Sirotkin K, Ward M, Sherry ST (2000) dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 28(1):352–355

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, Manolescu A, Karason A, Palsson A, Thorleifsson G, Jakobsdottir M (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39(12):1443–1452

Tishkoff SA, Verrelli BC (2003) Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. Curr Opin Genet Dev 13(6):569–575

Tomas C, Diez IE, Moncada E, Børsting C, Morling N (2013) Analysis of 49 autosomal SNPs in an Iraqi population. Forensic Sci Int Genet 7(1):198–199

Tyagi S, Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14(3):303–308

Tyagi S, Bratu DP, Kramer FR (1998) Multicolor molecular beacons for allele discrimination. Nat Biotechnol 16(1):49–53

Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A, Branicki W, Kayser M (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7(1):98–115

Yousefi S, Abbassi-Daloii T, Kraaijenbrink T, Vermaat M, Mei H, van't Hof P, van Iterson M, Zhernakova DV, Claringbould A, Franke L, M't Hart L (2018) A SNP panel for identification of DNA and RNA specimens. BMC Genomics 19(1):1–12

Ethical Governance of Forensic DNA Databases in Southeast Asia

11

M. C. A. De Ungria and E. B. Jimenez

Contents

M. C. A. De Ungria (✉)
DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines Diliman, Quezon, Philippines

Program on Biodiversity, Ethnicity, and Forensics, Philippine Genome Center, University of the Philippines Diliman, Quezon, Philippines
e-mail: madeungria@up.edu.ph

E. B. Jimenez
National Institutes of Health, University of the Philippines Manila, Manila, Philippines
e-mail: ebjimenez@up.edu.ph

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_9

Abstract

Over the past 36 years since its discovery, DNA profiling has contributed to improving forensic investigations by helping to identify human sources of biological samples. A critical part of the identification process is a comprehensive understanding of the genetic variation of forensically relevant DNA markers in human populations. Hence the continuous need to undertake DNA databasing research that is rigorous, intensive, and inclusive.

In DNA database research, the treatment of human subjects is commonly based on international ethical guidelines such as the 2013 Declaration of Helsinki, the 2016 Declaration of Taipei, and the 2016 CIOMS Guidelines. DNA research necessitates an understanding of the sensitive nature of DNA itself since its effects permeate through individual, familial, and community concerns. Informed consent, data protection, independence of ethics committees, funding and benefit-sharing, and public engagement are discussed using a Southeast Asian lens. Overall, the ethical governance of forensic DNA database research in the ten countries that make up the ASEAN must consider the extensive genetic diversity and vulnerabilities of different populations and cultural groups in the region.

Research stakeholders will realize that presumptions about ethical elements pervasive in Western developed countries – where a simple yes may signify consent – are not immediately applicable to all populations, including those in Southeast Asia where customs, traditions, cultures, and languages vary. Therefore, greater responsibility for public engagement and dialogue must be pursued to ensure that all efforts are exerted to "do no harm" and actively contribute to community knowledge generation and security.

Keywords

Forensic DNA databases · Biobanks; Southeast Asia; Research publication · Research ethics · Human subject protection; Declaration of Helsinki · Declaration of Taipei and the Council for International Organizations of Medical Sciences; Genetic diversity · Inclusivity · Cultural vulnerability

Introduction

The discovery of DNA fingerprinting by Sir Alec Jeffreys in 1985 revolutionized how criminal investigations are being conducted. DNA typing involves examining length variation of DNA repeat sequences and comparing these length variations between two samples, one of which is of known origin (Jeffreys et al. 1985). Since then, DNA profiling or DNA testing has become the gold standard or the most powerful tool for human identification.

From extremely time-consuming Southern blots and X-ray films, discovering the power of thermostable DNA polymerases in amplifying minute amounts of DNA commonly found in crime scene evidence has further influenced the rapid development of forensic genetics. The polymerase chain reaction, or PCR, is now an integral

component of DNA testing in different investigations ranging from simple paternity suits, complex kinship analysis, genealogy or ancestry tests, human remains identification, disaster victim identification, and criminal investigations.

Due to the size of the human genome, only discrete, polymorphic regions are considered necessary for the routine identification of samples/persons. Forensic DNA testing may employ different markers depending on the availability of reference samples that would be used for comparison. The main groups of DNA markers used are short tandem repeats (STRs) on autosomes (aSTRs), on the Y chromosome (YSTRs), and on the X chromosome (XSTRs), the hypervariable regions I and II (HV1 and HV2) of the mitochondrial DNA (mtDNA), and different types of single nucleotide polymorphisms (SNPs). Descriptions of these DNA markers have been aptly covered in other chapters of this book.

Initially, STR-DNA typing used the single DNA marker system. Later developments gave rise to commercial genotyping kits that amplify several STR loci simultaneously, commonly referred to as multiplex reactions. Besides increased efficiency and faster turn-around time, multiplex kits improved the overall discrimination capacity resulting from the same amount of genetic material. The increasing use of next-generation sequencing (NGS), also referred to as massively parallel sequencing (MPS), is fast-tracking the field of forensic genetics at a pace that has not been seen before. MPS allows the simultaneous sequencing of hundreds of DNA markers while differentiating between isoalleles – alleles with the same length but different sequences – thus improving the combined power of discrimination rather than simply relying on length variation. In forensic investigations, the strength of DNA evidence depends on the estimated rarity of an evidentiary DNA profile in the population of interest using a DNA database. Hence, the degree of polymorphism of DNA markers and the number of genotypes are critical factors that determine the database's usefulness in forensic investigations.

There are three types of DNA databases, depending on the samples they contain and the purpose of the database. Many law enforcement agencies have a national database containing samples from crime scenes, arrestees, convicted offenders, prisoners, and missing persons used for investigative and identification purposes. This type of DNA database retains all sample information, including the identity of human sources of samples, if they are known. A list of 132 countries with national databases or those planning to pass laws prepared by the Forensic Genetics Policy Initiative is available online (http://dnapolicyinitiative.org/). The second type of DNA database, referred to as reference population databases, is produced while studying the degree of polymorphisms of specific DNA markers. Knowledge of the diversity of populations must be characterized as comprehensively as possible, and representative samples from all subpopulations must be included in establishing a functional reference population DNA database. If stratification between groups is detected, adjustments to statistical estimates of matching DNA data must be made (National Research Council 1996). Because the purpose of reference DNA databases is to understand the extent of genetic polymorphisms within and across different groups or populations, samples are anonymized, for example, identity information is removed from the samples and stored in separate facilities which are not readily accessible. In some international databases, such as the YHRD (https://yhrd.org) and EMPOP (https://empop.online) maintained in

Germany and Austria, respectively, the DNA profiles submitted by research institutions from different counties were completely de-identified. Unlike law enforcement DNA databases, the DNA profiles in reference databases are summarized as group data and are not useful for individual identification. Hence, law enforcement investigative DNA databases and population reference databases may not be used interchangeably because each repository's original purpose is different.

More recently, commercial companies created the third type of DNA database for genealogical research, that is, searching and connecting with persons in the database who have similarities with one's DNA profile and could potentially be a relative. Persons who had their DNA profiles generated by direct-to-consumer testing (DTC) such as 23andMe, Ancestry DNA, My Heritage, and Family Tree DNA could upload these profiles with names and other personal details in a publicly accessible DNA database like GEDMatch (Greytak et al. 2019). Other persons can then access this commercial database, search for persons similar to one's DNA profile, and evaluate possible relationships. Because samples are not anonymized, any person, including law enforcement, can use the personal information from a genealogical database to identify and locate persons of interest (POI) in a process called genealogical testing (Greytak et al. 2019).

Privacy concerns in the extent of participants' consent and the use of anonymized genetic data in international databases by law enforcement have been raised (Schiermeier 2021). In the USA, genetic data uploaded in GEDMatch by individuals seeking to understand their genealogy have been used effectively in cold case investigations (Greytak et al. 2019). In both instances, the arguments for using non-law enforcement DNA databases in specific types of forensic investigations were justified. Hence, to use non-law enforcement DNA databases for forensic investigations without compromising the ethical requirement for consent, countries should pass legislation or promulgate national policies that include safeguards to balance genetic privacy concerns with public security.

This chapter aims to take the discussion further by reviewing the state of research into establishing reference population databases, focusing on the ten member nations of the Association of Southeast Asian Nations (ASEAN). In particular, the five ethical elements of population genetics research for forensics are presented through the lens of local researchers of the region. Available literature about the region was surveyed, and consultations with scientists and policy colleagues were made whenever possible. Recommendations to help the governance of research on reference DNA databases for forensic applications in the ASEAN member nations are discussed. These recommendations may also be relevant to countries that face similar challenges.

Research Involving Human DNA Databases

Research involving human DNA databases falls within a subset of research that international ethical guidelines call health-related research. These databases contain personal human data that are accessed to analyze information and generate

knowledge. The term health-related research may confuse some scientists, especially basic scientists, who work predominantly in laboratory environments and may or may not be involved directly with human participants from whom specimens are collected. But the term health-related is used to cover a wider range of research that uses human data (Council for International Organizations of Medical Sciences 2016).

The World Health Organization (WHO) defines health to include not just the physical absence of disease. Health embraces a more holistic consideration for a person's well-being influenced by factors that may be socially and culturally determined (World Health Organization 1946). The information contained in a single DNA sample contains the person's entire genetic code and his family or community. All these people related to that genetic information become vulnerable every time the information is accessed. While such harm may not be immediately physical, informational vulnerability can be a precursor to social harm. So long as the person's genetic information continues to be stored and used for research, injury from social harm can threaten persons' overall well-being.

Because of the broad and intergenerational impact of genetic data, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) International Declaration on Human Genetic Data granted it special status (United Nations Educational Scientific and Cultural Organization 2003). It states:

> . . .they [DNA data] can be predictive of genetic predispositions concerning individuals; they may have a significant impact on the family, including offspring, extending over generations, and in some instances on the whole group to which the person concerned belongs; they may contain information the significance of which is not necessarily known at the time of the collection of the biological samples, and they may have cultural significance for persons or groups.

Such status underscores the need to apply universally accepted ethical principles to assign specific responsibilities to stakeholders in the research enterprise. The stakeholders are the researcher, the research institutions under whose oversight the research is conducted, the funding agencies, the research ethics committees that approve the involvement of human participants, and scientific publications that disseminate the research results to the public domain.

Ethical Elements Applied to Genetic Research and DNA Databasing

Overall, applying ethical elements promotes an ethical governance system underscored by the importance of protecting human participants in research. That protection of the interests of human participants overrides the scientific agenda of research and even the sociopolitical aspirations of society in general. Nobody can argue against the scientific and social value of research in ensuring social progress. However, there is a "moral obligation to ensure that all research is carried out in a way that upholds human rights, and respect, protect, and are fair to study participants

and the communities in which the research is conducted" (Council for International Organizations of Medical Sciences 2016). Infringing these rights constitutes an injustice that scientific and social value cannot be justified, no matter how high.

The application of ethical elements may be ascertained during the different stages of the research. The process begins with the inclusion criteria or profile of target populations, rationalized through the scientific objectives of the study that will justify its social value (societal benefit). This step will be followed by recruiting genetic donors for the research, including explanations of direct individual and familial benefits (if any) and risks consisting of potential harms posed by the methods employed. Finally, the researcher must also discuss the outcome, for example, what will happen to the results or what future access and use can be authorized, including those by third parties. All these questions imply the establishment of protective mechanisms throughout the research.

These issues are not new. Various stakeholders have numerous publications on the governance of research involving genetic databases (Wickenheiser 2019). Recommendations highlight the need for "respectful protocols" when engaging individuals, drafting consents that take into consideration all potential uses of information, declaration of commercial interests, coding of interlinking of databases as part of privacy issues, data confidentiality, and security safeguards, and defensive strategies for access of "personal data by public health authorities, police, courts, employers, lenders, insurers, and subjects' relatives" (Lowrance 2001). Several papers assessed the adequacy of available governance frameworks (Gibbons et al. 2007; Hayry et al. 2007). And the European Observatory on Health Systems and Policies issued a policy brief in 2021 calling this domain "regulation of the unknown." It concluded that appropriate regulation is key to addressing the risk of misuse of genomic data and promoting anticipated benefits to people and society (Williams et al. 2021).

Regulations are arguably an important tool to provide appropriate research guidance, and regulations exist in varying forms in different countries toward this end. Regulatory requirements of each country are expected to be followed by researchers, whether they are local or foreign researchers. Concomitantly, substantial progress may be achieved by a commitment to a common set of standards that are acceptable to different countries. To this end, we are proposing a multiple stakeholder framework to facilitate the researcher's work toward ethical research involving human genetic databases. The framework is adapted from a guideline-based assessment of ethical, legal, and social issues, composed of five elements, namely: 1) informed consent; 2) data protection; 3) independence of ethics committees; 4) funding and benefit-sharing; and 5) public engagement (Austin et al. 2003). We propose to reappropriate these ethical elements toward an updated framework that applies to the proper and socially relevant use of genetic databases by citing illustrative examples from recent history that will show their continuing significance.

Informed consent. Consent is one of the most regulated aspects of health research, in conjunction with data protection. Informed consent is the cornerstone of ethical research (Tindana et al. 2020). It guarantees the freedom of a person to decide to be a research participant or refuse without penalty or coercion, directly or

indirectly. The consent process established the compass that guides research design directions and decisions.

The ethical acceptability of the consent process used in research is determined by the competence of a person to consent (commonly related to age and mental capacity), disclosure of a mandatory set of information that must be communicated to the research participant, ensuring that the information communication is conveyed in a form that will facilitate understanding (such as the use of local language and appropriate cultural context), absence of coercion whether the direct threat or undue influence (such as large sums of money), and unambiguous articulation of agreement (World Medical Association 2016). The fulfillment of these requirements is structured by regulatory requirements and cultural and social norms in an informed consent form.

Informed consent comes in many forms. The most relevant form for collecting, storing, and using biological materials and archiving genetic information is broad informed consent. This type of consent specifies: 1) the purpose of collection and the conditions and duration of storage; 2) the rules in accessing either the stored biological materials or genetic information; 3) the ways in which the donor can remain informed about future use of genetic information and specimen; 4) the foreseeable uses of the materials or information, that is, whether this is limited to the current study or to other future studies (which may be unknown at the time of consent); 5) whether such future uses are restricted only to a research purpose or may have commercial intentions; 6) whether the participant will receive compensation from commercial products developed from their biological specimens or genetic information; 7) the possibility of incidental findings and how they will be resolved; 8) the systems in place that will protect confidentiality as well as their limitations; and 9) whether biological specimens collected in the research will be destroyed at some point, and, if not, details about their storage (where, how, for how long, and final disposition) and possible future use, that participants have the right to decide about such future use, to refuse storage, and to have the material destroyed (Council for International Organizations of Medical Sciences 2016).

The archiving of genetic material and compilation of information are imbued with overlapping accountabilities both to the public and the people who donated the genetic material. Ethically suspicious use of a DNA database beyond the consent provided for by participants can eventually invalidate an otherwise socially and globally valuable scientific endeavor (Schiermeier 2021). However, with the appropriate safeguards included in a DNA law or national policy, the forensic use of non-law enforcement databases to balance genetic privacy concerns and public security can maximize the use of DNA technology. The increased effectiveness in policing because of powerful but costly DNA technologies is particularly valuable in low- and middle-income countries.

Data protection. We cannot overemphasize the reality that an individual's genetic information will carry sensitive private information that can transcend even generational boundaries. Therefore, as a matter of good research practice, researchers using genetic material and information must devise a coding system

and prevent unauthorized linking of such information to its human donor. Potential harm can come in many forms.

System vulnerability can be illustrated by de-anonymization attack experiments, which concluded that the severity of this threat goes hand in hand with anticipated progress in database technology (Seeman 2019). Unauthorized access in the absence of strict protocols for use, reuse, and future of both bio-banked specimen and databased information can result in unauthorized access outside the terms of reference surrounding the specimen and the data (Phillips et al. 2020). Once processed and analyzed, it would be difficult to withdraw information, even if access cannot be legitimately justified.

Then, there is the potential harm from incidental findings (Williams et al. 2021), so the database developers must put a mitigation plan in place and appropriate measures initiated for proper handling of such findings. Anonymizing is critical, and restricting access to the database is essential. Thus, international ethical guidelines stress the urgent compliance with protecting the confidentiality of information collected from participants.

Therefore, good research plans must anticipate as early as the developmental phase, the potential for unauthorized disclosure, intention for future or secondary use, and necessary approvals, customized to the available country regulations. Most countries have established regulations to protect data. According to the United Nations Conference on Trade and Development report, 66% of countries worldwide have data privacy legislation (https://unctad.org/page/data-protection-and-privacy-legislation-worldwide). Only Indonesia, Laos, Malaysia, the Philippines, Singapore, Thailand, and Vietnam have data privacy legislation in the ASEAN. Myanmar has drafted one, whereas Brunei and Cambodia provided no information.

Independence of ethics committees. Research ethics committees (REC) are known by different names in various countries and institutions such as ethical review board (ERB), ethical review committee (ERC), human research ethics committee (HREC), and institutional review board (IRB). No matter the label, the foundational element characteristic should be a group of individuals who undertake the ethical review of research protocols involving humans, applying agreed ethics elements (World Health Organization 2011). Notably, some IRBs do not include an ethics review in their procedures. Hence in this instance, the researchers must submit their proposal to a separate REC. It is also critical that the researchers know the scope of review on their proposal before initiating the project. Relevant country authorities must ensure that an adequate legal framework supports ethics review of health-related research, that RECs are established according to an official charter, that members include both scientific experts and nonscientific individuals who represent the aspirations of common people; and policies must ensure REC independence through the inclusion of non-affiliated members, management of conflict of interest (of researchers, institutions, and funding agencies), and preclusion of senior decision-makers of the institution to serve as members (World Health Organization 2011). The ethics committee is expected to be competent in both internationally agreed principles and guidelines and local regulations, customs, and language to sustain its function of human participant protection.

Funding and benefit-sharing. The benefit from research involving genetic databases is largely reckoned from the scientific and social value of the research because these projects can rarely provide a direct benefit to individuals. Research participants who donate biological specimens and agree to their use are entitled to know how their specimens and information will be used, and which entities are funding the project, whether these entities are local or foreign, commercial, or not for profit, and whether these entities can protect their interests. The key to a common understanding of these issues is a genuine informed consent process, documented in an informed consent form, already discussed above. In some areas, limited resources may potentially affect the level of ethics review; hence institutions must provide mechanisms to prevent this from happening (World Health Organization Western Pacific Region 2012).

Public engagement. Public engagement contributes toward a better understanding of science. Educating the public with the correct information through active public engagement is critical in the subsequent acceptance of the technology. Continuous public engagement goes hand in hand with the societal benefit of research. This way, people who participate in genetic research or give their specimen are empowered, at least with information, so that social value still outweighs the risks for the genetic donors.

Engaging the public should involve a meaningful and sustainable process (Wallerstein and Duran 2006). In some countries, researchers develop a long-term partnership with communities and networks, with a sustained visible presence to assure communities of their continuing transparency and willingness to be accountable for research projects that they conduct (Cargo and Mercer 2008).

Ethical Framework for Multiple Stakeholder Responsibility

Invoking ethical principles in a scientific activity can be made more meaningful considering their implications in defining the roles of various stakeholders that make research possible. The stakeholders include the researchers, research institutions, funding agencies, review ethics committees, and scientific journals. All have shared accountability in ensuring the social and scientific value of research is realized without compromising the well-being of research participants, whether in donating biological specimens or providing genetic information. This shared accountability applies to all genetic research fields, such as population studies, biomedical projects, and forensic genetics.

Role of the Researcher. Researchers are primarily responsible for designing the research plan and conducting the study following approvals related to its scientific and ethical acceptability, and eventually, for research outputs and outcomes. The researcher's responsibility is to verify applicable regulations to sites and populations affected by their projects and incorporate these regulatory and population-specific requirements in the research plan. In research involving genetic databases, a critical responsibility of the researcher is the obligation to comply with country regulations, including engagements or communications with target populations, country permits,

and applications for ethical review and approval. In case these critical oversight systems are not tightly regulated, foreign researchers collaborating with local research institutions must respect local requirements and "refrain from taking active or passive advantage of loopholes and weaknesses in the governance systems of another country" (D'Amato et al. 2020).

The role of the researcher is highlighted in ensuring compliance with the ethical requirements for data protection and informed consent because it is the researcher who engages the participation of people in the research project and who is accountable for research results.

Role of Research Institutions. The research institution has administrative oversight among affiliated researchers, and in this regard, has command responsibility to ensure compliance with scientific standards, regulatory norms, and ethical principles. In general, this means fostering an enabling environment that encourages responsible and respectful researchers, establishing scientific and ethical resources, and creating a monitoring system to ensure sustainability. The institutional responsibility serves as a sentinel system toward achieving the public good promised by scientific research.

Role of Funding Agencies. Funding is the bloodline for any research. Therefore, funding agencies are strategically positioned to promote research integrity and foster trust, demanding specific progress at any stage of the research cycle. Since funding agencies hold the power of the purse, this power carries with it the responsibility for uprightness and transparency. Fostering public trust is pivotal in sustaining future research.

Role of Research Ethics Committees. As part of the research plan, health-related research involving humans must be submitted to a research ethics committee (REC) for review of ethical acceptability or qualification for an exemption from ethical review depending on the nature of the research and applicable regulations (Council for International Organizations of Medical Sciences 2016). The work of RECs is, in general, a review of research-related documents, including qualifications of the researchers, not just in terms of scientific expertise but also in ethical competence. This review fulfills the primary responsibility of ethics committees to protect human participants in research. To fulfill this mandate, ethics committees must be properly constituted, according to the prescription in international guidelines and local policies; have standardized procedures that will allow predictable outcomes; and consolidate all considerations applicable to the study. The authority of the ethics committee is an extension of the authority of the body that created it. International ethical guidelines that are commonly cited in the procedures of most RECs are the World Health Organization Standards and Operational Guidance for Ethics Review of Health-related Research with Human Participants (World Health Organization 2011), World Medical Association Declaration of Helsinki (World Medical Association 2013), World Medical Association Declaration of Taipei (World Medical Association 2016), and the Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines for Health-related Research Involving Humans (Council for International Organizations of Medical Sciences 2016).

Table 1 highlights provisions relevant to genetic databases of three of these guidelines that are addressing a wide range of researchers – the Declaration of Helsinki, the Declaration of Taipei, and the CIOMS Guidelines – reckoned from the ethical elements of informed consent, data protection, independence of RECs, funding, and benefit-sharing, and public engagement. The table offers an overview of guideline-based responsibilities of stakeholders related to the conduct of research involving DNA databases. The Declaration of Helsinki and the Declaration of Taipei, underscored by the mandate of the World Medical Association, are primarily addressing physicians who may be doing research involving patients. The Helsinki document laid down the foundational ethical elements for biomedical research. The Taipei Declaration supplemented the Helsinki Declaration with the additional guidelines for collecting, storing, and using identifiable data and biological materials used in health databases and biobanks. The CIOMS guidelines particularized the provisions further by extending the principles to specific research contexts and emerging consensus. The WHO guidelines specifically formulated for RECs to ensure competent review harmonized with acceptable international guidelines were not included in the comparative analysis.

Role of Scientific Journals. Scientific journals safeguard data release to the public by ensuring that the researchers had observed accepted scientific methods and practical ethical elements (Fig. 1). Most journals rely on the Declaration of Helsinki as the benchmark document for protecting human participants in research. The Declaration of Helsinki first emphasized the ethical requirement of making research public during the assembly of the World Medical Association in 1975. Thus, most journals require authors to submit a statement that they had complied with the Declaration of Helsinki. This directive delineated the ethical obligations of authors, sponsors, editors, and publishers, including disseminating research results, accepting accountability for completeness and accuracy of reports, adhering to accepted guidelines for ethical reporting, and declaring conflicts of interest (World Medical Association 2016). However, the Declaration of Helsinki is addressed primarily to physicians conducting research on their patients based on the mandate of the WMA, thus emphasizing medical research in its provisions (Table 1).

However, decades after adopting the Declaration of Helsinki, research involving human participants had progressed substantially in objectives, methodology, and available technology. Health-related research involving humans now covers many disciplines beyond medical research, and concomitantly, a new ethical landscape had emerged with new issues and concerns. The Declaration of Helsinki remains relevant, but for studies beyond medical research, such as establishing and expanding reference DNA databases for forensic applications, the Declaration of Taipei and CIOMS represent the more applicable guidance documents. The "International Ethical Guidelines for Health-related Research Involving Humans" issued provisions for the collection, storage, and use of biological materials and related data, which can be cross-referenced with guidance for community engagement and assessment of scientific and social value vis-à-vis respect for human rights (Council for International Organizations of Medical Sciences 2016).

Table 1 Comparison of International Ethical Guidelines commonly adopted by Research Ethics Committees

	WMA Declaration of Helsinki (2013)	WMA Declaration of Taipei Research on Health Databases, Big Data, and Biobanks (2016)	CIOMS International Ethical Guidelines for Health-Related Research Involving Humans (2016)
Nature of organization formulating the guidelines	• Formulated by the World Medical Association (WMA) • An international organization of physicians • Established in 1947	• Formulated by the World Medical Association (WMA) • An international organization of physicians • Established in 1947	• Formulated by the Council for International Organizations of Medical Sciences (CIOMS) • An international NGO • Founded in 1949 under the auspices of WHO and UNESCO
Audience/subject of the guidelines	• Addressed primarily to physicians • Encourages those in medical research involving human subjects to adopt the principles	• Addressed primarily to physicians • Encourages those using data or biological material in health databases and biobanks to adopt these principles	• Addressed to all researchers • Collaborative relations with the United Nations and its specialized agencies, especially UNESCO and WHO
Purpose	• The objective of the document is captured in the following words: "Statement of ethical principles for medical research involving human subjects, including research on identifiable human material and data."	• Complements the Declaration of Helsinki • Focused on the collection, storage, and use of identifiable data and biological material for use in health databases and biobanks beyond individual patient care	• Synthesizes authoritative guidelines, documents, and literature on ethical frameworks for human subject research directed toward emerging viewpoints
Informed consent	• Must be voluntary and given by those with the capability to consent. Subjects must receive adequate information on: a) study aims and methods b) sources of funding c) potential conflicts of interest, institutional affiliations d) possible benefits and risks resulting from participating on e) post-study provisions and f) other relevant study aspects	Additional requirements to those of the Declaration of Helsinki: • Specific consent for the collection, storage, and use of identifiable data and biological material • Plans to store sample and information in a health database or a biobank for multiple and indefinite uses: a) purpose of the database or biobank b) risks and burdens associated with the collection, storage, and use of data and material	• Informed consent is a continuous researcher's duty and allows withdrawal at any time unless a REC has approved a waiver • Waiver of consent can be granted if research has important social value, pose minimal risk, and would not be feasible without it • Types of consent can be obtained from a research participant willing to provide specimen/s and related data: a) specific, informed consent for a

	Right to refuse to participate and withdraw consent without reprisal	c) nature of the data or material to be collected d) procedures for return of results including incidental findings e) rules of access to the database or biobank f) privacy protection g) governance arrangements h) any intention to make data and samples non-identifiable, which means the loss of the subject's option to withdraw consent and learn about the results of the study i) fundamental rights and safeguards established in this Declaration • Commercial use and benefit-sharing, intellectual property issues, and transfer to third countries/parties (if any)	particular use or b) broad informed consent for unspecified future use c) informed opt-out (in place of a full consent process) if future research is minimal risk or excludes contexts of heightened vulnerability • Specimens may be tissues, organs, blood, plasma, skin, serum, DNA, RNA, proteins, cells, hair, nail clippings, urine, saliva, or other bodily fluids; from living or dead human or bodily wastes or abandoned tissue. • Need for specific policies for the collection, storage, and access of biological specimens and data; in collaboration with local health authorities
Data protection	Every precaution must be taken to • protect the privacy of research subjects and • the confidentiality of their personal information	Recommends adopting procedures • to protect the dignity, autonomy, and privacy of the individuals and • to safeguard the data derived from the analysis of samples	• Storing biological material requires a guarantee of its confidentiality. Information from an analysis of the material could cause harm, stigma, or distress • Protection mechanism, to be included in the informed consent form, can be done by: a) anonymizing data released to researchers b) limiting access to the material • Specimen and data donors must be informed of the limits of confidentiality mechanisms of the protocol and the

(continued)

Table 1 (continued)

	WMA Declaration of Helsinki (2013)	WMA Declaration of Taipei Research on Health Databases, Big Data, and Biobanks (2016)	CIOMS International Ethical Guidelines for Health-Related Research Involving Humans (2016)
			consequences of a potential breach • A custodian is recommended to establish a governance system for data protection. The custodian must find ways to reduce the impact when large datasets are crossmatched, thus limiting the ability to anonymize data fully
Independence of research ethics committees (REC)	• Researchers must submit the research protocol for review and approval of an appropriate REC before the study begins. • The REC must have the following: a) qualified members b) transparent procedures independent of the researcher, the sponsor, and any other undue influence • The review must examine the protocol's compliance with the laws and regulations of the country where research will be done, and applicable international norms and standards • The review must be continuous while the study is ongoing. The REC must	Additional requirements to those of the Declaration of Helsinki: • The establishment of health databases and biobanks used for research and other purposes must be submitted for review • The REC must approve the use of data and biological material only after assessing the sufficiency of the subject's consent for the proposed research protocol	• Research involving humans must be submitted to a REC. • The REC must approve the research before the study begins • REC must be formally established with the following: a) qualified members b) adequate mandate to ensure timely and competent review c) governed by transparent procedures d) mechanisms to ensure independent operations • REC from different institutions or countries should establish efficient processing of multi-center research

	require the researcher to provide information for monitoring purposes • Regulations must not be allowed to reduce or eliminate any protections for research subjects guaranteed by the Declaration		• External studies must have an ethical review from both the host and the sponsoring institutions
Funding agencies and benefit-sharing	• Sources of funding, institutional affiliations, and conflicts of interest must be declared in research publications • Participation of vulnerable groups must be justified • The benefit from the knowledge, practices, or interventions that result from the research must be shared with the participants	• Benefit-sharing must consider the interests and rights of the communities concerned • Research and other health databases and biobanks related activities should contribute to the benefit of society and public health objectives	• Policies for managing conflicts of interest (COI) must be transparent and actively communicated to those affected to reduce potential harm • COI can be any of the following: a) professional b) academic c) financial interests • COI may be managed through a) education of researchers and REC b) disclosure to REC c) disclosure to participants d) mitigation of conflicts, e.g., limiting the involvement of those with serious conflicts • Collecting data from low-resource settings should include the following: a) collaboration with local health authorities b) the databank governance structure must have representation of the original setting

(continued)

Table 1 (continued)

	WMA Declaration of Helsinki (2013)	WMA Declaration of Taipei Research on Health Databases, Big Data, and Biobanks (2016)	CIOMS International Ethical Guidelines for Health-Related Research Involving Humans (2016)
Public engagement	• Research studies involving human subjects must be registered in a publicly accessible database before recruitment • Researchers should ensure that new information is recorded and made publicly available (as appropriate)	• Custodians of health databases and biobanks must be accessible and responsive to stakeholders • Health databases and biobanks must be governed following the principles of: a) protection of individuals: the rights of individuals prevail over science and other interests b) transparency: make public all useful information on health databases and biobanks c) participation and inclusion: custodians of health databases and biobanks must engage with individuals and their communities	• Public accountability is necessary to achieve the social and scientific value of health-related research • Researchers, sponsors, research ethics committees, funders, editors, and publishers share the obligation to comply with recognized publication ethics

An overview of the ethical requirements of scientific journals is typically outlined in the journal's guide to authors section, which contains instructions on publication ethics and the protection of humans that participated in the research. We had surveyed the author's guidelines across six science journals that frequently publish research on forensic DNA databases (Table 2).

Based on the policy scan, these journals require compliance with the Declaration of Helsinki, which encapsulates the general ethical considerations for human participants in research. These requirements include: a) REC review and approval; b) statement on the informed consent of participants; c) respect for privacy and protection of confidentiality; d) compliance with local regulations; e) consent for publication of identifiable information (if applicable); f) any conflict(s) of interest on the part of the authors; g) that all research procedures comply with relevant local laws and guidelines; and h) approvals from the relevant institutional or national committee(s). Authors are instructed to retain documentation or evidence of such consents, which should be provided to the journal as needed.

In addressing publication inconsistencies, clarificatory recommendations were proposed by D'Amato et al. (2020). These new guidelines anticipated the ethical issues of research involving genetics databases, particularly those used for forensic applications. The ethical guidance contained therein more or less reflects the issues articulated by other researchers working in various locations around the world. Additional authoritative guidelines were cited as applicable, such as the WMA Declaration of Taipei (World Medical Association 2016) and the CIOMS (Council for International Organizations of Medical Sciences 2016) guidelines. The central message remains constant: scientific publications must act as gatekeepers of information to ensure that what they endorse for public consumption is predicated on ethically acquired samples and biodata.

Research on Reference Databases on Forensically Relevant DNA Markers in Southeast Asia

The Association of Southeast Asian Nations or ASEAN is an international organization formed in 1967 by the Republic of Indonesia, Malaysia, the Republic of the Philippines, the Republic of Singapore, and the Kingdom of Thailand aimed to accelerate economic growth, social progress, and cultural development, and secondarily to promote regional peace and security based on the rule of law and the principles of the United Nations charter (https://asean.org/). Other countries in the region have since joined the organization: Brunei in 1984, Vietnam in 1995, Laos and Myanmar in 1997, and Cambodia in 1999 (Chandler and Overton 2021). In 2007, the member countries signed the ASEAN Charter, which entered into force in December 2008 following the ratification by all ten member states. The Charter established the organization's legal entity to promote economic and cultural exchange within the entire region. As of 1 July 2019, more than 600 million people lived in the ASEAN member states, comprising 8.5% of the world's population (Moon 2020).

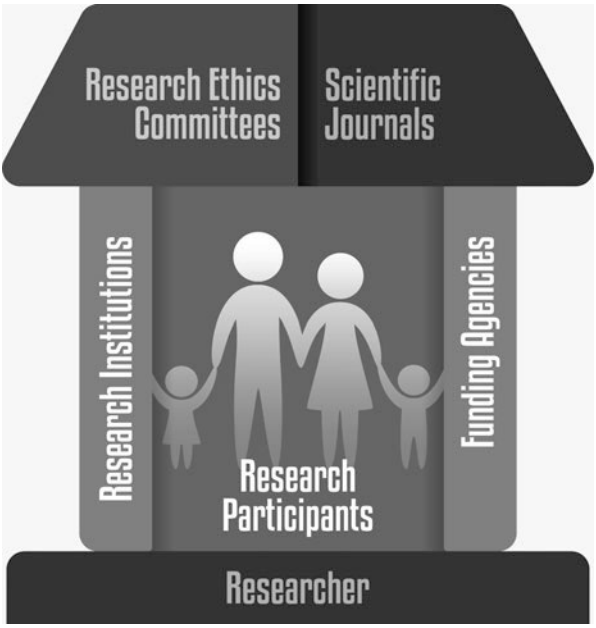


Fig. 1 Diagrammatic representation of the stakeholders of good scientific research. This figure represents the multi-stakeholder framework for the ethical conduct of research involving the use of genetic databases. The figure employs the metaphor of a house to represent the research enterprise signifying an area of safety, comfort, respect for persons, and empowerment for research participants sheltered within its confines. Five stakeholders comprise the building blocks of ethical research, namely 1) the researcher whose integrity lays the foundation for research and its benefit to society; 2) the research institutions which contribute the infrastructure as well as provide administrative oversight to researchers; 3) the funding agencies that support research activities; 4) the research ethics committees; and 5) the scientific journals as the roof that protects the house from external elements. The structure protects research participants from risks inherent in the conduct of research, supported by responsible researchers, research institutions, and funding agencies. The sturdiness of the structure is reinforced by competent research ethics committees that ensure the protection of human participants during research, and scientific publications that function as gatekeepers of information and products released to the public, by ensuring compliance with agreed scientific and ethical standards

Geographically close to each other, with some states sharing common borders (Fig. 2), the need to work together to maintain peace and order in the region is of utmost importance. Crime does not recognize borders; hence the formation of the ASEANOPOL in 1981 was a welcome development. This organization aims to enhance the professionalism of local police and push for stronger cooperation and policing by the same five original member nations of ASEAN was a welcome development (www.aseanapol.org).

With the development of DNA technology as an important forensic tool for human identification, a Regional East DNA Profiling Group (REAFD) was formed in Malaysia in 2007. The need to share information and use DNA testing procedures

Table 2 Survey of published literature on forensic DNA databasing research in Southeast Asia

		Brunei	Cambodia	Indonesia	Laos	Malaysia	Philippines	Singapore	Thailand	Vietnam	Total
Research	Total Number	2	1	6	2	17	23	5	30	5	91
Researchers	Local only	0	0	0	0	12	9	3	16	2	42
	Combined Local + Foreign	2	1	3	2	5	10	2	5	3	33
	Foreign only	0	0	3	0	0	4	0	9	0	16
Funding	Local only	0	0	0	0	12	6	3	11	2	34
	Combined Local + Foreign	0	0	0	0	1	9	0	2	1	13
	Foreign only	0	1	3	2	2	3	0	6	1	18
	No information provided[a]	2	0	3	0	2	5	2	11	1	26
Ethical Review before the country's ethics compliance cut off year[b]	Local only	1	0	0	1	0	1	2	0	2	7
	Combined Local + Foreign	0	0	0	0	0	2	0	0	1	3
	Foreign only	0	1	1	1	0	0	0	0	0	3
	No Information provided	0	0	2	0	7	14	3	15	1	42
Ethical Review after the country's ethics compliance cut off year	Local only	0	0	0	0	6	3	0	10	1	20
	Combined Local + Foreign	1	0	1	0	2	0	0	0	0	4

(continued)

Table 2 (continued)

		Brunei	Cambodia	Indonesia	Laos	Malaysia	Philippines	Singapore	Thailand	Vietnam	Total
	Foreign only	0	0	1	0	0	0	0	1	0	2
	No Information provided	0	0	1	0	2	3	0	4	0	10
Institutional Permission	Statement provided	0	0	1	0	5	4	0	1	0	11
	No Information provided	2	1	5	2	12	19	5	29	5	80
Informed Consent	Statement provided	1	1	2	2	7	8	1	8	4	34
	No Information provided	1	0	4	0	10	15	4	22	1	57
Conflict of Interest	Statement provided	1	1	4	0	6	4	0	10	3	29
	No information	1	0	2	2	11	19	5	20	2	62

Publication	Forensic Science International	0	0	0	0	2	3	4	3	0	12
	FSI: Genetics	0	0	1	0	0	2	0	3	1	7
	FSI: Genetic Supplement Series	0	0	1	0	0	1	0	2	0	4
	Journal of Forensic Sciences	0	0	1	0	6	3	0	8	0	18
	International Journal of Legal Medicine	0	0	0	0	7	1	0	5	2	15
	Legal Medicine	0	0	1	0	0	2	1	3	1	8
	Local Journals[a]	1	0	0	0	1	3	0	0	0	5
	Other Journals	1	1	2	2	1	8	0	6	1	22

[a]Local journals refer to journals that are published in the country where sampling was conducted

[b]The ethics compliance cut-off year used were as follows: Brunei (2009), Indonesia (2006), Malaysia (2006), Philippines (2013), Singapore (2015), Thailand (2007) and Vietnam (2013). Cambodia and Laos were not given cut-off years for ethics compliance, and Myanmar was not included in the table because there was no available publication on DNA databasing on samples collected locally

Fig. 2 Map of the location of the ASEAN member nations

that follow international standards became apparent after the Indian Ocean Tsunami in 2004. Research institutes, universities, and government laboratories of Brunei, Indonesia, Malaysia, the Philippines, Singapore, and Thailand discussed areas of cooperation such as research, proficiency testing, and accreditation to further promote forensic DNA technology in the region. In 2008, six national forensic institutes that provide DNA testing services to state police formed the Asian Forensic Science Network (AFSN). Dr. Barbara Remberg of the United Nations Office on Drug and Crimes (UNODC) and Dr. Jose Lorente of the Iberoamerican Academy of Criminalistics and Forensic Sciences (AICEF) contributed as external experts to the initiative. The six forensic science institutes were Brunei's Department of Scientific Services (DSS), Malaysia's Department of Chemistry (KIMIA), the Philippine's

National Bureau of Investigation (NBI), Singapore's Health Science Authority (HSA), Thailand's Central Institute of Forensic Science (CIFS), and Vietnam's Forensic Science Institute. In 2009, REAFD was absorbed by the DNA Work Group (DNAWG) of the Asian Forensic Science Network (AFSN). The DNAWG serves as a platform to share and promote research on forensic DNA technology in the region. DNA databasing, massively parallel sequencing (MPS), and ancestry inference are among the selected focus areas of DNAWG. Notably, research ethics to ensure the protection of human participants in research activities was not among the group's top focus areas.

Research governance must be sourced from other international organizations, given the need for ethics review in all projects involving human participants. In 2001, the WHO launched grassroots initiatives in the Asia Pacific region to address ethical gaps and challenges in global health research through the Strategic Initiative for Developing Capacity in Ethical Review (SIDCER) under the Special Program for Research and Training in Tropical Diseases (WHO-TDR). SIDCER established a network of ethical review committees, health researchers, and other partners, with a shared goal of protecting human subjects and local capacity building for ethics review (https://www.sidcer-fercap.org). Engagement with ethics committees was done through the SIDCER Recognition Program, which is essentially a voluntary audit using SIDCER criteria based on international ethical guidelines and applicable local regulations. The program is based in Thailand, and the first institutions that volunteered to join the program came from Thailand in 2005. Now across the Southeast Asian region, there is a total of 91 SIDCER recognized ethics committees collectively from Indonesia (13), Malaysia (4), Myanmar (1), Philippines (38), and Thailand (35). The SIDCER-recognized ethics committees are members of the Forum for Ethical Review Committees in Asia and the Western Pacific (FERCAP) together with other ethics committees from other Asian countries. Based on this critical mass of ethics review capacity in Southeast Asia, the region is adequately prepared to address its ethical review needs.

There has been considerable interest, particularly from surrounding countries, to study the genomic diversity of the different populations, many of which have retained their culture and language while remaining semi-isolated from the more cosmopolitan areas. Ethics policies in countries in mainland Southeast Asia (sometimes referred to as Indochina), for example, Cambodia, Laos, Myanmar, Peninsular Malaysia, Thailand, and Vietnam, as well as maritime Southeast Asia, for example, Indonesia, the Philippines, East Malaysia, Brunei, and Singapore, must include cultural sensitivity, dynamic public engagement, and science education when protocols on DNA databasing and biobanking for forensic applications are submitted for review. For scientists seeking to use their research output to establish a separate national database for law enforcement, the need to adopt an inclusive approach for all subpopulations and indigenous peoples that are citizens of their countries must be recognized. Policy and legal instruments governing forensic genetics research and biobanking are pressing concerns in a world where science needs to have crossed national borders to provide security for all.

The following section provides a brief overview of the development of the ethics review systems in each ASEAN member nation. Some countries directly benefited from the work of SIDCER-FERCAP that contributed toward the formulation of their national ethical policies for health research. A compilation of laws, regulations, and guidelines issued by the Office for Human Research Protections of the US Department of Health and Human Services (Office of Human Research Protections 2020) was reviewed together with published papers and online material to determine the most relevant policies for forensic DNA databases. The date of effectivity of the law, regulations, and guidelines were different for each country which was considered the "cut-off" year for "ethical compliance" used in Table 2. The impact of these local policies on forensic research into DNA databases is assessed using the ethical elements applied to genetic research and DNA databasing.

Brunei Darussalam

Brunei, with the official name of "Negara Brunei Darussalam" (State of Brunei Darussalam), is located on the northern coast of Borneo Island. It is surrounded by the Malaysian state of Sarawak. Brunei gained independence in 1984 from the British (Bee et al. 2021). It is inhabited by 467,000 (as of 2020), 65.7% of which are Malay, 10.3% of which are Chinese, 3.4% are indigenous (Dusun, Belait, Kedayan, Murut, and Bisaya), and 20.6% belong to other smaller groups and migrants from South Asia, Asia, and Europe (Bee et al. 2021). The population is relatively young and enjoys many benefits provided by the government, including free primary and secondary education and basic medical care. Literacy is high, and many would further pursue higher education in technical colleges or universities. The country boasts of several universities, with the University of Brunei Darussalam (UBD) being the most popular institution that offers degrees in biology and biomedicine.

Initially, Brunei did not have a national system for ethics review. In 2009, the Medical Health Research and Ethics Committee (MHREC) was established under the Ministry of Health (MOH) to review all research that requires the use of healthcare facilities (Tan et al. 2013). In collaboration with MHREC, universities also created institutional committees, known as UHREC or IHREC, to review their proposals and those initiated by foreign institutions, for example, UK-, Australian-, and New Zealand-based universities (Tan et al. 2013).

A literature survey showed that the DNA Laboratory, Department of Scientific Services (DSS) of the Ministry of Health reported a Brunei reference DNA database in 2008 (Cheong, Liew et al. 2008). The departmental director provided clearance for the study because Brunei did not have an ethics review committee at the time. The Brunei reference database was later expanded in 2018 (Liew, Riccardi et al. 2018) with DSS granted ethical approval for conducting the study in Brunei Darussalam and the University of Central Lancashire approved the work done in the United Kingdom (Table 2).

Notably, the DSS is the agency mandated to conduct DNA testing for the Royal Brunei Police Force (http://www.moh.gov.bn) and will likely take charge of the national forensic population database once support for its establishment is generated. In a 2016 global survey conducted by the International Criminal Police Organization, DSS reported that it did not have an operational DNA database for missing persons, arrestees, or convicted offenders (INTERPOL 2019).

Cambodia

With the official name of "Preahreacheanachakr Kampuchea" (Kingdom of Cambodia), Cambodia is in the southern portion of the Indochinese Peninsula, with 181,035 square kilometers and a population of 15,535,000 in 2020 (Chandler and Overton 2021). Thailand borders it to the northwest, Laos to the north, Vietnam to the east, and the Gulf of Thailand to the southwest. Over 95% of Cambodia's population is of ethnic Khmer origin because of the expansion of the Khmer Empire. Minority groups include Cham-Malays, Vietnamese, Laotians, Chinese, and indigenous peoples (Katu, Mnong, Jarai, and Rhade) of the rural highlands (Chandler and Overton 2021).

Cambodia had suffered one of the most horrific genocide in modern times (Klinkner 2008). Decades of murderous conflicts have created an imbalance in the male-female ratio of its population (Broadhurst et al. 2013). Cambodia remains one of the region's poorest countries, which continues to rely heavily on foreign aid for many of its governmental functions as it strives to build what has been lost (Chandler and Overton 2021).

In 2012, most of the country's research was externally funded without a unifying research agenda (World Health Organization Western Pacific Region 2012). Efforts to increase health research included establishing the National Ethics Committee for Health Research (NECHR) under the Ministry of Health, which reviews all research projects in Cambodia. There is no institutional REC; hence the responsibility for ethics governance is assumed by NECHR. Notably, most external funding agencies require an ethics review of a proposal (World Health Organization 2015). In conducting these reviews, the NECHR followed standard operating procedures originally written in 2002 and were revised in 2006 and 2008. These SOPs were not widely distributed, making it difficult for government agencies to fully implement these procedures (World Health Organization Western Pacific Region 2012). To date, Cambodia does not have a national law or regulation that requires researchers to have their research proposals undergo a local ethics review (World Health Organization Western Pacific Region 2012).

Moreover, the use of forensic science is limited in Cambodia. The country does not have a forensic DNA laboratory, let alone a functional offender and criminal DNA database under the Cambodian National Police (INTERPOL 2019). Samples recovered in investigations could be sent to Vietnam or Thailand. Population genetic studies on different Cambodian groups are also limited (Table 2).

However, there is interest in developing and including forensic science in its investigations, starting with identifying the casualties of the "Cambodian Genocide" (Klinkner 2008).

Indonesia

Indonesia, formally known as "Republik Indonesia," is the largest archipelago in the world, with about 17,504 islands (Legge et al. 2021). With an area of 1,916,907 square kilometers and a population of about 269,804,000 people, it is the world's fourth-most populous country and the most populous Muslim-majority country (Legge et al. 2021). Majority of its inhabitants reside in the six larger islands of Sumatra (Sumatera), Java (Jawa), Kalimantan, Celebes (Sulawesi), Lesser Sunda Islands (Nusa Tenggara), and Papua (Legge et al. 2021). Indonesia is an ethnically diverse country with around 1300 distinct ethnic groups, the largest group of which is the Javanese (40.1%), followed by the Sundanese (15.5%), then by the Malay (3.7%), Batak (3.6%), Madurese (3.0%), Betawi (2.9%), and other ethnicities (31.2%) (Legge et al. 2021). Because of its geographical location bridging the two continents, Asia and Australia, and its diverse population composed of numerous ethnic groups speaking 724 languages and dialects, researchers, local and international, have included Indonesian participants in their genetic studies (Table 2).

According to the Government Regulation of the Republic of Indonesia Number 41 of 2006, foreign researchers must get a government permit on the "Permit to conduct research and development for foreign universities, research and development institutes, companies and individuals." This law superseded Presidential Decree No. 100/1993 entitled Research by Foreigners listed in the 2020 International Compilation of Human Research Standards (Office of Human Research Protections 2020). In the same list, the Indonesian National Health Regulation No. 36 of 2009 also replaced Regulation No. 39/1995 on Health Research Development which governs public health research. Hence, we considered 2006 the cut-off year for implementing ethics clearance in Indonesia (Table 2).

In 2019, another law was passed establishing the Badan Riset dan Inovasi Nasional (BRIN) or the National Research and Innovation Agency, which was mandated to coordinate most of the research conducted in Indonesian government research centers. This law required foreign researchers to collaborate with local scientists, limit the exportation of Indonesian biological samples, and include sanctions on violators (Rochmyaningsih 2019). The law was part of a national response to two research studies considered "acts of biopiracy."

Notably, an institutional permit is not the same as an ethics clearance, although institutional review already includes an ethics component in some areas. The implementation of research ethics review requirements in Indonesia were placed under the Minister of Health, that the National Ethics Commission assists following the Regulation of the Ministry of Health (PMK) number 7 of 2016 (Fourianalistyawati et al. 2018). Researchers commonly apply for review to university-based institutional review committees. In 2018, 52.5% of Indonesian universities

were reported to have operational IRCs that also review the ethics component (Fourianalistyawati et al. 2018). In some areas, the absence of review committees is partly addressed by ethics committees in selected polytechnical colleges, research institutions, and teaching hospitals. A list of Indonesian ethics review committees that SIDCER-FERCAP recognizes for the quality of their ethical review can be found in https://www.sidcer-fercap.org.

Scientific papers reporting the generation of genotypes using DNA markers used for forensic applications were surveyed to determine compliance with institutional and ethical clearances required by Indonesian laws. Early databasing papers did not undergo ethics review in Indonesia because the ethics governance system was not established at that time (Table 2). Several scientific papers with foreign coauthors acknowledged the DNA Forensics Unit, Eijkman Institute for Molecular Biology Research in Jakarta. The laboratory conducts primary research in human identification across Indonesian populations, including indigenous cultural communities. Particularly for multicenter studies, Eijkman researchers provide the local perspective on Indonesian culture. Cultural sensitivity is an important factor in obtaining the free and prior informed consent of persons, and communities, whenever applicable. Hence the involvement of local researchers bridges the communication between the local population and the foreign researchers.

The Indonesian National Police has the Forensic DNA Laboratory of the Police Medicine Centre located in Jakarta that conducts DNA testing for forensic investigations (https://www.polri.go.id/) and has published its studies. Notably, Indonesia has been listed as one of the countries with a national database (http://dnapolicyinitiative.org/), although Interpol has not shared information on this database in its survey of 2019 (INTERPOL 2019).

Laos

Laos, or the Lao People's Democratic Republic, formally known as "Sathalanalat Paxathipatai Paxaxon Lao," is the only landlocked mountainous country in Southeast Asia (Lafont et al. 2021). Myanmar and Thailand surround it in the West, Cambodia in the South, Vietnam in the East, and China in the North. Its estimated population in 2020 was 7,226,000 that belong to more than 60 different ethnically and linguistically diverse groups. The Lao population had been conveniently grouped into three categories – Lao Loum (Lowland Lao), Lao Theung (Lao of the Mountain Slopes), and Lao Sung (Lao of the Mountain Tops) to simplify government administrative functions (Lafont et al. 2021).

There has been considerable interest, particularly from surrounding countries, to study the genomic diversity of different populations in Laos. However, there is little information about the ethical review of research involving humans before 2000 (World Health Organization Western Pacific Region 2012). In 1999, the National Institute of Public Health, or NIOPH (http://www.nioph.gov.la/), was established to manage the country's health research and training agenda (NIOPH-COHRED 2006). During this period, the Council on Health Research for Development (COHRED), a

global nonprofit organization (https://www.cohred.org), provided support to Laos and NIOPH in formulating a national health research master plan for the management of its national health research system (https://healthresearchweb.org/en/laos). This master plan which includes the ethical review process for health research was reviewed every five years since its initial formulation in 1992 (NIOPH-COHRED 2006). However, in 2010 report, Laos was no longer listed in the ten countries supported by COHRED, which may indicate either a simple discontinuation of available support for the country at that time, or that the country has reached its target outcomes and may not need further support (Council for Health Research for Development 2010).

In a 2012 report, Dr. Kongsap Akkhavong, then Acting Director-General of the National Institute of Public Health, Ministry of Health, confirmed that the challenges of maintaining an active ethical review program in Laos remains because of limited human resources and the lack of external support structure for the operations of the review process (World Health Organization Western Pacific Region 2012). At that time, Laos had two research ethics committees under the University of Health Sciences to cover university research and the National Ethics Committee for Health Research (NECHR) established in 2002 under NIOPH for all other health research in Laos (laohrp.com/index.php/hrp/index). Another major challenge in maintaining a fully operational ethics review system was the severe shortage of health professionals in Laos (Qian et al. 2016) that affected the management of members for these two committees.

Forensic genetics research had been conducted by foreign scientists interested in the great ethnic diversity of Laos populations. However, after 2012, studies reported ethics clearances obtained from foreign countries. As of 2020, Laos was not in the internal compilation on human standards, and updated information on the country's ethics review of proposals is not available (Office of Human Research Protections 2020). The Laos Police Force does not have a DNA forensic laboratory to handle any biological material, nor does it use any DNA database for its investigations (INTERPOL 2019).

Malaysia

Malaysia is composed of Peninsular Malaysia and East Malaysia on the island of Borneo (Ahmad et al. 2021). With an estimated population of 32,779,000 in 2021, it is the world's 43rd most populous country. About half of the population is ethnically Malay, with minorities of indigenous peoples, for example, Orang Asli, Semang and Senoi, Chinese, and Indians (Ahmad et al. 2021). The country's multiethnic and multicultural composition results in a diverse understanding of science across different groups, making some groups more vulnerable to the risks of participating than others. Hence, Malaysia turned to ethics review of all research as a means of protecting its human participants.

Ethics review in Malaysia started with the "Malaysian Guidelines for Good Clinical Practice (GCP)" in October 1999. This document was amended in 2004,

2011, and 2018 by the National Committee for Clinical Research (http://www.nccr.gov.my/). Although primarily focused on clinical research, NCCR also released the "Guideline for the Ethical Review of Clinical Research or Research Involving Human Subjects" in 2006, thus expanding the scope of the ethics review process to cover the collection of biological samples for research and bio-banking (National Committee for Clinical Research 2006; Office of Human Research Protections 2020). This development led to the modification of research guidelines in many institutions to require ethical clearance when human participants are involved. Hence for this review, 2006 was the critical year in Malaysia when guidelines for ethics review should have been followed (Table 2).

Research in Malaysia can be reviewed by the Medical Research Ethics Committee (MREC) of the National Institute of Health (https://www.nih.gov.my/mrec/) or corresponding committees in universities (World Health Organization Western Pacific Region 2012). These review committees include:

- Human Research Ethics Committee of the University Sains Malaysia (JEPeM) (http://www.jepem.kk.usm.my/)
- University of Malaya Research Ethics Committee (UMREC) (https://umresearch.um.edu.my/university-of-malaya-research-ethics-committee-umrec)
- Ethics Committee for Research Involving Human Subjects Universiti Putra Malaysia (Jawatankuasa Etika Universiti Penyelidikan Melibatkan Manusia or JKEUPM) (https://tncpi.upm.edu.my/research/research_ethics_evaluation/jkeupm_ethic_committee_for_research_involving_human_subject-39931)
- Research Ethics Committee of the Universiti Teknologi MARA (https://www.recuitm.org/) and
- Institutional Review Committee of the Universiti Kebangsaan Malaysia (https://www.ukm.my/portal/#).

The high quality of the ethics review process provided by MREC of MOH, and the three university RECS, for example, JEPeM, UMREC, and JKEUPM, was recognized by SIDCER-FERCAP (https://www.sidcerfercap.org). In compliance with national guidelines, researchers in forensic genetics reported their ethics clearance from these university RECs that have established ethics review procedures. Research that involves indigenous communities requires additional permissions from relevant institutions like the Jabatan Hal Ehwal Orang Asli, Malaysia (JHEOA) of the Department of Orang Asli Affairs (Table 2).

The testing of forensically relevant DNA markers and creation of reference DNA databases of different Malaysian groups was reported as early as 2001 by scientists of the Forensic Science Division of the Forensic Science Analysis Centre, Kimia Malaysia (https://www.kimia.gov.my). The preliminary research on reference DNA databases made by the Forensic Science Division of Kimia Malaysia became the basis for the Forensic DNA Databank of Malaysia (FDDM). FDDM was officially established under the DNA Databank Division of the Royal Malaysia Police in 2015 after enacting the Malaysian DNA Identification Act of 2009. By 2021, the forensic DNA database was reported to have reached 100,000 genotypes generated from

DNA of convicted offenders, detainees, suspected persons, drug dependents, crime scenes, missing persons, and volunteers who willingly provided their samples for forensic investigations (Rahman et al. 2021).

Myanmar

Myanmar, with the official name being "Pyihtaungsu Thamada Myanmar Naingngandaw" or the Republic of the Union of Myanmar, is bordered by Bangladesh and India to its northwest, China to its northeast, Laos and Thailand to its east and southeast, and the Andaman Sea and the Bay of Bengal to its south and southwest (Aung et al. 2021). It is the largest country in mainland Southeast Asia and the 10th largest in Asia by area, with about 54 million persons as of 2017. The country is ethnically diverse, with the government recognizing 135 distinct ethnic groups and the Burman forming an estimated 55.9% of the population. In contrast, 9.5% of the population is Karen, 6.5% is Shan, 2.3% is Mon, ethnolinguistically related to the Khmer, 2.2% Yangbye, 1.5% Kachin, and 2.5% Chinese. The remaining 19.6% includes all other persons in Myanmar, for example, Chin, Rohingya, Anglo-Indians, Gurkha, Nepali, and Anglo-Burmese (Aung et al. 2021). Myanmar is home to four major language families: Sino-Tibetan, Tai-Kadai, Austro-Asiatic, and Indo-European.

Briefly, research ethics committees (RECs) in Myanmar were established as early as 1980 (Oo et al. 2018). The first REC in the Department of Medical Research of the Ministry of Health and Sports (DMR, MoHS) is the most recognized. Its members contributed significantly to the formulation of the Operational Guidelines for Institutional Ethical Review Committee released in 2005. SIDCER-FERCAP also recognized the DMR ERC for the quality of ethics review it provides (https://www.sidcer-fercap.org). By 2018, there were 20 established RECs composed of 16 RECs associated with public universities, the DMR, and three at military institutions.

The US Office of Human Research Protection listed a 2016 Guideline for Submission to Ethics Review Committee and identified the Department of Medical Research and the Ministry of Health National Ethics Committee on Clinical Research as key organizations (Office of Human Research Protections 2020). However, the literature review on research in forensic DNA in Myanmar did not produce a single publication on work done within Myanmar. The paper on the mitochondrial DNA of 327 Burmese originally from Myanmar (Summerer et al. 2014) was not included because the procedure for sample collection in Thailand where these persons resided was not the same as in Myanmar (Table 2).

Myanmar came out of absolute military rule in 2011, only to return to it once again in February 2021. Progress in healthcare and scientific research, which saw a rapid increase during the interim years from 2011 to 2021, was again put on hold. The military declared a one-year state of emergency in the increasing civilian unrest amidst a global pandemic. Searches on the availability of forensic DNA laboratories in any university and the Myanmar Police Force did not produce any results. The

country's response to the 2019 Interpol Global Survey indicated that Myanmar did not have a forensic DNA database for forensic investigations (INTERPOL 2019). Myanmar was also not listed on the website of http://dnapolicyinitiative.org.

Philippines

The Philippines officially known as "Republika ng Pilipinas," is an archipelagic country consisting of about 7640 islands covering an area of 300,000 square kilometers and a population of around 111,109,000 people, making it the world's twelfth most populous country (Borlaza et al. 2021). Citing the 2010 census, used by Borlaza et al. (2021), 24.4% are Tagalog, 11.4% are Bisaya, 9.9% are Cebuano, 8.8% are Ilocano, 8.5% are Hiligaynon, 6.8% are Bikolano, 4% are Waray, and 26.2% are others. These others include the 110 groups of indigenous peoples, for example, Ibaloi, Kankana-ey, Ifugao, Agta, Ayta, Ati, Mamanwa, Aeta, Iraya, Hanunuo, Gubatnon, Sama, Yakan, that are recognized by the Philippine National Commission on Indigenous Peoples (https://ncip.gov.ph).

In 2013, the Philippines promulgated Republic Act 10532, entitled the "Philippine National Health Research System Act" (otherwise known as the PNHRS Law) (Office of Human Research Protections 2020). This law is to be jointly implemented by the Department of Science and Technology (DOST), the Department of Health (DOH), the Commission on Higher Education (CHED), and the University of the Philippines Manila. The law institutionalized the Philippine Health Research Ethics Board (PHREB) as the national policymaking body on health research ethics and mandated it to accredit and manage the registration of all research ethics review committees. The law categorically stated that "All research involving human subjects must undergo ethical review and clearance before implementation to ensure the safety, dignity, and well-being of research participants." Hence, 2013 is the year that made ethics review a requirement for all health and health-related research involving human participants (Table 2). Before promulgating this law, the "National Ethical Guidelines for Health Research" first released in 2006 and amended in 2013 were used for elective review of biomedical proposals (www.ethics.healthresearch.ph). In 2017, PHREB released a revised version of the guidelines, which included more sections on human research outside of the clinical/biomedical field (Office of Human Research Protections 2020).

Notably, in the Philippines, the scope of a project proposal, the extent of human participation, and the possible risks of participation are used to determine the level of review, for example, Tier 1, 2, or 3. In addition, the ethics process includes a procedure for the approval of requests for exemption from review or expedited review, depending on the research. For a forensic databasing project that includes sample biobanking and generation of genetic information about individuals and communities, Philippine procedures require the review to be conducted by at least a level 2 ERC (PHREB 2017). Additional requirements apply following the issuances from the NCIP when the research includes members of vulnerable populations, such as indigenous cultural communities/indigenous peoples (De Ungria and Jimenez 2021).

Researchers from the DNA Analysis Laboratory of the Natural Sciences Research Institute, University of the Philippines Diliman made early research efforts on databasing (Table 2), teaching, and public engagements to promote forensic DNA technology. The laboratory also worked with law enforcement agencies, for example, the National Bureau of Investigation and the Philippine National Police. It continues to advocate for establishing a national DNA database for law enforcement, provided there are sufficient safeguards and mechanisms to prevent unauthorized access to biological samples and unreasonable searches of DNA profiles (De Ungria and Jose 2010). In 2007, the Philippine Supreme Court promulgated the "Rule on DNA Evidence," which included directives for collecting and storing convicted offender samples (Philippines Supreme Court 2007). However, a broader regulatory tool in the form of legislation is needed to maximize the potentials of forensic DNA testing and databasing in the Philippines (De Ungria and Jose 2018). To date, several bills on DNA databases have been proposed in Congress but have not yet become law (http://dnapolicyinitiative.org).

Singapore

Singapore, formally known as the "Republic of Singapore," is a sovereign island city-state in maritime Southeast Asia, with a population of 5,709,000, occupying 728.6 square kilometers. With a multicultural population consisting of Chinese (74.7%), Malay (13.4%), Indian (9.0%), and others (3.2%) (Ho et al. 2021). Singapore has four official languages: English, Malay, Mandarin, and Tamil.

Bioethics in Singapore is well-structured and classified into three main systems: clinical ethics, research ethics, and biomedical ethics (Ho et al. 2014). The key organizations/institutions that determine the development of these systems are the Ministry of Health (MOH), the Bioethics Advisory Committee (BAC), and the Centre for Biomedical Ethics (CBmE), at the National University of Singapore). Biomedical and clinical research dominate the research environment because of the significant government investment in the development of drugs and medical devices and clinical trials. In 2004, the BAC came up with a report entitled "Research involving Human Subjects: Guidelines for IRBs [Institutional Review Boards]" (Bioethics Advisory Committee 2004). This report focused on the increasing public interest in human subjects and was a product of consultations with the Human Genetics Subcommittee (HGS) (Ho et al. 2014). The committee recommended reviewing all human biomedical research for approval by a recognized ethics review committee and requiring the accreditation of all ethics review committees (Bioethics Advisory Committee 2004). By 2015, BAC issued the Ethics Guidelines for Human Biomedical Research for all ERCs in Singapore (Bioethics Advisory Committee 2015; Office of Human Research Protections 2020). We identified 2015 as the year when ethics review in Singapore diverted its attention from strictly biomedical research to human health research (Table 2).

Research papers that reported the creation of reference population databases using forensically relevant DNA markers in Singapore were published from the

DNA Profiling Laboratory, Center for Forensic Science of the Health Science Authority. This laboratory was established in 1990 and provided DNA testing services for the Singapore Police Force (has.gov.sg/about-us/applied-sciences/forensic-biology). A second laboratory, the DNA Database Laboratory, was established with the Criminal Investigation Department of the Singapore Police Force in 2003. The second laboratory tasked to maintain the arrestee, convicted persons, and prisoner DNA database was established following Chapter 268 of the Registration of Criminal Acts of Singapore. In the 2019 Interpol survey, Singapore reported having a DNA database with 909,745 profiles (INTERPOL 2019).

Thailand

Thailand, known officially as "Ratcha Anachak Thai" or the Kingdom of Thailand, is located at the center of the Indochinese Peninsula with an estimated population of 66,664,000 (Hafner et al. 2021). Thai nationals make up 97.5% of the population. The remaining 2.5% of the population are Burmese (1.3%) and others (1.2%). Sixty-two ethnic communities are officially recognized in Thailand. The main groups arranged by the highest to lowest numbers are as follows: 1) Thai Lao and other smaller Lao groups, for example, the Thai Loei, Lao Lom, Lao Wiang/Klang, Lao Khrang, Lao Ngaew, and Lao Ti; 2) Khon Muang referred to as the Northern Thais; 3) Pak Tai or Southern Thais; 4) the Khmer Leu also called Northern Khmer; 5) and all other groups of Malay, Nyaw, Phu Thai, Suay, and Karen (Hafner et al. 2021).

The ethics review of research with human participants in Thailand started in 1975 (Panichkul et al. 2011). The first REC was the Human Experimentation Committee of the Medical Department, Chiang Mai University (HEC). At that time, the National Research Council of Thailand (NRCT) released the "Guidelines for Biomedical Research involving Human Subjects" distributed to different researchers, research hospitals, and universities. In 2007, the Forum for Ethical Review Committees in Thailand (FERCIT) released a revised set of guidelines entitled "The Ethical Guidelines for Research on Human Subject in Thailand" (Forum for Ethical Review Committees in Thailand 2007). The document set out the ethics for research on human subjects that covered studies that were not part of the traditional clinical trials and drug discoveries. In the same year, the National Research Council of Thailand (NRCT) issued the "Regulations on the Permission for Foreign Researchers to Conduct Research in Thailand B.E. 2550" was issued (National Research Council of Thailand 2007; Office of Human Research Protections 2020). Finally, in 2015, the NRCT recognized the need to elevate the guidelines to the level of national policies (National Research Council of Thailand 2015) that all ethics review committees must follow in research institutes, hospitals, and universities. In 2020, SIDCER-FERCAP recognized 30 ethics review committees in Thailand, following the international guidelines provided for by the WHO (www.sidcer-fercap.org). The year 2007 was considered the critical year in this study that marked the requirement for ethics compliance in Thailand (Table 2).

The issuance of the 2007 guidelines by FERCIT created greater awareness for the need to obtain ethics clearance for all research that involves human participants. Papers included information on the ethics clearance issued by the Faculty of Medicine of Mahidol University, Ethics Board of Prince of Songkla University, Institutional Review Board of the Faculty of Medicine of Chulalongkorn University, Human Experimentation Committee, Research Institute for Health Sciences in Chiang Mai University, Khon Kaen University Ethics Committee in Human, and the Naresuan University Institutional Review Board (Table 2).

The Central Institute of Forensic Science, Ministry of Justice (CIFS) provides the Royal Thai Police DNA testing services. It is mandated to establish the national DNA database (Boonderm et al. 2017). Creating the Thai reference database using volunteer samples was likely part of the in-house validation needed for the much larger work required to establish, monitor, and expand the national DNA database. As of 2019, the national DNA database has reached over 187,464 profiles from crime scenes, persons of interest, and prisoners (Boonderm et al. 2019).

Vietnam

Vietnam is officially known as "Cong Hoa Xa Hoi Chu Nghia Viet Nam" or the Socialist Republic of Vietnam. It is located at the eastern edge of the Indochinese Peninsula, with an estimated 97,591,000 inhabitants making it the world's fifteenth-most populous country (Buttinger et al. 2021). The dominant Viet or Kinh ethnic group constitutes 85.7% of the population. Fifty-four other ethnic groups are officially recognized, including the Hmong (Miao), Tho (Tay), Thai, Mong, Khmer, Nung, and the Hoa. The country's national language is Vietnamese, a tonal Austroasiatic language is spoken by most of the population. Vietnam's minority groups speak various languages to communicate among themselves and across different groups (Buttinger et al. 2021).

Vietnam has a four-tier hierarchical health system consisting of the national, provincial, district, and commune (World Health Organization Western Pacific Region 2012). Research institutes, medical schools, and the health service provision sector are included at the national level; hence an ethical review of a research project is national in scope.

Implementation of the ethical requirement for research involving human subjects remains challenging because of the country's weak health system (https://www.cohred.org). In the 2006 assessment, ethical concerns for human participants were considered less important by researchers in Vietnam. The main concerns were the low allocation of the national budget for health, the lack of support for researchers, and the need for capacity-building to enable researchers to turn research demands into feasible research proposals. In addition, because of the limited number of experts, the RECs, including those of the Ministry of Health, may conduct both technical and ethical reviews rather than separately (World Health Organization Western Pacific Region 2012).

In 2013, the "Regulation on Organization and Operation of Council of Ethics in Biomedical Research at grass-root level" under the Ministry of Health was promulgated (Ministry of Health 2013) (Office of Human Research Protections 2020). This law requires that all research with human participants in Vietnam undergo an ethics review. This promulgation marked 2013 as the critical year for ethics compliance in Vietnam (Table 2). After passing this law, the Ethical Committee of the Vietnam Academy of Science and Technology had been reviewing forensic DNA databasing protocols.

The Vietnam government has also allocated considerable resources to the Vietnam People's Police Force (PPF) of the Ministry of Public Security (http://en.bocongan.gov.vn/) for the creation of the national DNA database used by law enforcement (INTERPOL 2019).

Survey of Publications on Forensic DNA databases in SE Asia

The ethical requirements compiled from six forensic journals commonly publishing forensic DNA database papers were reframed to appraise 91 scientific articles for their compliance. These six journals include: 1) *Forensic Science International* (FSI); 2) *FSI: Genetics*; 3) *FSI Genetics Supplement Series*; 4) *Journal of Forensic Sciences*; 5) *International Journal of Legal Medicine*; and 6) *Legal Medicine*. All other submissions were classified as local for those published in the country where sampling took place. Forty-six percent of the articles did not have the mandatory ethical review statement before the country's designated cut-off year. This number was substantially reduced to 11% after the country-specific cut-off year indicating increased institutional compliance with international ethical requirements.

Similarly, local review of 8% before the cut-off year reached 22% post-cut-off year showing improvements in the national capacity for review and oversight. The publication of forensic DNA databases in local scientific journals (5%) may indicate researchers' choice to engage the people of their own countries as a way of giving back. These trends are encouraging and show evidence of harmonization with ethical publication standards. However, two main areas for improvement include issuing statements on the informed consent provided by participants and the researcher's declaration of conflict of interest, if any. The majority of the papers had no information on the informed consent of research participants (56%) and the conflict of interest of researchers, if any (68%), despite the long-standing mandates for these requirements.

Conclusions and Recommendations

With the rapid advancement in forensic DNA technology that allows researchers to generate human genetic data at a much larger scale and a much faster speed, there is an urgent need to re-evaluate the ethical governance of forensic genetics research.

This chapter revisited the principal roles of stakeholders, for example, researcher, research institution, funding agency, research ethics committee, and scientific journal, to ensure the protection of all human subjects, especially those coming from the more vulnerable populations. This discussion facilitated reflections into the sufficiency of the current ethical requirements in the face of these newer technologies.

The chapter revisited the tools available to researchers, including the requirements set by six journals that commonly publish forensic DNA databases. The authors recommend adopting the 2016 WMA Declaration of Taipei on research on databases, big data, and biobanks and the 2016 CIOMS guidelines while retaining the 2013 WMA Declaration of Helsinki requirement by all journals that publish any research on forensic DNA databases.

The five ethical elements, namely, informed consent, data protection, independence of research ethics committees, funding and benefit-sharing, and public engagement, were used to review the status of database research across the ten country members of the ASEAN. The status of research was ascertained through a survey of 91 scientific publications published from 1998 to the present. The data mined highlighted the importance of strengthening the ethical governance for research on forensic DNA databases to ensure human subject protection. This review is particularly important in the countries included here due to the extensive cultural, linguistic, and genetic diversity and the varying vulnerability of many groups and subpopulations in the region.

Strong ethical governance can be a cumulative outcome of all stakeholders fulfilling their individually designated responsibilities, underscored by a common set of agreed ethical elements, toward protecting human participants. Such a strong governance system, which relies on self-regulation and national policies, is essential to promoting research integrity. Country-level discussions should be initiated on the imposition of sanctions on researchers and institutions that are found to have willfully violated ethics guidelines, whether or not actual harm is caused to vulnerable populations subject of the research.

Hopefully, with continuing advocacy, research stakeholders will eventually realize that presumptions about ethical elements pervasive in Western developed countries – where a simple yes may signify consent – are not immediately applicable to all populations. Therefore, greater responsibility for public engagement (and empowerment) must be pursued to ensure that all efforts are exerted not only to "do no harm" but to actively contribute to making science truly at the service of the people.

References

Ahmad ZB, Lockard CA, Leinbach TR, Bee OJ (2021) Malaysia, Encyclopaedia Britannica

Aung MH, Steinberg DI, Aung-Thwin MA (2021) Myanmar, Encyclopaedia Britannica

Austin MA, Harding SE, McElroy CE (2003) Monitoring ethical, legal, and social issues in developing population genetic databases. Genet Med 5:451–457

Bee OJ, Damit MY, Thambipillai P (2021) Brunei, Encyclopaedia Britannica

Bioethics Advisory Committee (2004) In: Pin L (ed) Research involving human subjects: guidelines for IRBs. The Singapore Bioethics Advisory Committee, Singapore, p 165

Bioethics Advisory Committee (2015) Ethics guidelines for human biomedical research. Bioethics Advisory Committee, Singapore

Boonderm N, Suriyanratakorn D, Wongvoravivat C, Sangpueng S, Nettakul A, Waiyawuth W (2017) Effectiveness of CIFS DNA database in Thailand. Forensic Sci Int Genetics Suppl Ser 6: e585–e586. https://doi.org/10.1016/j.fsigss.2017.09.220

Boonderm N, Suriyanratakorn D, Sangpueng S, Wongvoravivat C, Waiyawuth W (2019) Utilization of the CIFS DNA database to monitor recidivism. Forensic Sci Int Genetics Suppl Ser 7: 685–687. https://doi.org/10.1016/j.fsigss.2019.10.138

Borlaza GC, Hernandez CG, Culliname M (2021) Philippines, Encyclopaedia Britannica

Broadhurst R, Bouhours T, Keo C (2013) Crime and justice in Cambodia. In: Liu J, Hebenton B, Jou S (eds) Handbook of Asian Criminology. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-5218-8_11

Buttinger J, Hickey GC, Turley WS, Jamieson NL, Osborne ME, Duiker WJ (2021) Vietnam, Encyclopaedia Britannica

Cargo M, Mercer SL (2008) The value and challenges of participatory research: strengthening its practice. Annu Rev Public Health 29:325–350. https://doi.org/10.1146/annurev.publhealth.29.091307.083824

Chandler DP, Overton LC (2021) Cambodia, Encyclopaedia Britannica

Cheong PY, Liew PVO, Ibrahim H, Budowle B (2008) A genetic database for DNA based forensic analysis in Brunei Darussalam. Brunei Darussalam J Health 3:1–6

Council for Health Research for Development (2010) COHRED annual report 2010: supporting the national health research systems in low- and middle-income countries. Council for Health Research for Development

Council for International Organizations of Medical Sciences (2016) International ethical guidelines for health-related research involving humans. Council for International Organizations of Medical Sciences and the World Health Organization, p 119

D'Amato ME, Bodner M, Butler JM, Gusmão L, Linacre A, Parson W, Schneider PM, Vallone P, Carracedo A (2020) Ethical publication of research on genetics and genomics of biological material: Guidelines and recommendations. Forensic Sci Int Genet 48, 102299. https://doi.org/10.1016/j.fsigen.2020.102299

De Ungria MCA, Jimenez EB (2021) Ethical challenges in genetics research with indigenous cultural communities/indigenous peoples in the Philippines. Philippine Genome Center, Quezon City, p 1. https://pgc.up.edu.ph/research-development/biodiversity-and-ethnicity/icc-ip-ethical-challenges-in-genetics-research/

De Ungria MC, Jose JM (2010) Forensic DNA profiling and databasing: the Philippine experience. In: Hindmarsh R, Prainsack B (eds) Genetic suspects: global governance of forensic DNA profiling and databasing. Cambridge University Press, Cambridge

De Ungria MCA, Jose JM (2018) Forensic science in challenging environments: the Philippine experience. In: Caguioa IAC et al (eds) Coalition Against the Death Penalty, Manila, pp 6–17

Forum for Ethical Review Committees in Thailand (2007) The ethical guidelines for research in human subject in Thailand 2007. In: Sueblinvong T et al (eds) p 62

Fourianalistyawati E, Uswatunnisa A, Mahdiannur MA, Saleky AP, Soebandhi S, Reni A, Kurniasih N (2018) Research development in Indonesia: Ethics committee in open science and collaboration era. J Phys Conf Ser 1114:012069. https://doi.org/10.1088/1742-6596/1114/1/012069

Gibbons SMC, Kaye J, Smart A, Heeney C, Parker M (2007) Governing genetic databases: challenges facing research regulation and practice. J Law Soc 34:163

Greytak EM, Moore C, Armentrout EM (2019) Genetic genealogy for cold case and active investigations. Forensic Sci Int 299:103–113

Hafner JA, Keyes EJ, Keyes CF (2021) Thailand, Encyclopaedia Britannica

Hayry M, Chadwick R, Arnason V, Arnason G (2007) The ethics and governance of human genetic databases: European perspectives. Cambridge University Press, New York

Ho CW-L, Chin JJL, Campbell A (2014) Singapore. In: Henk AMJ t H, Gordjin B (eds) Handbook of global bioethics. Springer Link, Dordrecht

Ho R, Winstedt RO, Kennard A, Leinbach TR (2021) Singapore, Encyclopaedia Britannica

INTERPOL (2019) Global DNA profiling survey results, p 20
Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable minisatellite regions in human DNA. Nature 314:67–73
Klinkner M (2008) Forensic science for Cambodian justice. Int J Transit Justice 2:227–243. https://doi.org/10.1093/ijtj/ijn015
Lafont P-B, Zasloff JJ, Silverstein J, Osborne ME, Dommen AJ (2021) Laos, Encyclopaedia Britannica
Legge JD, Leinbach TR, Adam AW, Wolters OW, Mohamad GS, McDivitt JF (2021) Indonesia, Encyclopaedia Britannica
Liew PVO, Riccardi LN, Afolabi AO, Goodwin W, Hadi S (2018) Optimization of a reduced volume PCR amplification for PowerPlex® Fusion kit using FTA™ cards and generation of population genetic data for Brunei population. Electrophoresis 39:2979–2990
Lowrance WW (2001) The promise of human genetic databases. BMJ 322:1009–1010. https://doi.org/10.1136/bmj.322.7293.1009
Ministry of Health (2013) On promulgation of Regulation on Organization and Operation of Council of Ethics in Biomedical Research at the Grass-Root Level. In: The Socialist Republic of Vietnam Ministry of Health (ed) No. 111/QD-BYT, Hanoi
Moon C-I (2020) ASEAN, Encyclopaedia Britannica
National Committee for Clinical Research (2006) Guideline on the use of human biological tissues for research. National Committee for Clinical Research, Malaysia
National Research Council (1996) The evaluation of forensic DNA evidence. National Academies Press, Washington, DC
National Research Council of Thailand (2007) In: Sutabutra T (ed) Regulations on the permission for foreign researchers to conduct research in Thailand. Office of International Affairs
National Research Council of Thailand (2015) National policy and guidelines for human research. National Research Council of Thailand, p 113
NIOPH-COHRED (2006) National Health Research System (NHRS) mapping in Laos, vol 8. National Institute of Public Health – Council on Health Research for Development
Office of Human Research Protections (2020) International compilation of human standards. US Department of Health and Human Services, pp 1–198
Oo Z-Z, Oo Y-T-N, Than M-M, Oo KZ, Wun M, Htun K-S, Silverman H (2018) Current status of research ethics capacity in Myanmar. Asian Bioeth Rev 10:123–132
Panichkul S, Mahaisavariya P, Morakote N, Condo S, Caengow S, Ketunpanya A (2011) Current status of the research ethics committees in Thailand. J Med Assoc Thailand 94:1013–1018
Philippine Supreme Court (2007) Rule on DNA Evidence. Manila, Philippines A.M. No. 06-11-5-SC
Phillips M, Molnar-Gabor F, Korbel JO, Thorogood A, Joly Y, Chalmers D, Townend D, Knoppers BM (2020) Genomics: data sharing needs an international code of conduct. Nature 578:31–33. https://doi.org/10.1038/d41586-020-00082-9
PHREB (2017) National Ethical Guidelines for Health and Health-related Research. Philippine Council for health and research development. PHREB, Manila
Qian Y, Yan F, Wang W, Clancy S, Akkhavong K, Vonglokham M, Outhensackda S, Østbye T (2016) Challenges for strengthening the health workforce in the Lao People's Democratic Republic: Perspectives from key stakeholders. Hum Resour Health 14. https://doi.org/10.1186/s12960-016-0167-y
Rahman MA, Ismail SA, Ngoh TN, Khan HO, Saedon NA, Lazim NHM (2021) Forensic DNA databank Malaysia (FDDM): 100,000 DNA profiles achievement. Sci Technol Public Policy 5:47–53
Rochmyaningsih D (2019) Indonesia gets tough on foreign scientists. Science 365:304
Schiermeier Q (2021) Crime-solving DNA databases faces ethical scrutiny. Nature 594:320–322
Seeman N (2019) Embrace data anonymity, not 'digital consent'. Nature 573:34–34. https://doi.org/10.1038/d41586-019-02627-z
Summerer M, Horst J, Erhart G, Weißensteiner H, Schönherr S, Pacher D, Forer L, Horst D, Manhart A, Horst B, Sanguansermsri T, Kloss-Brandstätter A (2014) Large-scale mitochondrial DNA analysis in Southeast Asia reveals evolutionary effects of cultural isolation in the multi-

ethnic population of Myanmar. BMC Evol Biol 14:17. https://doi.org/10.1186/1471-2148-14-17
Tan J, Tea AT, Pengiran Tengah DSNA, Yong AML (2013) Role of the Medical Health Research Ethics Committee (MHREC) in Brunei Darussalam. Brunei Int Med J 9:61–70
Tindana P, Depuur C, De Vries J, Seeley J, Parker M (2020) Informed consent in genomic research and biobanking: Taking feedback of findings seriously. Glob Bioeth 31:200–215. https://doi.org/10.1080/11287462.2020.1717896
United Nations Educational Scientific and Cultural Organization. (2003) International declaration on human genetic data, United Nations Educational Scientific and Cultural Organization (Ed.), United Nations Educational Scientific and Cultural Organization Geneva
Wallerstein NB, Duran B (2006) Using community-based participatory research to address health disparities. Health Promot Pract 7:312–323. https://doi.org/10.1177/1524839906289376
Wickenheiser RA (2019) A crosswalk from medical bioethics to forensic bioethics. Forensic Sci Int Synergy 1:35–44. https://doi.org/10.1016/j.fsisyn.2019.03.002
Williams GA, Liede S, Fahy N, Aittomaki K, Perola M, Helander T, McKee M, Sagan A (2021) In: Sagan A (ed) Regulating the unknown: A guide to regulating genomics for health policy-makers. World Health Organization, Denmark
World Health Organization (1946) In: World Health Organization (ed) Constitution of the World Health Organization. World Health Organization, New York
World Health Organization (2011) Standards and operational guidance for ethics review of health-related research with human participants. World Health Organization, p 56
World Health Organization (2015) List of National Ethics Committees
World Health Organization Western Pacific Region (2012) Expert consultation on optimization of health research ethics governance systems in the Western Pacific Region. World Health Organization Western Pacific Region, Manila, pp 1–69
World Medical Association (2013) Declaration of Helsinki- ethical principles for medical research involving human subjects. p 4
World Medical Association (2016) Declaration of Taipei on ethical considerations regarding health databases and biobanks. p 3

Using Laboratory Validation to Identify and Establish Limits to the Reliability of Probabilistic Genotyping Systems

12

Dan E. Krane and M. Katherine Philpott

Contents

Abstract

Probabilistic genotyping systems endeavor to aid human analysts in the interpretation of complex DNA mixtures. Laboratories use these complicated programs to attach statistical weights to results from evidence samples that cannot be confirmed by analysts using manual/conventional approaches. Before relying on these systems in case work, it is essential that the software be subjected to rigorous internal validation that establishes the limits of their reliability. Complex mixtures encountered in casework often have attributes that make interpretation challenging, including low template levels, unknown/high number of contributors, high levels of allele sharing, differential degradation, and so forth. Each laboratory's internal validation should critically evaluate its performance using

D. E. Krane (✉)
Biological Sciences, Wright State University, Dayton, OH, USA
e-mail: dan.krane@wright.edu

M. K. Philpott
Dept. of Forensic Science, Virginia Commonwealth University, Richmond, VA, USA

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_10

the system on samples that mimic or exceed the complexity with respect to these potential attributes, both individually and collectively of those encountered in the laboratory's casework. The results of carefully designed internal validation experiments must not only demonstrate a system's utility but also clearly establish and articulate the limits of reliability for use in casework.

Keywords

DNA profiling · Probabilistic genotyping · Validation · Complex mixture · Standard operating procedure

Introduction

Forensic laboratories are increasingly asked to test marginal and complex biological samples (Ford and Krane 2018). However, in the wake of a large-scale interlaboratory study, where analysts struggled to correctly interpret a complex DNA mixture with an ambiguous number of contributors and high levels of allele sharing (Butler et al. 2018), and well-publicized fallout following scrutiny of flawed mixture interpretation practices by several laboratories (Alexander 2015; Plohetski 2017), most forensic laboratories have moved away from conducting manual/conventional interpretation of complex DNA mixtures (Ford and Krane 2018).

Probabilistic genotyping software ("PGS") endeavors to untangle the results generated from these challenging yet increasingly common samples. Simply put, PGS systems use computer algorithms to infer plausible genotypes of possible contributors to a DNA profile. When reference DNA profiles are available the software can then attach a statistical weight (in the form of a likelihood ratio) to the inclusion or exclusion of possible contributors (ASB 2020). These PGS systems are advertised for their potential to "overcome the limitations of human review" (Cybergenetics 2021) and "resolve previously unresolvable DNA profiles" (ESR 2021) derived from complex DNA mixtures, i.e., "samples that contain comingled DNA from two or more contributors in which stochastic effects [associated with low level contributors] or allele sharing cause uncertainty in determining contributor genotypes" (Butler et al. 2021).

Because PGS systems: (i) purport to resolve DNA mixtures of such complexity that they exceed the bounds of human ability to interpret, and (ii) will attempt to deconvolve any DNA data presented to them, it is particularly important for forensic DNA analysts to understand the limitations of these systems so that they are not used beyond the bounds of their reliability in casework (Kelly et al. 2014).Validation testing the process of "push[ing] the system until it fails in order to understand the potential limitations-to define the scope of method (and interpretation) reliability" plays a fundamentally important role in establishing and informing analysts of these boundaries (Butler 2014). While validation is necessary across the sciences, its importance to forensics, where the consequence of an errant interpretation may be measured in human lives, is undeniable (OSAC Human Factors Committee 2020).

Validation in forensics generally falls into one of two categories: *developmental* and *internal*. It is important to recognize the distinction between the two because one cannot serve as a substitute for the other. Developmental validation is testing, usually conducted by the developer of a technique or approach, to ensure that the system is practically useful and that it functions as expected (ASB 2020). It should be informed by the developer's understanding of the limitations associated with the algorithms it is implementing. However, it should not be presumed that a tool or system works the same in the hands of its developer as it does for an analyst working in a different laboratory environment (Butler et al. 2021). Instead, a testing laboratory's internal validation must confirm that the system yields reliable results from data generated by the laboratory using its own procedures, reagents, instrumentation, and personnel.

Internal validation is "an opportunity for the user (e.g., DNA analyst) to understand performance of a method in their forensic laboratory environment rather than trusting the provider's (e.g., the software developer) claim that everything works fine" (Butler et al. 2021). It is conducted by a practitioner or laboratory that intends to use the system in casework, serves both to verify that the system works as intended in their hands and to establish the limitations of the system within that particular laboratory, and must be done using samples that mimic those encountered in the laboratory's casework (SWGDAM 2015; Coble et al. 2016). Unlike developmental validation, internal validation is not transferrable from one laboratory to another (Butler 2012). This should not be surprising, given that laboratory-specific parameters and limitations established during internal validation serve as the foundation for each laboratory's standard operating procedures and interpretation guidelines for use by its own personnel with its own facilities (ASB 2018; Butler et al. 2021). At the same time, laboratories are limited by the bounds of the system's developmental validation as they design their internal validation: Any conditions tested during internal validation that exceed the scope of developmental validation necessitate further developmental validation studies before they can be used to expand a laboratory's validated range (ASB 2020).

While this chapter concentrates on PGS limitations associated with sample complexity, software engineering norms additionally demand that validation of complex computer programs like PGS include tests that establish whether algorithms have been correctly implemented, which (like validation testing generally) necessitates falsifiable criteria (i.e., unambiguously articulated requirement specifications that can be tested against) (Adams et al. 2018; Adams 2018). However, discussions of PGS in the forensic biology community have consistently emphasized *genotyping* over a focus on the disciplines behind the adjoining words – *probabilistic* (statistics) and *software* (computer science). Some forensic standards and guidelines mention software testing in the context of developmental and internal validation testing (FBI 2020), including reference to software and hardware verification and validation standards such as IEEE Std 1012–2012 (ASB 2020), but internalize few of the principles from those standards and provide very little guidance to forensic laboratories that are not familiar with the rigors of software testing. The failure to apply well-established principles of software testing to validation of PGS systems is a

subject that demands the forensic community's attention and is increasingly drawing the attention of software engineers (Canellas 2021), but is beyond the scope of this chapter. Likewise, this chapter does not delve into the extent to which software bug patches and other updates to previously internally validated PGS systems necessitate new internal validation studies, but it is noteworthy that existing standards and guidelines do not provide clear guidance in this area.

This chapter's primary focus is on the internal validation of probabilistic genotyping systems for forensic casework use, and particularly assessing PGS performance in deconvolving DNA mixtures (i.e., determining the most likely combinations of alleles within and between loci for each contributor to a mixed sample). This chapter aims to clarify what it means to "push the system until it fails" so as to establish robust and objective boundaries that ensure the reliable use of PGS in casework, rather than provide a precise formula for study design. It also discusses the importance of reporting the results of validation testing in a way that is clear both to external reviewers and to analysts who will be using the technology in casework. "Every analyst that is interpreting DNA typing data should be familiar with and understand the validation studies. The performance characteristics and limitations of an instrument, a software program, and a DNA typing assay are important to understand in order to effectively interpret forensic DNA data" (Butler 2012).

Standards and Guidelines Governing Internal Validation

Approaching PGS validation testing as an investigation into the capabilities and limitations of the software rather than a series of tasks to be completed facilitates robust coverage of the various dimensions that impact system performance (Butler and Iyer 2020). Of course, broad categories of studies are required in the validation of any DNA analysis method –accuracy, precision, sensitivity, and specificity, but critical thinking about the dimensions to be evaluated within these broad categories is essential.

Published Guidance Documents

Published guidance documents tend to speak in broad terms about the importance of thoroughly evaluating PGS performance in "a range of situations that span or exceed the complexity of the cases likely to be encountered in casework," with special emphasis on the importance of evaluating the system's sensitivity (performance *vis a vis* true contributors) and specificity (performance *vis a vis* noncontributors) for complex DNA profiles (Coble et al. 2016). For example, the Scientific Working Group for DNA Analysis Methods (SWGDAM)'s validation guidelines state: "Each laboratory seeking to evaluate a probabilistic genotyping system must determine which validation studies are relevant to the methodology, in the context of its application, to demonstrate the reliability of the system and any potential limitations" (SWGDAM 2015). In particular, "complex mixtures and low-level

contributors should be evaluated thoroughly during internal validation, as the data from such samples generally help to define the software's limitations, as well as sample and/or data types which may potentially not be suitable for computer analysis" (SWGDAM 2015). These statements provide useful guidance to laboratories performing internal validation. They must: (i) craft their validation of PGS with an eye toward the uses to which the system will be put; (ii) thoroughly explore the dimensions that may impact performance; and (iii) practically identify limitations and the outer boundaries of reliability so that all analysts within the laboratory can consistently generate (and, ultimately, testify about) reliable test results.

Guidance documents consistently indicate that a validating laboratory is required to think critically about the uses to which it plans to put the software in casework. But those documents are less explicit regarding what might constitute a "thorough exploration" in the context of a laboratory's internal validation. For instance, the SWGDAM guidelines contain broad statements encouraging performance-driven evaluation that are followed by a list of dimensions that "should" be addressed, "where applicable" (SWGDAM 2015). Many laboratories include this list and cross-reference it with relevant sections of their internal validation (Butler and Iyer 2020). For example, section 4.1.6 of the SWGDAM guidelines lists various dimensions that should be addressed with respect to mixed samples, including 4.1.6.5, "sharing of alleles among contributors." Read together with the introductory language encouraging thorough examination, particularly involving complex mixtures, section 4.1.6.5 provides guidance about one particular aspect of complex mixtures that should be specifically evaluated by any forensic laboratory that plans to use PGS with complex DNA mixtures in case work. From a performance-driven perspective, it is quite important to include an evaluation of varying degrees of allelic overlap, both because mixtures with high levels of allele sharing are known to be exceptionally challenging and prone to erroneous interpretation under a traditional interpretation framework (Butler et al. 2018), and because situations where high allelic overlap could occur such as the possibility of multiple family members contributing to a DNA mixture are commonly encountered in forensic casework.

Unfortunately, some forensic laboratories have focused on the list in SWGDAM's section 4.1.6 at the expense of the general principles articulated elsewhere in the guidelines. Such an approach can result in superficial, task-driven studies that treat the list of dimensions as an optional checklist (Butler and Iyer 2020). This has resulted in wide discrepancies in validation efforts by laboratories with respect to SWGDAM guideline section 4.1.6.5. Some laboratories have validated their PGS with samples that have high levels of allele sharing (e.g., mixtures comprised of relatives) and include detailed analysis of the system's performance on these challenging samples. Other labs have endeavored to clear a much lower bar by simply including (and not necessarily critically evaluating) mixed samples with varying amounts of allele sharing across loci. Still other labs have emphasized that SWGDAM only provides guidelines, not requirements, and opt out of any consideration of allele sharing for validation purposes even though they may encounter samples with allele sharing in their forensic work. Interestingly, laboratories in each

of these three categories point to the very same language in section 4.1.6.5 as support for the approaches they have taken.

The absence of any enforcement mechanism in connection with guidelines such as those developed by SWGDAM effectively renders laboratories self-policing with respect to the quality of their validation – at least until admissibility challenges about their work are raised in court. While some criminal justice stakeholders have pointed to accreditation as a testament to the quality of laboratories' validation efforts, the focus of a forensic laboratory accreditation process is limited to an assessment of conformity with a set of minimum standards. These standards require laboratories to have a procedure in place for validation, but do not critically evaluate study design or the representativeness of samples used for validation in light of the intended use of the method or process being validated (ANAB 2019). This is not intended as a criticism of the accreditation process, but a recognition of the reality of its scope. Audits associated with accreditation rely upon a limited sampling of case work and processes and cannot be expected to effectively evaluate whether case work use exceeds the scope of internal validation. Moreover, accreditation agencies acknowledge that they cannot enforce the SWGDAM guidelines because they are couched as recommendations ("should") rather than requirements ("shall") (Texas Forensic Science Commission 2017); as a result, major technical problems may evade detection by the accreditation process (Garcia 2019). For example, accreditation and routine audits failed to call attention to significant deficiencies in Austin Police Department Forensic Science Division's ("APD") internal validation of its DNA mixture interpretation protocol; it was only after a chance discovery of case work where, for example, low-level DNA data was erroneously reported as incriminating, and that APD's internal validation was reexamined and found to be inadequate (Texas Forensic Science Commission 2016).

Because laboratories cannot rely upon accreditation as a robust quality check for their validation studies, it is incumbent that they themselves think critically about how to meaningfully explore the dimensions that impact performance in order to establish and fully appreciate the limitations of PGS systems. Recently, the Organization of Scientific Area Committees and American Academy of Forensic Sciences Standards Board promulgated standards that provide laboratories and practitioners more prescriptive guidance in this area. For example, laboratories are required to conduct sensitivity and specificity studies (i.e., exploration of true contributors and noncontributors) on representative samples that include compromised DNA samples subject to the effects of low template, allele sharing, and differential degradation (ASB 2020). Annexes to standards further clarify, for example, that investigations of allele sharing require consideration of mixtures with particularly high levels of allele sharing, such as mixtures containing DNA from two or more close relatives (ASB 2018).

Framed as requirements rather than recommendations, the ASB standards are intended to add flesh to the bare-bones requirements of the Federal Bureau of Investigation's *Quality Assurance Standards for Forensic DNA Testing Laboratories* (ASB 2018). If the internal validation performed by a laboratory to develop its existing interpretation and comparison protocols does not comply with the

requirements, laboratories are advised to supplement that earlier validation. These ASB standards certainly constitute an improvement to prior guidelines, both in form and substance. However, at this time the ASB standards (like the SWGDAM guidelines) have no enforcement mechanism. Time will tell how many laboratories will voluntarily adhere to these standards, or whether they will be formally incorporated into the audit/accreditation process. The fact that they constitute a consensus of experts in the field and are the product of a process that elicited and considered feedback from the public may make them difficult to disregard – especially when raised during the course of admissibility challenges in trials.

Validation Workshops and Technical Merit Evaluation by NIST Scientists

For decades, scientists from the National Institute of Standards and Technology (NIST) have provided a service to the forensic biology community by conducting accessible trainings on validation and other topics, as well as reviewing and publishing on the scientific bases of forensic methods. These scientists have filled a critical gap in most guidance documents by connecting the spirit of thorough assessment of PGS performance via evaluation of challenging and marginal samples with an articulation of how this translates to the dimensions that are relevant to analysis of complex DNA mixtures (Butler and Iyer 2020; Butler et al. 2021). Emphasizing that understanding the limitations of PGS requires exploration of "the 'edges' and beyond" (Butler and Iyer 2020), they encourage laboratories and practitioners to reject a checklist mentality and think critically about the purpose of each experiment in designing their studies.

In particular, consideration must be given to *how* to explore the dimensions that impact the reliability of mixture interpretation on a representative range of case-type samples, a concept NIST scientists refer to as "factor space" coverage (Butler and Iyer 2020; Butler et al. 2021). These scientists' guidance encourages laboratories to be thoughtful in light of intended use and fleshing out limitations: For example, many laboratories fail to test mixture ratios more extreme than 20:1, even though more extreme ratios are common in routine case work. Likewise, even though allele sharing is well known to impact mixture deconvolution and particularly high levels of allele sharing can be expected to arise in common case work situations, the design of many laboratories' validation studies fails to account for these realities. Exploring samples with overlapping alleles and artifacts is essential too: Mixtures with minor alleles in stutter positions must be evaluated to determine the limits of the software's reliability because the ability to discriminate between real alleles and stutter artifacts is critical to the performance of PGS systems.

Equally important to a meaningful exploration of the dimensions that affect complex mixture interpretation is an appreciation that these factors/dimensions do not exist in isolation. They can (and do) overlap to multiplicatively compound the complexity of interpreting a mixture, with the most challenging samples combining issues related to number of contributors and low template and mixture ratio *and*

allele sharing, and so forth. Thus, evaluation of these dimensions should not be treated as separate silos. Reliability of PGS for casework requires well-designed studies that evaluate mixtures that challenge the system along multiple dimensions at once (Butler and Iyer 2020; Butler et al. 2021). Indeed, internal validation should specifically include mixtures that *exceed* the complexity of samples anticipated in casework, not only in order to learn the breaking point in the system, but also as a practical way of ensuring that case-specific reliability is understood without having to navigate the edges of validation (Butler et al. 2021).

In short, a well-designed PGS internal validation study requires thoughtful consideration of the factors that are well understood to make mixture deconvolution challenging – often to the point of being uninterpretable (at least by human experts). The results of those experiments must then be critically evaluated – not just from the perspective of demonstrating a system's utility but also in a way that establishes limitations for casework use. Unfortunately, the checklist mentality described above, and overreliance on the audit/accreditation process to expose insufficiencies, can undermine well-intentioned efforts. Sample PGS validation plans provided by some textbooks attempt to provide more concrete advice to laboratories and practitioners, but neglect fundamental factors that are known to impact mixture deconvolution, such as allele sharing and DNA quality issues like degradation (Butler and Iyer 2020). Best scientific practice demands that any laboratory validating PGS systems carefully considers the range of sample types seen in casework and particularly the most challenging DNA mixtures that may be encountered, and design validation testing to exceed that (Butler and Iyer 2020; Butler et al. 2021). The next section endeavors to apply the spirit of performance-driven evaluation described above to some of the most critical dimensions impacting mixture deconvolution in order to further flesh out what it means to "push the system until it fails" in the context of PGS and complex DNA mixtures.

Key Principles

- Accreditation does not ensure a robust quality check of validation processes and cannot be relied upon to discover inadequacies.
- Laboratories must think critically about how they establish and fully appreciate the limitations of PGS.
- Validation studies must evaluate mixtures that mimic the most complex casework (i.e., challenge the system along multiple dimensions at once).
- Pushing the system until it fails not only allows a laboratory to fully appreciate the limitations of PGS, but also creates a practical path to understanding case-specific reliability without having to test every possible permutation of factors.

Pushing the System Until It Fails

Probabilistic genotyping systems do not replace human evaluation of test results. Rather, they are tools that can be used to assist analysts in the interpretation of DNA test results. The need for their assistance is generally greatest for samples that are

least suitable for analysis with conventional statistical approaches, such as those with a large and/or unknown number of contributors where allelic dropout may have occurred. The analysts that use these tools must understand the limits of their reliability in the same way that they understand the limits of the rulers that they use to measure the dimensions of evidence samples or the limits of the commercially available test kits that they use to generate DNA profiles.

The validation of any DNA interpretation methodology or tool should be designed to assess its precision, accuracy, sensitivity, and specificity for a broad range of samples that reflect in all respects the types of samples/results that will be encountered in casework. For PGS systems, *precision* studies assess the variation in statistical weights (expressed as likelihood ratios) generated from repeated analyses of the same input data (ASB 2020; SWGDAM 2015). Evaluation of *accuracy* generally assesses the system's ability to generate "correct" likelihood ratios by comparing PGS results with those of human analysts and those that have been generated by other reliable methods (ASB 2020; SWGDAM 2015). Thus, PGS accuracy studies are typically limited to sample types that are within human ability to interpret, i.e., single-source samples and simple mixtures, and are therefore of limited value in establishing the accuracy calculations for complex mixtures (Butler et al. 2021). This shortcoming in accuracy assessments translates directly to a need for a higher level of scrutiny in PGS sensitivity and specificity studies.

Sensitivity studies evaluate factors that can lead the system to deliver false negatives (i.e., where a true contributor to a sample is determined to be a non-contributor). *Specificity* studies assess factors that can cause the system to deliver false positives (i.e., a known noncontributor to a sample is included as a possible contributor) (ASB 2020; SWGDAM 2015) and require DNA profiles to be assessed against robust databases of known noncontributors. These studies are a vital element of PGS validation, particularly in the absence of any direct accuracy assessments for complex mixtures. This is especially so given that these systems are used to support assertions regarding the presence or absence of DNA from persons of interest on items of evidence in criminal cases. Assessments of accuracy, sensitivity, and specificity can only be made by using samples where ground truth (e.g., known contributors and known noncontributors to samples) is known (SWGDAM 2015).

The specificity of a method is usually expressed in terms of the magnitude and frequency of false positives; for probabilistic genotyping, the magnitude would be the reported likelihood ratio for known noncontributors that are erroneously included as potential contributors to a DNA profile (i.e., produce a likelihood ratio favoring inclusion (>1)). Because the frequency of false positives for PGS is known to vary under different conditions (e.g., more false positives at lower template levels), calculation of one universal false positive rate is generally not possible or appropriate. Rather, the tendency of the system to produce false positives under different conditions (e.g., low template, high allele sharing, etc.) must be assessed. Sensitivity assessments requires similar exploration of the tendency of the system to produce false negatives ($LR < 1$) under different conditions.

All probabilistic genotyping approaches are built upon models that attempt to describe the relationship between a response (e.g., allelic dropout) and an explanatory variable (e.g., peak height) in a DNA test result. However, it must be understood

that these models are only approximations of the real relationships, and robust testing of these models is crucial. A scatterplot might support the proposition that there is a linear relationship within a tested range of values, but it would be inappropriate to assume that the nature of the relationship and how well the model performs outside of the range of data collected during a laboratory's internal validation studies is known. Establishing that a ruler works well in measuring things to the nearest 1 cm should not be extrapolated to confidence that it also provides accurate measurements of objects that are only 0.01 cm long. Similarly, it cannot be presumed that assessments of sensitivity, specificity, precision, and accuracy will apply to samples outside the range of those used in PGS validation work.

Dimensions That Must Be Evaluated Individually and Collectively

Currently popular probabilistic genotyping systems will generally report a likelihood ratio as output regardless of the quality of the data it is provided as input. However, the interpretation of forensic DNA test results becomes dramatically more difficult as the complexity of the DNA profile increases. Internal validation studies must be designed to thoroughly evaluate the factors or dimensions that make DNA profile interpretation challenging, and in so doing identify specific characteristics of evidence samples, test results, and/or software diagnostics that fall outside the range of values for which the system has been shown to be reliable. Superficial evaluation of the reliability of PGS output (such as what might be assumed to be self-reporting of the system through the magnitude of likelihood ratios themselves) must be strongly discouraged and not be allowed to take the place of the critical evaluation of multiple factors that internal validation studies have identified as being important.

Factors that complicate and cause greater uncertainty in the interpretation of DNA test results may include one or more from the following nonexhaustive list (Butler et al. 2021):

- Suboptimal amounts of template DNA
- Skewed ratios of the amount of DNA from different contributors to a mixture
- A large and/or unknown number of contributors
- Allele sharing between two or more contributors to a mixture (as well as between true contributors and noncontributors)
- Degradation (including varying degrees of degradation between contributors) of template DNA and inhibitors of DNA amplification\

Just as DNA profiles encountered in casework may be suffering from the combined effects of more than one of these complicating factors, internal validation must include DNA profiles that are challenging along multiple dimensions in order to test the outer limits of PGS reliability (Butler and Iyer 2020).

Suboptimal Amounts of Template DNA

All commercially available DNA profiling test kits describe a range of DNA template quantities, generally between 0.5 and 1.0 nanograms, from which optimal results can be reliably obtained for single-source samples. It has been appreciated since 1992 that use of suboptimal amounts of template DNA can be expected to lead to a number of stochastic effects (e.g., peak height imbalance to the point of allelic dropout; drop-in; exaggerated stutter artifacts) that can make the interpretation of test results challenging (Walsh et al. 1992). In simple terms, stochastic effects arise from sampling errors that can, to some extent, be modeled by probability distributions within the most sophisticated probabilistic genotyping systems in lieu of traditional signal thresholds. However, studies by PGS developers and testing laboratories alike have found that false positives and false negatives become more prevalent with lower levels of template DNA, even for these sophisticated systems (e.g.,Bright et al. 2014; Palm Beach County Sheriff's Office Laboratory 2017). Moreover, if the alleles within a DNA profile are at a level that is dominated by background noise, then a biostatistical interpretation for these alleles should not be attempted (Gill et al. 2006).

Thus, it is critical that testing laboratories evaluate the sensitivity and specificity of PGS for low-level contributors as part of their internal validation (i.e., likelihood ratios produced for trace contributors and noncontributors), particularly given that the accuracy of PGS output cannot be directly measured for complex mixtures that exceed the bounds of human ability to interpret. Each laboratory must specifically determine the minimum quantity of an individual contributor's template DNA for which it has demonstrated the ability of the PGS system to consistently distinguish known contributors from known noncontributors. Likelihood ratios generated for a potential contributor whose template levels fall below this level should not be considered to be reliable. It is inappropriate to assume that lower levels of contributor template than those tested during internal validation will produce accurate and reliable results.

Variation in Contributor Ratios

Continuous PGS systems primarily utilize peak height information in DNA testing results to deconvolve mixtures. There are very different challenges to reliable deconvolution at either end of the spectrum of mixture ratios. At one extreme, peak height information is of little help in deconvolving mixtures where all contributors are responsible for adding the same amount of template DNA to a sample; thus, mixtures comprised of multiple individuals contributing similar amounts of template DNA must be specifically evaluated during internal validation. At the other extreme, the peaks associated with those of a minor contributor may be masked by the contributions of other contributors or be comparable to commonly encountered forward or reverse stutter artifacts associated with a major contributor's peaks (Butler and Iyer 2020). As with contributors displaying substantial allelic dropout,

PGS systems may have difficulty distinguishing true contributors from non-contributors where there is extensive masking (Bright et al. 2014). Minor contributors in highly skewed mixture ratios are also likely to simultaneously manifest the complications associated with suboptimal amounts of template DNA described above. It is therefore essential that internal validation data be used to determine specific limits of reliability both in terms of a minor contributor's relative contribution and in terms of the quantity of template DNA originating from that contributor.

Section 4.1.6.1of the SWGDAM guidelines specifically suggests that testing laboratories include various contributor mixture ratios as part of their internal validation of a PGS system, providing as an example ratio of "1:1 through 1:20, 2:2:1, 4:2:1, 3:1:1, etc." (SWGDAM 2015). Many laboratories take the quoted ratio values quite literally, and consequently it is relatively rare for laboratories to conduct internal validation on mixtures more skewed than 20:1 (Butler and Iyer 2020). Laboratories should see the ratios provided by SWGDAM as an example, not a prescription for all circumstances. If a laboratory plans to use a PGS system on mixtures more skewed than 20:1 casework, the laboratory needs to validate its PGS system on similarly skewed mixtures in order to understand the outer boundaries of reliability. As with suboptimal DNA template, each laboratory must determine the minimum contributor ratio for which it has demonstrated the ability of the PGS system to consistently distinguish known contributors from known noncontributors. Where a potential contributor's ratio of the mixture falls below this level, associated likelihood ratios cannot be assumed to be reliable.

A Large and/or Unknown Number of Contributors (NOC) to a Sample

At the polymorphic loci used for human identification, individuals will typically have either one (if they inherit the same alleles from mother and father) or two (if they inherit different alleles from mother and father) alleles. The simplest explanation for a locus where three or four alleles are observed is to say that the tested sample was a mixture of at least two individuals. By the same token, the simplest explanation for a locus where five or six unique alleles are observed is to say that the tested sample was a mixture of at least three individuals. However, the process of determining the exact number of contributors (NOC) to a sample is complicated by the fact that it is possible for individuals to have alleles in common. In short, the potential for individuals to have "overlapping" or "shared" alleles can lead to an underestimate of the number of contributors to a sample (Fig. 1) (Paoletti et al. 2005; Coble et al. 2015). As a result, it is important to note that in this context "*at least* three individuals" should not be taken to mean "three individuals" or even "most likely three individuals." Conversely, stochastic effects such as allelic drop-in and exaggerated stutter can lead to an overestimation of the number of contributors.

Currently popular PGS programs require analysts to input an actual (not an estimated or minimum) NOC for each hypothesis that is to be evaluated. Some research reports that both over and underestimating the NOC to a mixture can result in false positives (i.e., likelihood ratios over 1 for known noncontributors) (Benschop et al. 2015; Bille et al. 2019). At the same time, one multilaboratory analysis by the developers of a widely used PGS system suggests that at least one

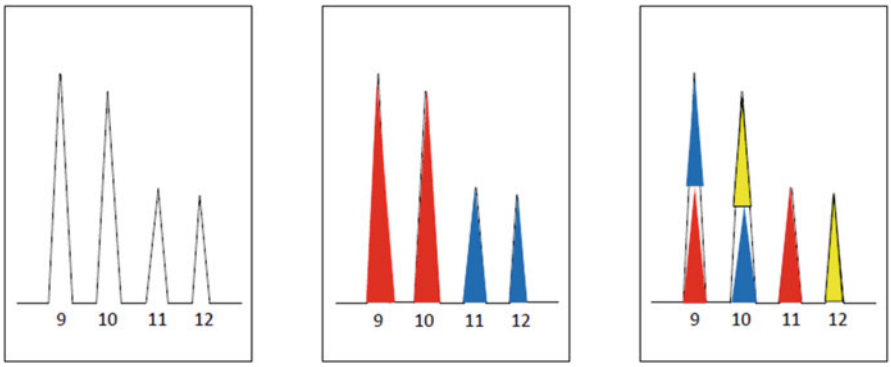


Fig. 1 Demonstration of uncertainty in NOC due to allele sharing. The first panel shows an evidentiary (i.e., unknown) mixture at a hypothetical locus. One possibility is that the mixture is comprised of two contributors, as represented in the second panel (i.e., one contributor (red) possesses a 9, 10 and the second (blue) possesses an 11, 12). However, if contributors to a mixture "share" alleles, meaning two or more individuals possess the same allele at a particular locus, the allelic contributions stack upon one another, forming one peak in the electropherogram. In the third panel, although there are three contributors to the mixture, each possessing two alleles, only four alleles are detected: the genotype of one (red) is 9, 11; the second (yellow) is 10, 12; and the third (blue) shares a 9 allele with the first contributor and a 10 allele with the second contributor. Allele sharing can lead to underestimation of the NOC that may impact downstream interpretation efforts

version of that system can produce reliable results using "the apparent number of contributors" (effectively, the observed minimum number of contributors) to a sample for mixtures known to contain DNA from six individuals but which can be explained by a minimum of at least four contributors (Bright et al. 2018).

It is therefore necessary for a testing laboratory to perform its own internal validation to determine not only the maximum number of contributors for which its probabilistic genotyping system can provide reliable likelihood ratios that appropriately distinguish between contributors and noncontributors, but also the means by which it determines the number of contributors to input into its system and the implications associated with an analyst entering both correct and incorrect values (i.e., over- and underestimating) (ASB 2020;SWGDAM 2015). Further, if a testing laboratory intends to use a PGS system on mixtures of up to N contributors, it should specifically evaluate mixtures of at least N + 1 contributors that are underestimated by one contributor. Demonstration that a probabilistic genotyping system yields results with a low rate of false positives or false negatives for samples with just a few contributors should *not* be taken as a demonstration that the system is a reliable tool for casework samples with larger or unknown/contested numbers of contributors.

Allele Sharing Between Contributors to a Sample, and with Noncontributors

Crime scene samples are often collected from locations and under circumstances where closely related individuals may be considered to be additional and/or alternative contributors of DNA. Most people share some number of alleles (Fig. 2). Allele

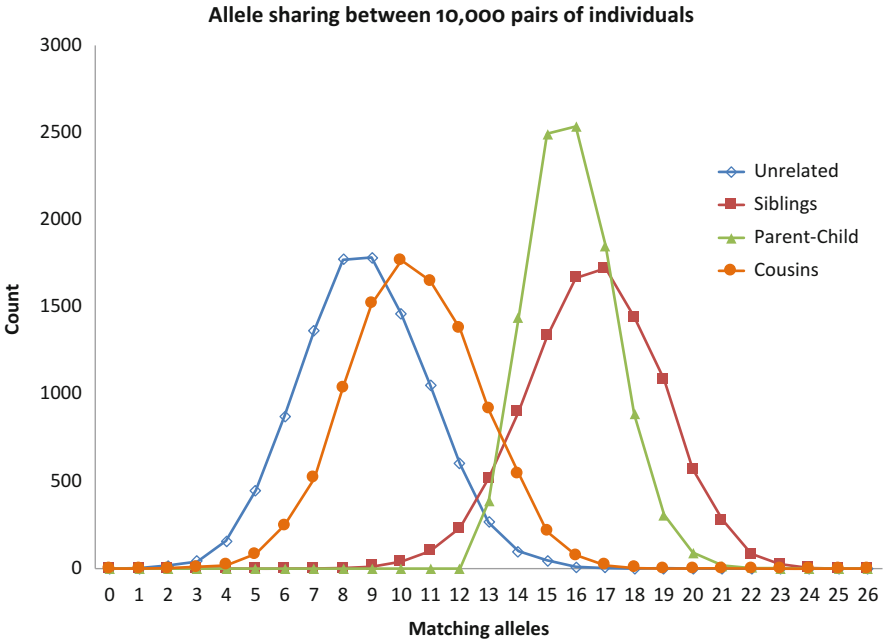


Fig. 2 Pairwise allele sharing between simulated individuals with four different degrees of relatedness: unrelated, sibling, parent/child, and cousin. Genotypes at the 13 CODIS loci were generated using NIST Caucasian allele frequencies. Ten thousand pairs of simulated individuals were compared at each locus (matching homozygotes were scored as having two matching alleles) for each of the four kinds of relatedness

sharing between pairs of close relatives (e.g., siblings) is likely to be much greater than between randomly chosen, unrelated pairs of individuals.

Allele sharing can lead not only to underestimates of the number of contributors to mixed DNA samples, it can also give rise to potentially misleading characterizations of difficult-to-interpret test results. It is not sufficient for a testing laboratory to include DNA mixtures with varying degrees of allele sharing as part of their internal validation without actually evaluating how the levels of allele sharing impact the performance of the PGS system in their hands. Laboratories must intentionally include, and specifically assess PGS performance upon mixtures with high levels of allele sharing in their internal validation. If laboratories "do not explore the complexity induced by allele sharing," their analysts "may inadvertently extrapolate validation results and apply methods beyond the limits of the validation studies conducted" (Butler et al. 2021). One way to explore high levels of allele sharing is to test mixtures of DNA from close relatives (ASB 2018; Butler and Iyer 2020).

Including mixtures of related individuals also allows the laboratory to explore one of the most challenging scenarios for a PGS system: a mixture including two or more close relatives assessed against a related noncontributor (simulated or real) as the person of interest. A growing number of public crime laboratories have specifically

examined the likelihood ratios that result from this kind of situation as part of their internal validations, obtaining "false positives" (also known as misleading likelihood ratios) many orders of magnitude higher than those produced when an unrelated noncontributor is assessed as the person of interest. For example, one laboratory's internal validation found that for mixtures of unrelated individuals assessed against unrelated noncontributors, the highest false positive obtained was on the order of 10^4 (ten thousand). When mixtures that included related individuals were assessed against unrelated noncontributors, the highest false positive rose to 10^6 (million). Assessed against a noncontributor who was related to the mixture contributors, the highest false positive vaulted to 10^{17} (hundred quadrillion). In fact, the noncontributor frequently produced higher likelihood ratios than true contributors to the mixture (Palm Beach County Sheriff's Office Laboratory 2017).

Such seemingly impressive likelihood ratios for known noncontributors in internal validation work raises serious questions about exactly how likelihood ratios are conveyed to participants in the criminal justice system. They are also a testament to the absolute necessity of including strong cautions within a laboratory's interpretation guidelines that even the most sophisticated PGS systems may produce extremely high "false positive" or "misleading" likelihood ratios for noncontributors who are related to two or more contributors to a mixture, suggesting extremely strong support for a false hypothesis – namely, the inclusion of the noncontributor to the mixture.

Degradation/Inhibition and Differential Degradation

Some sophisticated PGS systems endeavor to model the effects of DNA degradation. Those models typically also apply to the effects of inhibition of PCR amplification of template DNA in that both inhibition and degradation tend to give rise to a "ski-slope effect" on electropherograms whereby amplified DNA fragments exhibit progressively smaller peak heights as the length of the amplified region increases (Fig. 3).

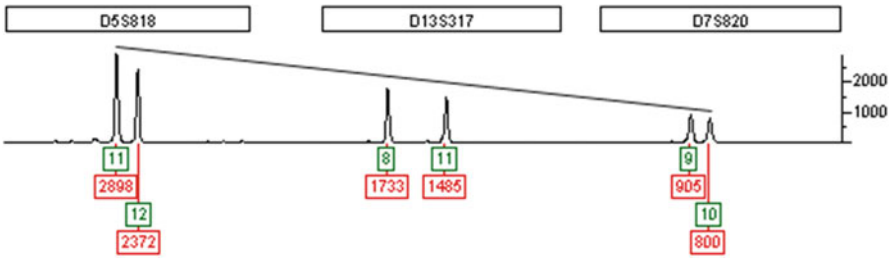


Fig. 3 Peak heights become progressively smaller from left to right on electropherograms for DNA that has been degraded and/or inhibited. The horizontal axis represents the size of amplified DNA fragments with the smallest appearing on the left. Loci names are shown in boxes above the electropherogram. Allele designations are shown in green boxes below peaks while peak heights (in relative fluorescent units, RFUs) are shown below peaks in red boxes. (Figure from Thompson et al. 2003.) Note that this figure displays a single-source sample; when there are multiple contributors to a sample in different states of degradation (as well as different amounts of template DNA), evaluation of degradation becomes exponentially more difficult

The practical implication of these degradation/inhibition models is that the probability of allelic dropout is not the same across all tested loci. It is therefore necessary to evaluate the performance of PGS systems with samples exhibiting widely ranging degrees of degradation/inhibition. As with suboptimal DNA template and extreme mixture ratios, PGS systems' ability to distinguish between true contributors and noncontributors diminishes as the degradation/inhibition becomes more extreme. However, it is more challenging to quantify the level at which degradation/inhibition effects are problematic or outside the bounds of validation testing, than it is to do the same with template levels and mixture ratios. Thus, through its internal validation the laboratory must identify sufficiently detailed qualitative indicators in the mixture profile and quantitative indicators in the PGS output to permit an analyst to recognize a sample or a test result where degradation/inhibition levels may negatively impact the reliability of PGS results.

It is worth noting that not all contributions to a mixed DNA sample will necessarily be degraded to the same degree. Some contributions may have been added to the mixture days or even years after others and it is possible that DNA from some biological materials could persist longer than others. An exploration of the impact of "differential degradation" on the reliability of a PGS system is therefore important as well (ASB 2020; Butler et al. 2021).

Additional Factors That Complicate the Interpretation of Mixed DNA Samples

This chapter does not attempt to provide an exhaustive list of factors that testing laboratories should evaluate in the context of their internal validation of a probabilistic genotyping system. Many testing laboratories have found it necessary and useful to evaluate other factors such as the implications of using allele frequency databases that are not representative of reasonable alternative contributors; the reliability of a probabilistic genotyping system when confronted with common technical artifacts such as spikes (with associated "bleed through") and blobs; and additional factors that are identified as probabilistic genotyping models become increasingly sophisticated.

Factor independence should not be assumed until demonstrated though rigorous multifaceted internal validation by a testing laboratory; for example, mixture ratio effects often have implications for other factors such as template quantity and degradation. Regardless of whether they are demonstrated to be independent, several of these complicating factors often impact crime scene samples simultaneously. For internal validation to be representative of casework, it must include samples designed to suffer the combined effects of multiple factors at once.

After validation data is generated and standard operating procedures are created based on that data (as discussed further below), the protocols must be verified using a different set of samples than were used during validation testing (ASB 2018), and should include a range of sample types (Butler and Iyer 2020). It is possible that the boundaries of validated reliability for a laboratory will expand (or contract) over

time as more experiments are conducted, whether in response to a perceived need due to a change in the type of samples encountered in the laboratory's casework, or as new test kits, equipment, or updated versions of probabilistic genotyping systems are put into service. Any revisions to the standard operating procedures based on supplemental validation experiments must be similarly verified.

Key Principles

- PGS systems all use models that are only approximations and robust testing of those models is crucial.
- A system's accuracy on results that humans can interpret cannot be taken as proof of accuracy for results that are too complex for human interpretation.
- False positive rates must be documented under different circumstances, and the intersection of multiple complexity factors must also be investigated.
- Assessments of sensitivity, specificity, precision, and accuracy cannot be extrapolated beyond the range of conditions used in internal validation studies.
- PGS systems may have difficulty distinguishing true contributors from non-contributors where there is extensive masking.
- Laboratories must specifically assess PGS performance on mixtures with high levels of allele sharing (such as occurs between closely related individuals).

Using and Sharing Validation Data and Results

It is of vital importance that results generated during a laboratory's validation testing of PGS, and associated limitations of the software, be communicated in a manner that is clear to analysts who will be using and testifying about the technology, and that the underlying data be fully accessible to external reviewers. Regrettably, a recent foundational review by NIST scientists found that there was "not enough publicly available data to enable an external and independent assessment of the degree of reliability of probabilistic genotyping software (PGS) systems," particularly given the lack of "established and accepted criteria for reliability with complex mixtures involving contributors containing low quantities of DNA template or where there is a high degree of allelic overlap between contributors" (Butler et al. 2021). This underscores the need for individual laboratories to be transparent with their studies and resulting data.

Incorporating Validation Results into Standard Operating Procedures

The results of a testing laboratory's internal validation constitute the foundation of its written standard operating procedures (ASB 2018, 2020). These documents should give analysts specific guidance about characteristics of samples and test results that make them unsuitable for interpretation or comparison with the laboratory's PGS software, and features that indicate PGS output is potentially unreliable (ASB 2018;

Butler et al. 2021). To ensure accuracy as well as consistency in the work of all analysts, those protocols and interpretation guidelines must include unambiguous and detailed guidance, tightly connected to the laboratory's validation, regarding the specific limits to the system's robustness and reliability in the context of complicating factors that make the interpretation of DNA test results challenging (ASB 2018, 2019). It is essential that testing analysts have a clear understanding of the limitations of the PGS software (ENFSI 2017).

The concept of including complexity and data quality thresholds based on validation data in DNA interpretation guidelines is not new (Butler 2015). Robust thresholds can help analysts identify samples that are too complex or too poor quality to reliably resolve, making interpretability decisions more repeatable and reproducible. Thresholds may be based on the relative number of false positives and false negatives under certain conditions (Butler and Iyer 2020; Butler 2015). These thresholds and criteria are necessary to establish a laboratory's "acceptable degree of reliability" (Butler et al. 2021), which may vary from lab to lab (Butler and Iyer 2020) and possibly even from one criminal offense to another. The variability between laboratories on this point underscores the importance of transparency: For any sample that is deemed interpretable, an analyst (or someone reviewing the work product of an analyst) should be able to look to the standard operating procedures for the rate of false positives and false negatives and other indicia of reliability under conditions comparable to that observed in a casework sample (Butler and Iyer 2020; Butler et al. 2021). The false positive rate "should be characterized as a limitation of the software," and should be clearly communicated in the standard operating procedures (Butler 2015). Along with the rate, the range of false positive likelihood ratios (i.e., LRs higher than 1) produced under different conditions in validation should be clearly communicated. Including examples of DNA profiles considered unsuitable for interpretation and/or comparison alongside information about false positives may assist analysts in understanding, and thus appropriately communicating, the limitations of the PGS software as applied in their laboratory (ENFSI 2017). Gray areas associated with advice like "proceed with caution" should be avoided so as to minimize variability in application between analysts, opportunities for subjective reporting of findings that are susceptible to bias, and overstatement of the probative value of results.

Standard operating procedures should also clearly articulate the boundaries of validation testing, or put differently, the limits of demonstrated reliability. Clearly defined, objective thresholds help analysts identify in a reproducible manner casework samples that are more marginal and/or complex than those tested in validation. For example, most laboratories have a bright-line number-of-contributor cutoff for DNA mixtures based on the maximum number of contributors evaluated during internal validation. In these circumstances, if there is a reasonable possibility that the number of contributors to the mixture exceeds this value, the mixture is deemed to be "uninterpretable." (As stated above, due to the challenges of accurately deducing the number of contributors to complex mixtures, validation testing should include mixtures comprised of at least one more contributor than the laboratory intends to interpret in casework.) Standard operating procedures should include similar cutoffs for other dimensions explored in validation such as:

Mixtures with contributors donating as little as 10% and 100 pg of template DNA were tested during internal validation. If PGS analysis of a mixture associates a person of interest with a contributor donating less than this amount of DNA (in terms of ratio or quantity), the sample is outside the scope of validation and should be deemed uninterpretable.

Unambiguous guidance prevents analysts reporting and/or testifying to conclusions that are not scientifically supportable. Put simply, analysts "cannot support any conclusion drawn for results that are more complex than [their] validation data" (Rudin and Inman 2012). Because the end consumer of PGS technology – judges and juries – are not equipped to discern the limits of its reliability, it is essential that these limitations are appreciated and clearly communicated by testing analysts. "As DNA profiling continues to grow more sensitive, and it is used in more investigations, the need for accurate communication between scientists and nonscientists only grows – both to ensure that their expectations of the technology are realistic, and its limits are properly understood" (SENSE about SCIENCE and EUROFORGEN 2017). Relatedly, it is essential that all parties (i.e., prosecution and defense) understand the limits of reliability of PGS systems so that they can, for example, appropriately push back on any overstatement by testifying analysts. As such, validation performance results for samples that are similar in complexity to the sample in the case at hand should be included in the case file and report as well (Butler et al. 2021).

Transparency of Validation Data for Independent Review

While detailed standard operating procedures tightly connected to internal validation studies are essential to the valid use of PGS systems for DNA interpretation, independent review of validation data is a critical quality check on the validity of the validation studies themselves. Robust independent review requires full transparency of validation data; transparency, in turn, is a basic tenet of science (Nosek et al. 2015). As the National Academies of Sciences, Engineering, and Medicine put it in a consensus report, "research conducted openly and transparently leads to *better science*. Claims are more likely to be credible – or found wanting – when they can be reviewed, critiqued, extended, and reproduced by others" (National Academies of Sciences 2018). This applies no less to the forensic sciences than science in general; if anything, transparency must be a greater priority in the forensic realm, where undetected errors can cause and have caused vast injustice (Chin et al. 2019). Thus, summaries of PGS internal validation studies and underlying data should be "freely share[d] with academics, fellow researchers, and all other parties" (OSAC Human Factors Committee 2020). Concerns about genetic privacy can be addressed by anonymizing alleles, e.g., using letters in place of numeric allelic designators (Butler and Iyer 2020; Butler et al. 2021). This permits reviewers to assess complicating factors like allele sharing without implicating privacy concerns.

The validation summary serves as the bridge between the validation studies and the standard operating procedures, and as such must detail how the data from the studies led to the parameters in the interpretation protocol (ASB 2018). NIST's review of publicly available validation summaries unfortunately found that many

fail to provide the data and information "necessary to assess the degree of reliability and performance under potential case scenarios" (Butler et al. 2021). The summary should assist analysts and reviewers alike "in understanding the limitations of the system and where potential risks may exist" (Butler and Iyer 2020). Conclusory statements without accompanying data are insufficient, as are descriptions of trends and data plots that provide no correlation between individual data points and the samples used to generate them (Butler and Iyer 2020; Butler et al. 2021). Meaningful independent review of internal validation requires access to the following nonexhaustive list of information (Butler and Iyer 2020; Butler et al. 2021):

- Sample identifier
- Sample number (if replicate)
- Number of contributors
- Number of contributors used for PGS analysis (i.e., apparent NOC)
- Target template amounts (total and per contributor)
- Degradation status (per contributor, if differentially degraded)
- H_1 (Hp) true (i.e., is the POI a contributor)? Y/N
- Person of interest (POI) position (if H_1 true), including mixture proportion
- LR or $\log_{10}$(LR) for POI reported by PGS system
- Mixture electropherogram results
- POI profile (whether known contributor or noncontributor)
- DNA profiles for each known contributor (whether or not POI)
- Size of database for noncontributor (i.e., H_2 true) testing
- Analytical threshold used for analysis
- PGS parameters and settings

This data is necessary for an external reviewer to assess, for example, degrees of allelic dropout and allele sharing between mixture contributors, and robustness of noncontributor testing. In short, it allows the reviewer to determine exactly what sample conditions were explored by the laboratory during validation and correlate those conditions with the PGS-reported LRs to assess reliability in a meaningful and nuanced way. If this data is not provided in the validation summary (e.g., in the form of appendices), it must at very least be accessible via the underlying validation data and materials. Good science dictates that the full set of data and materials underlying the laboratory's validation be made publicly available (e.g., through the laboratory's website) or copies be made available upon request (Nosek et al. 2015; National Academies of Sciences 2018).

Key Principles

- Internal validation constitutes the foundation of a laboratory's written standard operating procedures/interpretation guidelines.

- Analysts must have a clear understanding of the limits of the reliability of the PGS system they are using.
- Standard operating procedures should include clearly defined criteria and thresholds to instruct analysts what samples/results are not suitable for interpretation using PGS.
- Robust independent review requires full transparency of internal validation data.

Conclusions

Probabilistic genotyping promises to address one of the most pressing problems associated with the practical application of forensic DNA profiling by attaching reliable statistical weights to test results where complicating factors such as allelic dropout and an unknown number of contributors make DNA data unsuitable for interpretation by human analysts. Coble et al. are correct in saying "Determination of the limits of the software is important to establish the types of profiles that are suitable for handling by the laboratory. Probabilistic software, especially for low-level DNA mixtures may allow a laboratory to widen the scope of their casework in terms of the type of evidence handled. However, there may also be a temptation to submit all complex mixtures to particularly versatile software" (Coble et al. 2016). The results of carefully designed internal validation experiments must be critically evaluated not just from the perspective of demonstrating a system's utility but also in a way that clearly establishes limitations for use in casework.

References

AAFS Standards Board (ASB) (2018) ANSI/ANAB standard 020, First ed. Standard for validation studies of DNA mixtures, and development and verification of a laboratory's mixture interpretation protocol. Academy Standards Board. https://asb.aafs.org/wp-content/uploads/2018/09/020_Std_e1.pdf. Accessed 12 Jun 2021

AAFS Standards Board (ASB) (2019) ANSI/ANAB standard 040, First ed. Standard for forensic DNA interpretation and comparison protocols. Academy Standards Board. http://www.asbstandardsboard.org/wp-content/uploads/2019/10/Std_040_e1.pdf. Accessed 12 June 2021

AAFS Standards Board (ASB) (2020) ANSI/ASB standard 018, First ed. Standard for validation of probabilistic genotyping systems. http://www.asbstandardsboard.org/wp-content/uploads/2020/07/018_Std_e1.pdf. Accessed 12 Jun 2021

Adams N (2018) What does software engineering have to do with DNA? Champion 4:58–66

Adams N, Koppl R, Krane D et al (2018) Letter to the editor – appropriate standards for verification and validation of probabilistic genotyping systems. J Forensic Sci 63:339–340

Alexander KL (2015) National accreditation board suspends all DNA testing at D.C. crime lab. Washington Post. https://www.washingtonpost.com/local/crime/national-accreditation-board-suspends-all-dna-testing-at-district-lab/2015/04/26/2da43d9a-ec24-11e4-a55f-38924fca94f9_story.html. Accessed 12 Jun 2021

ANAB (2019) ISO/IEC 17025:2017- forensic science testing and calibration laboratories accreditation requirements. https://anab.qualtraxcloud.com/ShowDocument.aspx?ID=12371. Accessed 24 Apr 2021

Benschop CCG, Haned H, Jeurissen L et al (2015) The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures. Forensic Sci Int Genet 19:92–99
Bille T, Weitz S, Buckleton JS, Bright J-A (2019) Interpreting a major component from a mixed DNA profile with an unknown number of minor contributors. Forensic Sci Int Genet 40:150–159
Bright J-A, Taylor D, Curran J, Buckleton J (2014) Searching mixed DNA profiles directly against profile databases. Forensic Sci Int Genet 9:102–110
Bright J-A, Richards R, Kruijver M et al (2018) Internal validation of STRmix™ – a multi laboratory response to PCAST. Forensic Sci Int Genet 34:11–24
Butler JM (2012) Review of urban legends. In: Advanced topics in forensic DNA typing: methodology. Elsevier/Academic Press, Walthan, p 190
Butler JM (2014) Validation overview. In: Presented at the NIST DNA analysis webinar series: validation concepts and resources – part 1, Online, 6 Aug 2014
Butler JM (2015) A complexity threshold? In: Advanced topics in forensic DNA typing: interpretation. Elsevier, Boston, pp 176–177
Butler JM, Iyer HK (2020) Validation: principles, practices, parameters, performance evaluations, and protocols. In: Presented at the 31st international symposium on human identification, Online, 18 Sep 2020
Butler JM, Kline MC, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. Forensic Sci Int Genet 37:81–94
Butler JM, Iyer H, Press R, et al (2021) DNA mixture interpretation: a NIST scientific foundation review. NIST. https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8351-draft.pdf. [DRAFT] Accessed 12 Jun 2021
Canellas M (2021) Defending IEEE software standards in Federal Criminal Court. Computer 54(6): 14–23
Chin JM, Ribeiro G, Rairden A (2019) Open forensic science. J Law Biosci 6:255–288
Coble MD, Bright JA, Buckleton JS, Curran JM (2015) Uncertainty in the number of contributors in the proposed new CODIS set. Forensic Sci Int Genet 19:207–211
Coble MD, Buckleton J, Butler JM et al (2016) DNA Commission of the International Society for forensic genetics: recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. Forensic Sci Int Genet 25:191–197
Cybergenetics (2021) What can Cybergenetics do for you? https://www.cybgen.com/services/. Accessed 7 Apr 2021
ENFSI (2017) Best practice manual for the internal validation of probabilistic software to undertake DNA mixture interpretation. https://enfsi.eu/wp-content/uploads/2017/09/Best-Practice-Manual-for-the-internal-validation-of-probabilistic-software-to-undertake-DNA-mixture-interpretation-v1.docx.pdf. Accessed 12 Jun 2021
ESR (2021) STRmix. Resolve more DNA mixtures. https://www.strmix.com/strmix. Accessed 7 Apr 2021
Federal Bureau of Investigation (FBI) (2020) Quality assurance standards for forensic DNA testing laboratories. https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf. Accessed 14 Jun 2021
Ford S, Krane D (2018) The dawning of a new era in DNA profiling. Champion 5:40–49
Garcia L (2019) Hearing: raising the bar: progress and future needs in forensic science. https://www.youtube.com/watch?v=jrNIUTJK8hI. Accessed 12 Jun 2021
Gill P, Brenner CH, Buckleton JS et al (2006) DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. Forensic Sci Int 160:90–101
Kelly H, Bright J-A, Buckleton JS et al (2014) A comparison of statistical models for the analysis of complex forensic DNA profiles. Sci Justice 54:66–70
National Academies of Sciences (2018) A vision for Open Science by design. In: Open Science by design: realizing a vision for 21st century research. The National Academies Press, Washington, DC, pp 107–120

Nosek BA, Alter G, Banks GC et al (2015) Promoting an open research culture. Science 348:1422–1425

OSAC Human Factors Committee (2020) Human factors in validation and performance testing of forensic science. NIST OSAC for Forensic Science. https://www.nist.gov/system/files/documents/2020/05/22/OSACTechSeriesPub_HF%20in%20Validation%20and%20Performance%20Testing%20of%20Forensic%20Science_March2020.pdf. Accessed 12 Jun 2021

Palm Beach County Sheriff's Office Laboratory (2017) Internal validation of STRmix v2.4. PBSO Laboratory. http://www.pbso.org/qualtrax/QTDocuments/4228.PDF. Accessed 12 Jun 2021

Paoletti DR, Doom TE, Krane CM et al (2005) Empirical analysis of the STR profiles resulting from conceptual mixtures. J Forensic Sci 50(6):1361–1366

Plohetski T (2017) Austin police DNA lab closed amid forensics commission's concerns. Austin American-Statesman. https://www.statesman.com/news/20170110/austin-police-dna-lab-closed-amid-forensics-commissions-concerns. Accessed 12 Jun 2021

Rudin N, Inman K (2012) The discomfort of thought – a discussion with John Butler. News of the California Association of Criminalists:8–11

Scientific Working Group on DNA Analysis Methods (SWGDAM) (2015) Guidelines for the validation of probabilistic genotyping systems. SWGDAM. https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_22776006b67c4a32a5ffc04fe3b56515.pdf. Accessed 12 Jun 2021

SENSE about SCIENCE, EUROFORGEN (2017) Making sense of forensic genetics. https://senseaboutscience.org/wp-content/uploads/2017/01/making-sense-of-forensic-genetics.pdf. Accessed 12 Jun 2021

Texas Forensic Science Commission (2016) Final audit report for austin police department forensic services division DNA section. Texas Tribune. https://static.texastribune.org/media/documents/APD_Audit_Final_report_071116.pdf. Accessed 12 Jun 2021

Texas Forensic Science Commission (2017) August 18, 2017 meeting. Timepoint 4:24:12. https://www.youtube.com/watch?v=-p_30-20kQI. Accessed 12 Jun 2021

Thompson W, Ford S, Doom T et al (2003) Part I: evaluating forensic DNA evidence: essential elements of a competent defense review. Part 1. Champion 27(3):16–25

Walsh P, Erlich H, Higuchi R (1992) Preferential PCR amplification of alleles: mechanisms and solution. PCR Meth Appl 1(4):241–250

Potential of DNA Technique-Based Body Fluid Identification

13

Aditi Mishra, Ulhas Gondhali, and Sumit Choudhary

Contents

Abstract

Determination of the type of body fluid evidence recovered at the crime scene is very crucial to forensic investigators, since it facilitates the reconstruction of crime scene the sample donors with actual criminal acts. So far, several conventional methods such as alternative light sources, immunological tests, protein catalytic activity tests, chemical tests, catalytic tests, spectroscopic methods, and microscopy have been developed for identification of body fluids. Majority of these conventional methods are presumptive in their nature and thus possess some limitations regarding specificity, sensitivity, etc. Driven by the importance for forensic applications, body fluid identification methods have been extensively developed in recent years. To overcome the limitations of conventional analysis, molecular genetics-based methods using DNA methylation profiling has been implemented for body fluid identification in forensic context. DNA methylation

A. Mishra (✉) · S. Choudhary
School of Forensic Science and Risk Management, National Security and Police University of India, Rashtriya Raksha University, Gandhinagar, Gujarat, India
e-mail: aditimishra@rsu.ac.in

U. Gondhali
Jindal Institute of Behavioural Sciences, O.P. Jindal Global University, Sonipat, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_11

that occurs at CpG dinucleotides-5′-position of the pyrimidine ring of cytosine residue has proved to be useful markers for identification of blood, semen, saliva, vaginal secretions, and menstrual blood samples. Interestingly, several studies have been conducted that have reported variations at tissue-specific DNA methylation and gene expression. With the time, several tDMRs (tissue-specific differentially methylated regions) have been identified. These tDMRs are found to be specific for each body fluid or cell type and further validated using different samples. In this chapter, we have focused on potential of DNA methylation as a forensic marker and its utility for advanced investigative leads from forensically relevant body fluids.

Keywords

Forensic science · Body fluid identification · Epigenetics · DNA methylation

Introduction

The detection and identification of biological samples found at the scene of crime have an important aspect in criminal investigations by providing valuable evidence (Kayser and De Knijff 2011). Determining the type and origin of most popular body fluids such as blood, saliva, semen, menstrual blood, and vaginal fluid would assist in finding the pivotal information to deduce the action that cause the deposition of these cellular material at the crime scene, thus providing the link between suspect and actual criminal acts (Virkler and Lednev 2009). In order to identify the presence or absence of body fluids at the crime scene, the suspected samples undergo for several forensic analysis including DNA testing which is considered as significant step in a wide range of investigations. It has been seen that occasionally knowing the identity of a fluid could be enough to influence the result of a case. However, many body fluids are invisible to the naked eyes, present in very minute quantities, and found as in mixtures or have similar appearance to other fluids. Therefore, absolute confirmation or identification of body fluids is very much required even when the identities of stains are obvious/known to the forensic experts. The absolute identification of body fluids is helpful to prove or disapprove any facts needed in the court of law. This is particularly important with the possible occurrence of mixtures (An et al. 2012).

The presence of these body fluids is useful in reconstruction of the crime scene and tells about the sequence of events which occurred. For example, seminal or vaginal fluid could indicate the possibility of sexual activity, presence of blood stains could indicate murder, or any physical struggle, saliva stains could be the indicative of bite marks, burglary etc. Over the years, numerous methods such as chemical or enzymatic tests, protein catalytic activity tests, immunological or serological tests, biochemical assay, spectroscopic methods, and microscopy have been developed for identification of forensically relevant body fluids. These tests

have been classified into two categories such as presumptive and confirmatory test (Virkler and Lednev 2009).

Historically, tests based upon chemical or enzymatic methods were often considered as presumptive, whereas those based upon microscopic or immunological assay were confirmatory in nature (An et al. 2012). However, these methods have several limitations like lower sensitivity, reduced specificity, and less stability (Virkler and Lednev 2009). Body fluid samples are likely to have DNA material that is essential in identification of the perpetrator and connect them to the victim and to the crime scene. These presumptive and confirmatory tests also interfere with the cellular material present inside the samples and usually have incompatibility with downstream DNA analysis. Several presumptive tests are available for the body fluids, but very few of them have the confirmatory test (Harbison and Fleming 2016). Most of these confirmatory tests are performed in the laboratory and not at the scene of crime. If any presumptive tests can be performed on field, there are high chances of getting false-positive or false-negative result for the body fluids. This could result in the rejection of evidence and create false interpretation. Therefore, these conventional methods are not suitable for positive identification of certain biological fluids in forensics. Due to the limitations, their interpretation must be done very carefully, and the positive confirmation of a certain body fluid requires further testing, expect the spermatozoa detection via microscopy.

One of the major challenges faced with these tests are destruction of the samples. These are considered as destructive methods.In some cases, where small amount of biological evidence are present at the crime scene, it is highly recommended that they are examined in the most efficient way possible by nondestructive methods (Powers and Lloyd 2007). This is because all biological fluids contain DNA evidence, and with nondestructive methods, the DNA evidence can be preserved for the conclusive finding. Therefore, it is necessary to develop identification tests that will preserve this valuable data. The other limitations of these present methods are that specific test can only detect the specific body fluid, so the examiner needs to decide the test based on the fluid likely to be present there. Thus, universal confirmatory methods are required that could be able to identify any kind of body fluids to an unknown stain encountered from the crime scene.

During the last years, various methods have been approached for the investigation of these body fluids and have much broader aspect than the conventional methods (as shown in Table 1). They are mostly recognized as presence of specific markers, namely, expression of messenger RNAs (mRNAs), microRNAs (miRNAs), and detection of differential DNA methylation levels. These molecular markers are being reported as alternative methods to body fluid identification in the last decade (Sijen 2015).

To keep pace with recent improvements in the discriminating power of DNA typing, precise body fluid identification is required now more than ever. In this paper, we discussed the current and emerging methods for identifying forensically relevant body fluids.

Table 1 Different methods for the identification of body fluids

Body fluid	Enzymatic/ chemical	Serological	Microscopic	Molecular biological
Saliva	Colorimetry Phadebas® SALIgAE® Blue starch agarose plate method	Immunodiffusion Immunoelectrophoresis ELISA Immunochromatography RSIDTM-Saliva		mRNA microRNA DNA methylation
Semen	Acid phosphatase test	Immunodiffusion Immunoelectrophoresis Immunochromatography SPERM HY-LITERTM SERATEC®PSA Semiquant RSIDTM-Semen	Baecchi staining Corin-Stockis staining Oppitz staining Hematoxylin and eosin staining	mRNA microRNA DNA methylation
Vaginal fluid		Immunodiffusion Immunoelectrophoresis	Papanicolaou staining Lugol's staining	mRNA microRNA DNA methylation
Blood	Kastle-Meyer test Benzidine test Leucomalachite Green test Fluorescein test Colorimetry test Hemastix®	Immunoelectrophoresis Immunological (antihemoglobin test) Rapid immunoassay	Teichmann test Takayama test Wagennar test	mRNA microRNA DNA methylation
Urine	Nessler's DMAC Urease/ bromothymol blue Jaffe test Salkowski test	Immunological Sandwich ELISA SP radioimmunoassay	Visualize cells (no stains)	mRNA microRNA DNA methylation

Biological Background of Molecular Markers

mRNA Profiling

To overcome the limitations of conventional body fluid identification method, the incorporation of various messenger RNA (mRNA) markers was suggested for their identification purposes (Bauer 2007). There are total 210 distinct cells types in the adult body, and each one of them is specialized to perform different functions (Milo et al. 2010). Despite the different cell type, their genome material will be the same with exception in haploid and anucleated cells. However, their physical characteristics such as shape, size, and metabolic activity vary from each type. In the stage of embryonic development, the formation of cell type is highly controlled and directed by the mRNA. Terminally differentiated cells, whether they comprise blood, epithelial cells lining the oral cavity, ejaculated spermatozoa, or epidermal cells from the

skin, become such during a developmentally regulated program in which certain genes are turned off (i.e., transcriptionally silent) and turned on (i.e., are actively transcribed and translated into protein) (Ballantyne et al. 2018).

Hence, this generates a unique pattern of gene expression for each cell type and could be indicated with the presence of large amount of mRNA transcripts. The determination of type and abundance of targeted mRNAs allows a positive identification of the body fluids encountered at the crime scene. This is how the principle of mRNA profiling methods is utilized in identification of body fluids. RNA profiling also has the potential to discriminate the forensically relevant body fluids and tissues utilizing process compatible with the contemporary DNA analysis (Zubakov et al. 2008). The European DNA Profiling Group (EDNAP) performed collaborative exercises and tested on mRNA profiling analysis for body fluid such as blood, seminal fluid, saliva, and menstrual and vaginal fluid (CVF) and supports that mRNA/DNA co-extraction analysis could be utilized for positive identification of biological fluid in routine casework (Haas et al. 2014).

So far, three mRNA-based identification methods such as capillary electrophoresis (CE), quantitative RT-PCR (qRT-PCR), and high-resolution melt (HRM) have been developed and approached (Hanson and Ballantyne 2013). The specificity and sensitivity of the conventional serological methods were addressed with these developments but raised new concerns like isolation of mRNA, digestion of genomic DNA contaminant, synthesis of cDNA with reverse transcription, amplification, and separation and identification of the fragments. These concerns were achieved with the development of capillary electrophoresis and high-resolution melting methods (Conti and Buel 2012). A novel-based mRNA multiplex system was developed that has ability to identify all the possible body fluid mixtures (Xu et al. 2014). The mRNA-based biological biomarkers are considered specific, able to differentiate between menstrual blood and vaginal secretions. Studies have been conducted that showed the detection of semen with or without spermatozoa through mRNA markers (Moawad et al. 2018). These mRNA biomarkers have been used routinely utilized in the forensics for identification of body fluids. However, they also have some limitations like RNA instability or more prone to degradation, expensive primers and probes, detrimental environmental conditions such as heat and humidity, and unavailability of commercial kits (Hanson et al. 2018). Due to instability of mRNA molecules, they might not be detectable in some actual forensic case samples. Also, the concentrations of RNA in forensic samples are low which makes mRNA analysis method more challenging in identification of body fluids. These limitations of mRNA make them unsuitable for forensic purposes and less recommended for forensically relevant body fluids. Several researchers have confirmed the stability of RNA in dried forensic stains and, hence, could be successfully utilized in aged biological stain samples (Setzer et al. 2008).

Detection of mRNA transcripts using RT-PCR methods has not shown consistent results or sometimes remains undetected. Hence, these have been improved with introduction of next-generation sequencing. With the advent and evolution of NGS, targeted multiplexed RNA sequencing assay could be possible, due to incorporation of multiple biomarkers for each biological fluid (blood, semen, saliva, vaginal

secretions, menstrual blood, and skin) in a single targeted assay (Hanson et al. 2018). Despite the success of multiplex NGS RNA assay for degraded samples of blood, semen, saliva, vaginal material, menstrual fluid, and skin, they are not likely to be used for typical forensic casework sample. Therefore, presently there is no clear solution to this problem except to avoid analyzing degraded RNA. Further, it is hypothesized that miRNAs might be more suitable for the analysis of challenging and degraded forensic samples due to their small (18 to 24 nucleotides) size and the stabilizing effect of Argonaute proteins (Rocchi et al. 2021).

miRNAs Profiling

In the last 10 years, miRNA profiling has gained popularity within forensic community and has been given consideration for body fluid identification. They are noncoding small RNA, typically size of 20–25 nucleotides in length (Lagos-Quintana et al. 2001). Due to this, miRNAs have higher stability and are less prone to RNase degradation. Extracellular miRNAs are found in exosomes and microvesicles, although some miRNAs are also binded with proteins, especially Argonautes (Sohel 2016). Research has demonstrated that they detect specific body fluid expression profiles for each cell type. Hence, these advantages of miRNA make them a potential biomarker in identification of forensically relevant body fluids. In forensics, they have been extensively studied in five body fluids such as blood, menstrual blood, semen, vaginal secretion, and saliva. miRNAs are small regulatory RNAs that regulates the gene expression at posttranscriptional level. There have been various studies which have also demonstrated several miRNAs as being specific for tissue expression pattern (Wang et al. 2013).

Currently, the miRNA database has ~38,589 entries of which 1,917 are registered as human (www.mirbase.org). In a single miRNA, multiple mRNA targets could be present followed by multiple miRNAs at any given target; therefore, a key limitation is their specificity, for both body fluid and species. At first, several studies have been demonstrated to discover miRNAs that target specific body fluid, but none of them were successful so far. Hanson and colleagues (2009) conducted first forensic-based study to report miRNA profiling and their expression in forensically relevant body fluids. Using qRT-PCR analysis (miScript SYBR Green PCR kit, Qiagen), 452 human miRNAs in semen, saliva, blood, vaginal secretions, and menstrual blood were explored. A total of nine miRNAs markers such as miR451, miR16, miR135b, miR10b, miR658, miR205, miR124a, miR372, and miR412 were characterized and differentially expressed in these body fluids. They were also able to detect these body fluids using as little as 50 pg of total RNA.

In 2010, Zubakov et al. also examined the expression of 718 human miRNAs in forensic samples using comprehensive microarray platform. The results recognized 14 specific miRNA markers for each body fluid. Later on, subsequent validation using reverse transcription quantitative PCR (RT-qPCR) suggested only blood and semen miRNA candidates for future forensic applications. Another research was conducted by Seashols-Williams et al. (2016) that highlighted the identification of

forensically relevant body fluids with miRNA markers using next-generation sequencing (NGS) platform. They generated a large data set, with the subsequent development of candidate miRNAs for further research. Using their findings some more panels of miRNA were identified in six forensically relevant bodies with the help of whole miRNome massively parallel sequencing. Although several studies have been conducted so far highlighting the importance of miRNA profiling for forensic samples, their outcomes were not consistent when different methods and technology were applied. Further, it has been also suggested that more panels of miRNA are required for each of the forensically biological samples. In addition to this, some body fluids such as vaginal fluid and menstrual blood are being less explored with miRNA profiling (Glynn 2020). To implement this method into forensic applications, rigorous methodological validation and accurate models are needed.

Epigenetics

Epigenetics is the study of various reversible modifications to the genome that produce heritable changes without changing the underlying DNA sequence. One such common epigenetic modification is known as DNA methylation that involves the addition of a methyl group to the cytosine in DNA (Nustad et al. 2018). The most potential sites for DNA methylation in mammals are CpG dinucleotides (Ghosh et al. 2010); however, few other methylation sites such as CpA, CpT, and CpC have also been reported (Pinney 2014). Depending upon the cell type and chromosomal position, these CpG dinucleotides are found in three different states such as methylated, semimethylated, and unmethylated (Reinius et al. 2012; Tammen et al. 2013; Kader and Ghai 2017).

Higher density of CpGs is seen in exons, while introns and intergenic sequences have lesser amount of CpGs (Bestor et al. 2015). Majority of the CpGs are methylated in mammals, and unmethylated stretches of CpG dinucleotides known as CpG islands (CGIs) also exist in methylated states. CGIs are typically located in promoter regions and control gene expression via DNA differentiated methylation levels (Lövkvist et al. 2016). In humans, the sizes of CpG islands (CGIs) are usually 300–3000 base pairs long with a CpG content of above 50%. In mammals, DNA methylation contributes to embryonic development and participates in various epigenetic regulation such as genomic imprinting, gene regulation, X chromosome inactivation, aging, and genome stability (Ghosh et al. 2010; Lalruatfela 2013). Any abnormalities or disruption in these DNA methylation process leads to the various diseases such as genetic disorders, cancers, psychological or neurological disorder, and autoimmune diseases (Robertson 2005).

Epigenetic modification of the epigenome is established during early embryonic development and is maintained during cell division. During the period of embryonic development, methylation patterns of the germline and somatic cell lines are established. The dynamics of the DNA methylation pattern changes once the fertilization occurs, and hence these established patterns are the collective result of

establishment and maintenance of methylation during embryonic development. It has been also established that DNA methylation patterns can be stored or inherited and might influence the cell fate as well as overall phenotype. Initially, the egg released during ovulation carries the maternal DNA methylation pattern. The sperm that will fertilize with the egg carries the paternal DNA methylation pattern. During the early embryonic development, the level of DNA methylation could be increasing or decreasing. The female egg and male sperm carry their own DNA methylation pattern. During fertilization, a zygote is formed, and the maternal and paternal genome and epigenome are contained in one cell. The paternal epigenome then rapidly undergoes genome-wide active DNA demethylation and remains demethylated following multiple rounds of cell division (Schmitz et al. 2011; Messerschmidt et al. 2014; Stricker and Götz 2018).

The maternal epigenome undergoes gradual passive demethylation. By the morula stage, the methylation pattern of the embryo is essentially erased. At the blastocyst stage, de novo methylation is initiated. De novo methylation is where new methylation patterns are re-established in growing embryo. During differentiation, each cell type acquires a unique DNA methylation pattern that is subsequently maintained during each cell division, thereby resulting in specific methylomes which may enable prediction of cell type. Although the DNA methylation patterns are relatively stable, changes in methylation can potentially occur during early development due to environmental influences (Christensen et al. 2009). For example, the mother environment exposure to smoke, alcohol, nutrition, and drugs may lead to random epimutations in developing embryo and even in the germ cells of the developing embryo. Besides these environmental factors, changes in DNA methylation also occur due to genetic variants such as single-nucleotide polymorphisms (SNPs) (Van Dongen et al. 2016). DNA methylation plays an important role in development and differentiation by controlling gene expression through changes in chromatin structure. Thus, regulating gene expression during normal development and cellular differentiation or proliferation is very crucial.

Majority of the CGIs display low methylation; however, small percentage does obtain methylation during development. Also, a large number of CGIs are differentially methylated across tissues and cell types (Illingworth and Bird 2009; Mendizabal and Yi 2016). DNA methylation is found in CGI shelves (regions 2–4 kb from CGIs), CGI shores (regions 0–2 kb from CGIs), and other open sea regions (CpG sites harbored in the genome randomly). CGI with flanking region exhibits stable methylation and takes part in gene regulation. Numerous individual CpG sites and genomic regions have been identified by epigenome-wide association studies (EWAS) that show different methylation patterns between tissue and body fluids. These are called differentially methylated sites (DMSs) and tissue-specific differentially methylated regions (tDMRs), respectively (Eckhardt et al. 2006; Lokk et al. 2014). tDMRs are primarily observed at the boundaries of CpG islands, and both CpG and G/C content are lower than that of surrounding regions. tDMRs afford cells with an epigenetic memory by generating cell type-specific hypo- and hypermethylation patterns (Igarashi et al. 2008; Kader and Ghai 2015). Even though established role of tDMRs is not fully known, these differentially methylated

regions are thought to function by either attracting or preventing the binding of specific factors in a methyl- dependent manner. Studies have showed major links between gene silencing and tDMRs (Ohgane et al. 2008; Cohen et al. 2011). Based on differential methylation pattern, DMSs and tDMRs have tendency to distinguish between tissues and fluids.

DNA Methylation Profiling Method

The collection and successive analysis of human body fluids and tissue from crime scene are one of the imperative steps of a forensic investigation. The transfer of various body fluids from human body to objects may direct and lead to reveal events of physical confrontation. Similarly, events of sexual assaults can be revealed by studying dynamics of transfer of semen, vaginal fluid, and menstrual blood. Body fluids can help in the reconstruction of crime scene and linking suspects to crime scene. Identification of body fluids along with linkage to suspects is a challenging task. They may be contaminated, mixed with body fluid of another individual, present in trace quantity which is insufficient for any test or undergone degradation due to environmental factors, all of which limits the precise identification (Kayser and De Knijff 2011; An et al. 2012; Weyermann and Ribaux 2012).

In addition to variations at the mRNA and miRNA level, an alternative method that utilizes tissue-specific differential DNA methylation was introduced for forensically relevant biological fluids. A summary of various body fluid-specific microRNA markers and tissue differentially methylated regions are shown in Table 2. Interestingly, there have been several studies conducted that have reported variations at tissue-specific DNA methylation and gene expression (An et al. 2012; Sijen 2015). In fact, tissue of origin is the primary difference in DNA methylation profiles from different samples, regardless of their origin from the same or different individuals (Byun et al. 2009; Jiang et al. 2015). Whole epigenome profiling of DNA methylation on different tissues indicated several tDMRs in the mammalian genome. In addition to this, DNA methylation levels were more consistent between the same tissues from different individuals than between different tissues from the same individual (Lee et al. 2016). The possibilities of tissue specificity DNA methylation have been intended for examination of different biological body fluids frequently encountered in the crime scenes. This process takes place in the epigenome, the multitude of chemical compounds that surrounds the DNA and affects the expression of genes within the genome. The major epigenetic modifications involve DNA methylation, histone modification, and chromatin structure. Among all the three modifications, DNA methylation epigenetic marks stand in extracted DNA. This characteristic makes the DNA methylation marks best compatible with standard forensic procedures. DNA methylation that occurs at CpG dinucleotides-5′- position of the pyrimidine ring of cytosine residues inhibit gene expression by affecting chromatin structure. Depending upon the cell types, different methylation patterns and chromosome segments exist and known as tissue-specific differentially methylated regions (tDMRs). These tDMRs show different DNA methylation profiles

Table 2 Summary of body fluid-specific microRNA markers and tissue differentially methylated regions in literature

miRNA marker	Targeted body fluid	Other body fluids detected	DNA marker/ CpG ID	Body fluid	Technique
miR-26a	Saliva	N/A	cg06379435	Blood	Bisulfite pyrosequencing
miR-96	Saliva	N/A	cg08792630	Blood	Bisulfite pyrosequencing
miR-1	Saliva	Menstrual blood, vaginal	BL1/ cg06378435	Blood	Multiplex methylation SNaPshot
miR-135b	Saliva	N/A	BL2/ cg01543184	Blood	Multiplex methylation SNaPshot
miR-138	Saliva	N/A	BL-I/ cg04011671	Blood	10-plex MSRE-PCR
miR-141	Saliva	N/A	BL-II/ cg18454288	Blood	10-plex MSRE-PCR
miR-145	Saliva	N/A	cg23521140	Semen	Bisulfite pyrosequencing
miR-182	Saliva	N/A	cg17610929	Semen	Bisulfite pyrosequencing
miR-200c	Saliva	N/A	SE1/ cg17610929	Semen	Multiplex methylation SNaPshot
miR-203	Saliva	Vaginal secretions	SE2/ cg26763284	Semen	Multiplex methylation SNaPshot
miR-203a	Saliva	Vaginal secretions, semen	SE3/ cg17621389	Semen	Multiplex methylation SNaPshot
miR-205	Saliva	Menstrual blood, semen, vaginal secretions	SE-I/ cg05261336	Semen	10-plex MSRE-PCR
miR-208b	Saliva	N/A	SE-II/ cg07485723	Semen	10-plex MSRE-PCR
miR-381	Saliva	N/A	cg26107890	Saliva	Bisulfite pyrosequencing
miR-431	Saliva	N/A	cg181421427	Saliva	Bisulfite pyrosequencing
miR-450b-5p	Saliva	N/A	SA1/ cg09652652	Saliva	Multiplex methylation SNaPshot
miR-518c	Saliva	N/A	SA-I/ cg09652652	Saliva	10-plex MSRE-PCR
miR-583	Saliva	N/A	SA-II/ cg09107912	Saliva	10-plex MSRE-PCR

(continued)

Table 2 (continued)

miRNA marker	Targeted body fluid	Other body fluids detected	DNA marker/ CpG ID	Body fluid	Technique
miR-622	Saliva	N/A	MB/ cg18069290	Menstrual blood	10-plex MSRE-PCR
miR-658	Saliva	N/A	cg01774894	Vaginal secretions	Bisulfite pyrosequencing
miR-1228	Saliva	N/A	cg14991487	Vaginal secretions	Bisulfite pyrosequencing
miR-16	Venous blood	Menstrual blood, semen, saliva, skin, vaginal secretions	VF1/ 09765089	Vaginal secretions	Multiplex methylation SNaPshot
miR-16a	Venous blood	Semen, vaginal secretions	VF2/ 26079753	Vaginal secretions	Multiplex methylation SNaPshot
miR-20a	Venous blood	N/A	VG/ cg15402210	Vaginal secretions	10-plex MSRE-PCR
miR-106a	Venous blood	N/A	–	–	–
miR-126	Venous blood	N/A	–	–	–
miR-142	Venous blood	Semen	–	–	–
miR-142-3p	Venous blood	Saliva, semen, vaginal secretions	–	–	–
miR-150	Venous blood	N/A	–	–	–
miR-200b-3p	Venous blood	N/A	–	–	–
miR-412	Venous blood	Menstrual blood	–	–	–
miR-451	Venous blood	Menstrual blood, saliva	–	–	–
miR-451a	Venous blood	Menstrual blood	–	–	–
miR-486	Venous blood	N/A	–	–	–
miR-10a	Semen	Vaginal secretions	–	–	–
miR-10b	Semen	Menstrual blood	–	–	–
miR-17	Semen	N/A	–	–	–

according to the cell or tissue type. Thus, determination of DNA methylation profiles at a specific CpG site of tDMR would help in identification of the cell type of forensically biological DNA samples (Frumkin et al. 2011; Forat et al. 2016).

In 2011, Frumkin and colleagues reported the first introduction of DNA methylation profiling to forensic science and evaluated the DNA methylation expression in forensically relevant biological fluids. They reported 15 differentially methylated genomic loci in samples such as saliva, semen, blood, urine, vaginal secretion, and skin epidermis using methylation- sensitive restriction enzyme-PCR (MSRE-PCR) method. These restriction enzymes can easily digest the DNA samples followed by amplification of specified loci of genomic DNA using fluorescently labeled primers, capillary electrophoresis, and automatic signal detection method. Hence, this procedure can be successfully included in forensic laboratories like other techniques such as SniPS (SNPs) and STRs which would positively determine the source tissues in 50 DNA samples from blood, saliva, semen, and skin epidermis.

Using the previously reported approach, Wasserstrom et al. (2013) advanced his study by developing a kit known as DNA source identifier (DSI)-SemenTM. This assay was designed to replace the microscopic examination of spermatozoa for semen identification in forensic casework samples. Five genomic loci were selected to detect the semen-specific DNA methylation patterns using MSRE-PCR method. A validation study of (DSI)-SemenTM kit was also conducted with 135 samples of different body fluids and 33 real-time forensic casework samples. Hence, the results were reliable and showed positive result for semen and also established their competences to detect semen utilizing as little as 31 pg of template DNA input.

Later, Lee et al. (2012) examined the potential of five previously examined tDMRs for blood, semen, saliva, menstrual fluid, and vaginal secretion using bisulfite sequencing approach. They proposed two tDMRs: two for semen-specific marker and other for vaginal fluid specific marker. The tDRMs for USP49 and DACT1 genes are considered as a semen-specific marker by exhibiting semen-distinctive hypomethylation, and the PFN3 tDMR was recognized as a vaginal fluid-specific marker. Madi and colleagues conducted the similar study and explained tissue-specific DNA methylation in forensic samples such as blood, semen, saliva, and epithelial cells. They examined the methylation arrays at the ZC3H12D and FGF7 loci using bisulfite modification and pyrosequencing approach. A very few genomic loci were examined to find out whether methylation pattern at ZC3H12D and FGF7 loci has potential to differentiate sperm from other forensic samples or not. However, locus C20orf117 and the BCAS4 can distinguish blood and saliva samples from other biological samples, separately. The outcome of these studies also indicates that the DNA methylation approach could be used as a valuable tool for identification of forensically relevant body fluid samples; however further validation studies including more markers will be required for actual casework applications.

The recent advancements in epigenetics have enabled genome-wide DNA methylation profiling analysis. DNA methylation was assessed with the Illumina's Human Methylation450 (450 K) BeadChip method that measured methylation at $\geq$450,000 CpG sites. Using 450 K BeadChip array method, researchers have also examined some CpG markers that displayed differential DNA methylation patterns in different body fluids. Another similar study was designed by Park and colleagues with Human Methylation 450 K bead array to report differentially

methylated DNA among blood, saliva, semen, and vaginal secretions samples. The authors reported eight CpG sites (two CpG sites, cg06379435 and cg08792630, for blood; two sites, cg26107890 and cg20691722, for saliva; two cg23521140 and cg17610929 for semen; and two sites, cg01774894 and cg14991487, for vaginal secretions) in these body fluids. Later, these body fluid-specific DNA methylation markers were also validated using pyrosequencing analysis. This study showed large methylation differences between blood and semen markers; however, their saliva and vaginal secretion markers showed small or inter-individual differences from other body fluids.

Another study was conducted by Lee et al. (2015) using the Illumina Infinium Human Methylation-450 BeadChip array to identify differentially methylated DNA profiles among 42 body samples, i.e., saliva, blood, vaginal fluid, semen, and menstrual blood. The authors selected the 64 potential body fluid-specific CpG sites. The group also examined these CpG sites in 151 more body fluids using bisulfite sequencing and identified the subset of CpG markers. Next, a multiplex methylation SNaPshot (methylation-sensitive single-nucleotide primer extension, MS-SNuPE) reaction was developed that indicated methylation status of eight body fluid-specific CpG sites (cg17610929, cg26763284, and cg17621389 for semen; cg06379435 and cg01543184 for blood; cg09765089-231d and cg26079753-7d for vaginal fluid; and cg09652652-2d for saliva). However, upon testing only one CpG site (cg17621389) showed potential as semen- specific hypo-methylation, while seven other candidates showed target body fluid-specific hyper-methylation. The multiplex developed in this study could positively distinguish between blood, saliva, semen, and the body fluid which originates from female reproductive organ. Previously conducted studies showed cg17610929 as a semen-specific marker and cg06379435 as a blood-specific marker.

Forat et al. (2016) identified potential body fluid-specific CpG sites between blood, menstrual blood, saliva, vaginal fluid, and semen using HumanMethylation27 (27 K) and 450 K BeadChip arrays. Both bisulfite sequencing and MS-SNuPE were used to analyze methylation levels of the markers and reported two CpG sites (cg26285698 and cg03363565) for blood; one (cg09696411) for menstrual blood; two (cg21597595 and cg15227982) for saliva; two (cg14991487 and cg03874199) for vaginal fluid; and two (cg22407458 and cg05656364) for semen. Studies conducted so far have suggested that molecular genetics-based methods using DNA methylation profiling could be a valuable tool for identification of body fluids, but further validation studies including more markers will be required for better applications to forensic caseworks.

DNA-Compatible Cell-Specific Identification

An alternative method for forensically relevant body fluid identification also involves classification of individual cells by labeling them at the protein, DNA, and RNA level coupled with microscopy. Immunohistochemistry technique has been developed to identify epidermal cells and distinguishing the vaginal and mucosal

cells (buccal and vaginal) using cytokeratin (Paterson et al. 2006). Mucosal cells could be directly distinguished from epidermal cells, since low expression of each cytokeratin was found in the other type of cell. In a study reported by Thorogate et al. (2008), immunofluorescent technique was introduced to spatially recognize on- site human blood. This method used fluorescently tagged monoclonal mouse antihuman antibodies and detect one erythrocyte (glycophorin A) and three nucleated leukocytes (*CD45*, myeloperoxidase, histone 1). The multiple washing steps made this method impractical for forensic caseworks. The application of XY fluorescent in situ hybridization (FISH) method coupled with laser microdissection (LMD) has also been explored for body fluid identification. They have been used to determine the male or female origin of epithelial cells and subsequent recovery of DNA profiling from target cells by labeling them with specific fluorescent (Lynch et al. 2015).

ESR1 (estrogen receptor 1) was detected using fluorescently labeled monoclonal antibodies and showed its ability to distinguish between buccal and vaginal epithelial cells when mRNA profiling failed to do so. This suggests that mRNA expression and tissue-specific protein expression do not necessarily correlate. RNA suspension-fluorescent in situ hybridization is being used to characterize and detect the epithelial cells with the help of fluorescently labeled LNA probe for keratin 10. Next, epithelial cells are extracted using LMD and then profiled with DNA. Even though these labeling methods might be specific, prolonged time span taken to selectively gather the labeled cells by LMD highlights that these techniques are likely to be restricted to targeted cases rather than for general applications (Fleming et al. 2013; Williams et al. 2014).

Conclusion

Even today, majority of the forensic laboratories used conventional methods to identify body fluids, while current forensic DNA profiling based on STRs allows personal identification of biological sample donors. With the time and advancements, several other techniques have been developed in order to add more informative layers to the forensically relevant body fluid samples. These emerging techniques are primarily based on molecular genetic approaches such as mRNA and miRNA profiling and DNA methylation. In this chapter, all the three emerging techniques have been discussed and explained their advantages and disadvantages. mRNA markers have been most rigorously investigated, and the number of specific markers is enough for the identification of forensically relevant body fluids. Despite having high tissue specificity and sensitivity, mRNA markers are not considered for body fluid identification. mRNAs are easily prone to heat and humidity and does not allow simultaneous analysis of multiple markers specific for various tissues in a single multiplex system. DNA methylation technique has been considered as a valuable tool for forensic body fluid identification, because the same DNA samples are utilized for methylation profiling as well as DNA typing. DNA methylation-based body fluid identification has many advantages when compared to mRNA-based methods, such as higher specificity, the ability to be inserted into current

forensic DNA-based testing protocols, and the applicability to old cases where only DNA extracts are available. It also allows the simultaneous analysis of multiple markers specific for various tissues in a single multiplex system. We believe that it is now possible to identify more informative markers for body fluid identification using various high-throughput screening technologies. To incorporate DNA methylation profiling into forensic caseworks, a greater number of markers need to be identified. With the advancements in technology, more informative markers for body fluid identification could be achieved. Soon, forensic community should improve a great deal because of the continued advances in genetics, epigenetics, and molecular biology, and hence, the extraction of more information from forensically relevant biological samples will be possible.

References

An JH, Shin K-J, Yang WI, Lee HY (2012) Body fluid identification in forensics. BMB Rep 45:545

Ballantyne J, University of Central Florida, United States of America (2018) Identification of forensically relevant fluids and tissues by small RNA profiling

Bauer M (2007) RNA in forensic science. Forensic Sci Int 1(1):69–74

Bestor TH, Edwards JR, Boulard M (2015) Notes on the role of dynamic DNA methylation in mammalian development. Proc Natl Acad Sci 112(22):6796–6799

Byun H-M, Siegmund KD, Pan F, Weisenberger DJ, Kanel G, Laird PW, Yang AS (2009) Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. Hum Mol Genet 18(24):4808–4817

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS Genet 5(8):e1000602

Cohen NM, Kenigsberg E, Tanay A (2011) Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell 145(5):773–786

Conti T, Buel E (2012) Forensic stain identification by RT-PCR analysis. BiblioGov

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38(12):1378–1385

Fleming R, Harbison S, Lin M-H (2013) New RNA methods for the identification of body fluids and cell types. Forensic Sci Int 4(1):e87–e88

Forat S, Huettel B, Reinhardt R, Fimmers R, Haidl G, Denschlag D, Olek K (2016) Methylation markers for the identification of body fluids and tissues from forensic trace evidence. PLoS One 11(2):e0147973

Frumkin D, Wasserstrom A, Budowle B, Davidson A (2011) DNA methylation-based forensic tissue identification. Forensic Sci Int Genet 5(5):517–524

Ghosh S, Yates AJ, Frühwald MC, Miecznikowski JC, Plass C, Smiraglia D (2010) Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. Epigenetics 5(6):527–538

Glynn CL (2020) Potential applications of microRNA profiling to forensic investigations. RNA 26(1):1–9

Haas C, Hanson E, Anjos M, Ballantyne K, Banemann R, Bhoelai B, Borges E, Carvalho M, Courts C, De Cock G (2014) RNA/DNA co-analysis from human menstrual blood and vaginal secretion stains: results of a fourth and fifth collaborative EDNAP exercise. Forensic Sci Int Genet 8(1):203–212

Hanson EK, Ballantyne J (2013) Rapid and inexpensive body fluid identification by RNA profiling-based multiplex High Resolution Melt (HRM) analysis. F1000Research 2:281

Hanson E, Ingold S, Haas C, Ballantyne J (2018) Messenger RNA biomarker signatures for forensic body fluid identification revealed by targeted RNA sequencing. Forensic Sci Int Genet 34: 206–221

Harbison S, Fleming R (2016) Forensic body fluid identification: state of the art. Res Rep Forensic Med Sci 6:11–23

Igarashi J, Muroi S, Kawashima H, Wang X, Shinojima Y, Kitamura E, Oinuma T, Nemoto N, Song F, Ghosh S (2008) Quantitative analysis of human tissue-specific differences in methylation. Biochem Biophys Res Commun 376(4):658–664

Illingworth RS, Bird AP (2009) CpG islands – 'a rough guide'. FEBS Lett 583(11):1713–1720

Jiang R, Jones MJ, Chen E, Neumann SM, Fraser HB, Miller GE, Kobor MS (2015) Discordance of DNA methylation variance between two accessible human tissues. Sci Rep 5(1):1–8

Kader F, Ghai M (2015) DNA methylation and application in forensic sciences. Forensic Sci Int 249:255–265

Kader F, Ghai M (2017) DNA methylation-based variation between human populations. Mol Gen Genomics 292(1):5–35

Kayser M, De Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12(3):179–192

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science 294(5543):853–858

Lalruatfela B (2013) On DNA methylation: an introductory review. Sci Vision 13:1–7

Lee HY, Park MJ, Choi A, An JH, Yang WI, Shin K-J (2012) Potential forensic application of DNA methylation profiling to body fluid identification. Int J Legal Med 126(1):55–62

Lee HY, Lee SD, Shin K-J (2016) Forensic DNA methylation profiling from evidence material for investigative leads. BMB Rep 49(7):359

Lokk K, Modhukur V, Rajashekar B, Märtens K, Mägi R, Kolde R, Koltšina M, Nilsson TK, Vilo J, Salumets A (2014) DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol 15(4):1–14

Lövkvist C, Dodd IB, Sneppen K, Haerter JO (2016) DNA methylation in human epigenomes depends on local topology of CpG sites. Nucleic Acids Res 44(11):5123–5132

Lynch L, Gamblin A, Vintiner S, Simons JL (2015) STR profiling of epithelial cells identified by X/Y-FISH labelling and laser microdissection using standard and elevated PCR conditions. Forensic Sci Int Genet 16:1–7

Mendizabal I, Yi SV (2016) Whole-genome bisulfite sequencing maps from multiple human tissues reveal novel CpG islands associated with tissue-specific regulation. Hum Mol Genet 25(1): 69–82

Messerschmidt DM, Knowles BB, Solter D (2014) DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev 28(8):812–828

Milo R, Jorgensen P, Moran U, Weber G, Springer M (2010) BioNumbers – the database of key numbers in molecular and cell biology. Nucleic Acids Res 38(suppl_1):D750–D753

Moawad AM, Abdelsalam MH, Abdelfattah A, Sabry D, Atef A (2018) Differential genes expression biomarkers for menstrual and peripheral blood stains analysis. Egypt J Forensic Sci 8(1):1–6

Nustad HE, Almeida M, Canty AJ, LeBlanc M, Page CM, Melton PE (2018) Epigenetics, heritability and longitudinal analysis. BMC Genet 19(1):99–108

Ohgane J, Yagi S, Shiota K (2008) Epigenetics: the DNA methylation profile of tissue-dependent and differentially methylated regions in cells. Placenta 29:29–35

Paterson SK, Jensen CG, Vintiner SK, McGlashan SR (2006) Immunohistochemical staining as a potential method for the identification of vaginal epithelial cells in forensic casework. J Forensic Sci 51(5):1138–1143

Pinney SE (2014) Mammalian non-CpG methylation: stem cells and beyond. Biology 3(4): 739–751

Powers LS, Lloyd CR (2007) Method and apparatus for detecting and imaging the presence of biological materials. United States patent US 7,186,990

Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlén S-E, Greco D, Söderhäll C, Scheynius A, Kere J (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PLoS One 7(7):e41361

Robertson KD (2005) DNA methylation and human disease. Nat Rev Genet 6(8):597–610

Rocchi A, Chiti E, Maiese A, Turillazzi E, Spinetti I (2021) MicroRNAs: an update of applications in forensic science. Diagnostics 11(1):32

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR (2011) Transgenerational epigenetic instability is a source of novel methylation variants. Science 334(6054):369–373

Seashols-Williams S, Lewis C, Calloway C, Peace N, Harrison A, Hayes-Nash C, Fleming S, Wu Q, Zehner ZE (2016) High-throughput miRNA sequencing and identification of biomarkers for forensically relevant biological fluids. Electrophoresis 37(21):2780–2788

Setzer M, Juusola J, Ballantyne J (2008) Recovery and stability of RNA in vaginal swabs and blood, semen, and saliva stains. J Forensic Sci 53(2):296–305

Sijen T (2015) Molecular approaches for forensic cell type identification: on mRNA, miRNA, DNA methylation and microbial markers. Forensic Sci Int Genet 18:21–32

Sohel MH (2016) Extracellular/circulating microRNAs: release mechanisms, functions and challenges. Achiev Life Sci 10(2):175–186

Stricker SH, Götz M (2018) DNA-methylation: master or slave of neural fate decisions? Front Neurosci 12:5

Tammen SA, Friso S, Choi S-W (2013) Epigenetics: the link between nature and nurture. Mol Asp Med 34(4):753–764

Thorogate R, Moreira JC, Jickells S, Miele MM, Daniel B (2008) A novel fluorescence-based method in forensic science for the detection of blood in situ. Forensic Sci Int Genet 2(4): 363–371

Van Dongen J, Nivard MG, Willemsen G, Hottenga J-J, Helmer Q, Dolan CV, Ehli EA, Davies GE, Van Iterson M, Breeze CE (2016) Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat Commun 7(1):1–13

Virkler K, Lednev IK (2009) Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. Forensic Sci Int 188(1–3):1–17

Wang Z, Zhang J, Luo H, Ye Y, Yan J, Hou Y (2013) Screening and confirmation of microRNA markers for forensic body fluid identification. Forensic Sci Int Genet 7(1):116–123

Wasserstrom A, Frumkin D, Davidson A, Shpitzen M, Herman Y, Gafny R (2013) Demonstration of DSI-semen – a novel DNA methylation-based forensic semen identification assay. Forensic Sci Int Genet 7(1):136–142

Weyermann C, Ribaux O (2012) Situating forensic traces in time. Sci Justice 52(2):68–75

Williams E, Lin M-H, Harbison S, Fleming R (2014) The development of a method of suspension RNA-FISH for forensically relevant epithelial cells using LNA probes. Forensic Sci Int Genet 9: 85–92

Xu Y, Xie J, Cao Y, Zhou H, Ping Y, Chen L, Gu L, Hu W, Bi G, Ge J (2014) Development of highly sensitive and specific mRNA multiplex system (XCYR1) for forensic human body fluids and tissues identification. PLoS One 9(7):e100123

Zubakov D, Hanekamp E, Kokshoorn M, van IJcken W, Kayser M (2008) Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. Int J Legal Med 122(2):135–142

Lee HY, An JH, Jung S-E, Oh YN, Lee EY, Choi A, Yang WI, Shin K-J (2015) Genome-wide methylation profiling and a multiplex construction for the identification of body fluids using epigenetic markers. Forensic Sci Int Genet 17:17–24

Hanson EK, Lubenow H, Ballantyne J (2009) Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. Anal Biochem 387(2):303–314

Overview of Familial DNA and Forensic Phenotyping

14

Samuel D. Hodge and John Meehan

Contents

Abstract

DNA has radically transformed criminal investigations and significantly improved the operation of the criminal justice system. This genetic roadmap is continually evolving in criminal, civil, and medical applications as new DNA applications are routinely advanced.

DNA fingerprinting is a scientific technique used to establish a link between biological evidence and a person under criminal investigation. The DNA specimen is compared with the sample collected from the suspect. If the two DNA profiles match, one may presume that the evidence came from that suspect. If the two samples do not match, then the person may be excluded as the donor. The

S. D. Hodge (✉)
Temple University, Philadelphia, PA, USA

J. Meehan
Beasley School of Law at Temple University, Philadelphia, PA, USA
e-mail: tuj93141@temple.edu

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_12

primary limitation of DNA fingerprinting is that it is not 100% accurate. Mistakes can happen because of contamination, falsification, and chain of custody issues. Incorrect testing procedures may also cause false-positive or false-negative results.

DNA evidence has proven beneficial to both the prosecution and defense, but it is not free from controversy. Experts will occasionally disagree about the analysis and statistical meaning of test results, but as new forensic techniques are advanced, ethical and legal issues are frequently asserted. Two new forensic techniques, familial DNA and forensic phenotyping, have proven to be the most controversial.

Even though large quantities of data are stored in government databases, there are still times when no match is found to link DNA from a crime scene or victim with the DNA profiles in the databases. In these situations, several countries permit familial DNA searches. In conducting these types of searches, law enforcement officials can explore DNA databases for genetic information that points to a relative of the individual whom they seek to identify. When a search for a precise match to a DNA sample proves unsuccessful, a familial DNA search may reveal a partial match which signifies a blood relative such as a parent, sibling, or cousin.

Critics retort that the technique violates the constitutional protections against unreasonable searches and seizures because the technique utilizes the DNA of innocent family members. Concern is also expressed that the search will cause an unequal number of Hispanics and African-Americans to these searches because minorities have traditionally been arrested at disproportionately higher rates than other racial groups.

Forensic DNA phenotyping is publicized as providing much promise by utilizing DNA left at a crime scene to predict a suspect's eye, hair, and skin color, gender, age, and ancestry. This technique is an attempt to convert genetic DNA into a person's physical appearance. Advocates maintain that by ascertaining how genetic DNA characteristics translate into one's physical appearance, it is possible to reverse-engineer DNA into a physical profile. Some critics of forensic phenotyping assert that it can only offer a general outline of the suspect's appearance. There is also a lack of peer-reviewed studies supporting the science.

This chapter will explore the use of DNA in criminal investigations and offer a detailed examination as to why familial DNA and forensic phenotyping have proven to be so controversial.

Keywords

Deoxyribonucleic acid (DNA) · Forensic examinations · Junk DNA · Familial DNA · DNA profiling · DNA phenotyping · Racial profiling · Right to privacy

Every family has its secrets but nothing reveals the truth like DNA.
— Danielle Trussoni

The world of forensics has dramatically changed during the past several decades. Solving crimes has become almost futuristic in its approach, and long-standing techniques such as fingerprinting and blood splatter analyses seem antiquated (Forensic Outreach n.d.; Forensics Colleges n.d.). These procedures have been replaced with automated fingerprint identification technology and video spectral comparators, which allow law enforcement personnel to immediately evaluate a fingerprint found at a crime scene with a broad database and permit criminalists to analyze a piece of paper, visualize hidden notes, and "lift" indented writings (Forensics Colleges n.d.). Long established techniques like bite mark identification and hair analysis, have now been challenged as lacking a valid scientific foundation.

The use of DNA has transformed criminal investigations and significantly improved the operation of the criminal justice system (In Brief n.d.). This genetic roadmap is constantly evolving in criminal, civil, and medical applications as new DNA uses are routinely advanced (Forensic Colleges n.d.). Familial DNA and forensic phenotyping are two of the latest and more controversial developments in crime-solving applications. This chapter will explore their development, use, and the legal, social, and ethical considerations that both techniques present.

The Basics of DNA

The use of deoxyribonucleic acid (DNA) in criminal investigations was accidentally discovered in 1984 when Sir Alec Jeffreys, a British geneticist, evaluated blood samples from a group of relatives and discovered that a DNA profile could be fashioned to differentiate the genetic makeup between people (Matheson 2016). According to the Embryo Project Encyclopedia, the scientist then sets out to learn how inherited diseases are spread through families, but his work became the building block for DNA fingerprinting, and it had immediate use in forensic investigations by allowing criminalists to identify suspects based upon their genetic makeup. Prior genetic applications only permitted the police to exclude suspects and not to identify individuals.

The Building Blocks for the Human Body

DNA, or deoxyribonucleic acid, is the chemical name for the long molecule that contains the hereditary makeup of individuals and practically all other species (Genetics Home Reference 2020b). The name is derived from its construction, consisting of a sugar and phosphate backbone which have bases protruding out from its foundation. Therefore, "deoxyribo" denotes the sugar, and the nucleic acid signifies the phosphate and the bases (National Human Genome Research Institute n.d.-a). When put together, DNA provides the genetic code for an individual and encompasses the outline for constructing the proteins that are vital for the body to operate (Panneerchelvam and Norazmi 2003; Newman 2018). In *Cobey v. State,* the Maryland Court of Special Appeals noted that the nucleus of each cell contains forty-

six chromosomes collected in sets of twenty-two, plus two sex chromosomes – X for female and Y for male. The DNA of humans has approximately three billion bases, and more than 99% of them are the same in all people (Genetics Home Reference 2020b).

The structure of DNA looks like a double helix, and it resembles a twisted chain (Smith and Gordon 1997). The sides of the chain are made up of duplicate progressions of phosphate and deoxyribose sugar molecules (Smith and Gordon 1997). The Supreme Court of Maryland in *Diggs v. State*, 531 N.E.2d 461 (1988), explained that DNA is chiefly located in the nucleus of cells and is accordingly called nuclear DNA. A small segment of DNA can also be found in the mitochondria, where it is dubbed mitochondrial DNA (mtDNA) (Genetics Home Reference 2020b). The genetic elements are then contained in a coded arrangement consisting of four nitrogen parts: adenine (A), guanine (G), cytosine (C), and thymine (T) (Genetics Home Reference 2020a). Their sequencing regulates the substances available for building and maintaining an organism, very similar to the manner that the letters of the alphabet are positioned to form words and phrases (Genetics Home Reference 2020a). Every person's genetic makeup is different, but most DNA coding is the same among people (Science Learning Hub 2005).

The Use of DNA in Forensics

DNA was first applied in forensics in the late 1980s, and it has since helped solve an untold number of crimes (Cronen 2008). As noted in *Haskell v. Harris*, 669 F.3d 1039, this is not surprising since DNA evidence is deemed more dependable than many other forms of crime scene evidence, and genetic identification is deemed more precise than fingerprint detection. A forensic examination is initiated by procuring a DNA specimen from the crime scene or a victim (Chin et al. 2018). A small number of genetic markers are then identified by utilizing assembled chemical sequences called primers, which adhere to similar DNA sequences of interest in the specimen (Chin et al. 2018). A series of primers joined to the DNA sample permit the original specimen to be enlarged so that one can determine if a DNA profile is present (Chin et al. 2018).

In *Haskell v. Harris,* the court noted that after the DNA sample is sent to a laboratory for analysis, a profile is produced by analyzing 13 genetic markers labeled "junk DNA," which are not connected with any specifically identified genetic traits. The National Institute of Justice's website on analyzing DNA evidence posits that a short tandem repeat (STR) technology is then used to identify the repetitive progression of base pairs at all of the 13 markers. The differences in the number of sequences at every marker create a unique profile that may be utilized when employed for recognition purposes (Matheson 2016). The possibility that two people will have the same sequences on all 13 markers are remote.

Databases that contain a set of numbers that depict the summed-up STR repeats in each allele for a particular set of loci were soon created to store these individual sequences (Arnaud 2017). In 1998, the FBI created the most well-known database

known as the Combined DNA Index System (CODIS). The court in *Haskell* went on to note that many other databases have been created since that time. The specimen is then compared to the other DNA profiles maintained in various databanks. If the DNA does not match the sample acquired from the victim or crime scene, the suspect may be removed from the suspect pool.

DNA Fingerprinting

DNA fingerprinting is a scientific method used to find a link between biological evidence and a person under criminal investigation. The DNA specimen collected from a crime location is compared with a DNA sample from a suspect (National Human Genome Research Institute n.d.-b). If the two DNA profiles match, it may be concluded that the evidence came from that suspect. On the other hand, if the two samples do not match, then the suspect may be excluded as the donor (National Human Genome Research Institute n.d.-b).

At one time, state law restricted the retrieval and storage of DNA to only those convicted of murder and rape due to the gravity of these offenses and the probability that DNA evidence could be discovered at such crime scenes (Rothstein 2014). This restriction was short-lived as laws were later passed that permitted the collection of DNA in most criminal matters. (Debus-Sherrill and Field 2017). Databases, such as CODIS, are currently used by the government to store DNA profiles, which permit forensic units around the United States to disclose and compare genetic evidence (Debus-Sherrill and Field 2017). According to the FBI, as of May 2020, the National DNA Index (NDIS) has more than 14,287,909 offender profiles, 4,058,033 arrestee profiles, and 1,040,573 forensic profiles.

Limitations of DNA Evidence

Scientists who examine the information employed in DNA fingerprinting maintain that it is rare that the procedure, if done correctly, would link an innocent person to a crime (Kolata 1992). Therefore, the police frequently gather genetic material when an individual is believed to be a suspect in a crime (Norrgard 2008). The main limitation of DNA fingerprinting is that it is not 100% accurate. Errors can occur because of contamination, falsification, and chain of custody issues. Incorrect testing procedures may also generate false-positive or false-negative results (Gaille 2018). Because DNA fingerprinting is considered a fact-based bit of evidence, it is conceivable for false results to be misconstrued as factual outcomes (Gaille 2018).

The initial application of DNA profiling focused on a way to establish paternity (Cormier et al. 2005). Its first criminal court application occurred in 1986 when a scientist employed DNA to dispute a confession provided by a young man accused of two murders in England (Cormier et al. 2005). The biologist demonstrated that the suspect was not the perpetrator of the crime, and the real culprit was eventually

caught through the employment of DNA testing (Cormier et al. 2005). Soon thereafter, a Florida court in *Andrews v. State*, 533 So. 2d 841 (1988), was asked to consider the admissibility of "genetic fingerprint" evidence to identify the offender of a sexual battery.

As noted in *Commonwealth v. Crews, 640 A.2d 395 (1994),* the Pennsylvania Supreme Court *opined* DNA evidence is now routinely admitted in judicial proceedings as an independent "science-based identification method" to establish a person's guilt, free someone wrongfully convicted of a crime, or determine paternity (Ming et al. 1994; Norrgard 2008).

The Right to Collect DNA from Suspects

A major development in the collection of DNA evidence occurred in 2013 when the Supreme Court of the United States in *Maryland v. King,* 569 U.S. 435, upheld the right of the police to routinely collect DNA from a person arrested, but not yet convicted of a serious offense, and to enter that person's DNA profile into its databases. *King* involved an assailant who had concealed his face when he broke into a woman's home and raped her. No one could identify the criminal, but the police obtained a sample of his DNA from the victim. Six years later, King was taken into custody for threatening several individuals with a weapon. At that time, a DNA sample was obtained from King, and it was found to match the DNA from the earlier rape case. The suspect was indicted for the sexual assault, but he challenged the collection of his DNA as a violation of his Fourth Amendment rights against unlawful search and seizure.

The Supreme Court noted Maryland law permitted law enforcement to obtain DNA samples from "persons charged with violent crimes, including first-degree assault." The Supreme Court agreed with this procedure by opining that it is a key advance in the procedures long employed by law enformenet to satisfy legitimate police concerns. The *King* decision has resulted in more than half of the states enacting laws permitting "test-on-arrest" depending on variables such as "qualifying offenses, the timing for taking a sample, conditions for destroying samples, and [the] use [of samples] in court proceedings" (Strutin 2015).

Recent DNA Controversies

DNA evidence has proven beneficial to both the prosecution and defense, but it is not free from controversy (Thompson et al. 2012). Experts will occasionally disagree about the analysis and statistical meaning of test results, but as new forensic techniques are advanced, ethical and legal issues are frequently asserted (Cesar et al. 2016). Two new forensic techniques, familial DNA and forensic phenotyping have proven to be the most controversial.

Familial DNA

Despite the huge amount of information stored in various government databases, there are still times when no match is found to link DNA from a crime scene or victim with the DNA profiles in the various databases. In these situations, several countries permit familial DNA searches (FDS) (Debus-Sherrill and Field 2017). In conducting an FDS, law enforcement can search DNA databases for genetic information that points to a relative of the individual whom they seek to identify (Murphy 2013).

In those situations where the hunt for a precise match to a DNA sample proves futile, a search of familial DNA may reveal a partial match which signifies a blood relative such as a parent, sibling, or cousin (FindLaw 2019a). Familial searching is based on the idea that biologically related people such as siblings or parents will have more similar genetic information than non-biologically related people (Liberty 2015).

The United Kingdom was the first country to search for familial DNA matches and prosecute the arrested individual (Debus-Sherrill and Field 2017). With the assistance of comprehensive databases, British authorities mainly use FDS to resolve cold cases (Murray et al. 2017). For example, British authorities apprehended a 1970s serial murderer after DNA from the crime scene exposed a familial connection to the killer's son who had been convicted for theft of a vehicle (McBride 2016).

Meanwhile, the United States has been slow to adopt the technique. Recently, however, there has been a growing attentiveness in this technique to aid criminal investigations. Advocates talk about its ability to help in the identification and conviction of criminals, prevent crime, resolve cold cases, absolve improperly convicted defendants, and enhance public safety (Debus-Sherrill and Field 2017). Its use in criminal matters has continued to proliferate because of scientific advances and increasing laws that allow DNA collection from a wider array of people coming into contact with the justice system (Debus-Sherrill and Field 2017).

The FBI currently forbids searches of NDIS to discover a familial match (Debus-Sherrill and Field 2017). Several states permit familial DNA searches (Rainey 2018). More specifically, California, Colorado, Texas, and Virginia allow both familial searching and partial matching, while 19 jurisdictions sanction partial matching either through expressed permission or lack of direct prohibition (Debus-Sherrill and Field 2017). A recent analysis, however, reported that only 12 states solved criminal cases using FDS. (Rainey 2018). As of 2018, the 12 states that had done so were Arizona, California, Colorado, Florida, Minnesota, New York, Ohio, Texas, Utah, Virginia, Wisconsin, and Wyoming (Rainey 2018). Conversely, at least two jurisdictions – the District of Columbia and Maryland – have statutorily prohibited this technique (FindLaw 2019a). This limited acceptance may be partly due to a variety of ethical, legal, and logistical factors that have been asserted by a host of legal bodies, civil liberties activists, and scholars (FindLaw 2019a).

The test itself is performed by criminalists who explore the databanks for genetic information linked to a relative of a suspect whom they are trying to identify (FindLaw 2019a). For instance, if the suspect's sister has been arrested and her DNA is included in the computer system, a familial DNA search may connect the police with the sibling and, in turn, the suspect. (FindLaw 2019a).

FDS generated a fair amount of publicity when it was used to solve the Golden State Killer case in 2018. Between 1976 and 1986, the Golden State Killer had murdered over a dozen people and raped more than 15 victims (Reigle 2019). Despite the extensive publicity and horrific nature of the crimes, the identity of the Golden State Killer remained unknown for decades. This changed on April 24, 2018, when criminalists cracked the case and apprehend Joseph DeAngelo, a 72-year-old former police officer, as the elusive killer. The case was solved through the use of familial DNA testing (Reigle 2019). DeAngelo's identity was ascertained because a family member fortuitously submitted their own genetic DNA to GEDmatch.com. Although the police had saved DNA from the Golden State Killer's crime scenes, they could not match that evidence with anyone's DNA in the databases. When investigators uploaded the DNA evidence into an ancestry website, they obtained a match with the suspect's great-great-great-grandparent (Reigle 2019). Based on the identity of the distant relative, investigators were able to craft a detailed family tree that included DeAngelo. They then conducted surveillance of DeAngelo and recovered a discarded tissue containing his DNA (Reigle 2019). Fortunately, DeAngelo's recovered DNA sample matched the DNA of the infamous Golden State Killer. DeAngelo eventually pled guilty to the charges (Reigle 2019).

Running DNA evidence through large databases can often result in an erroneous match because various specific markers can be possessed by a large number of people (Romano 2018). The case of Michael Usry, a filmmaker from New Orleans, provides an unfortunate example. Usry gained the attention of the Idaho Falls Police Department because of the violent nature of his films (Mustian 2015). The police suspected the filmmaker of killing a woman in a 1996 case that had garnered much interest because it was claimed that the wrong person had been convicted (Mustian 2015). The key issue was the absence of a DNA match from the crime scene with anyone in the various government databases (Mustian 2015).

The police conducted a familial DNA search in a publicly accessible database through *Ancestry.com* and discovered a partial match with Usry's father. The authorities believed that this match demonstrated that one of the father's relatives was responsible for the crime (Mustian 2015). The court ordered the commercial site to perform a genetic search of its files, and the police narrowed their inquiry to three men in the father's family tree (Mustian 2015). The police then obtained a warrant to search Usry property and collected his DNA. Upon further testing, it was ascertained that the filmmaker's DNA did not match the evidence collected at the crime scene. Usry's experience, however, focused much attention on the difficult balance between an individual's right to privacy and public safety (Mustian 2015).

Similarly, before investigators identified Joseph DeAngelo as the Golden State Killer, they had misidentified a 73-year-old Oregon man as being a suspect. Investigators were led to this Oregon man based on DNA information found on the online

genealogical website. An Oregon judge ordered the man to provide a DNA sample based on the fact that he shared a rare genetic marker with the Golden State Killer. However, the Oregon man's DNA turned out to not be a match to the Golden State Killer's DNA.

The advantage of DNA profiling is based upon its specificity. A minute amount of DNA at a crime scene can produce adequate material for analysis (Brennan 2018). Advocates of familial DNA searches, therefore, assert that it is an innocuous way of generating leads for law enforcement officials hoping to solve cold cases (Rosenberg 2017). Homeland Security News Wire in 2011 announced that a partial DNA match of crime scene evidence to a genetic profile in a DNA database can direct police to relatives of possible suspects. Proponents further maintain that the technique could increase the number of identified suspects by 40%, thereby helping police resolve more open crimes.

According to the Homeland Security News Wire, critics retort that the technique violates the Fourth Amendment protection against unreasonable searches and seizures because the technique utilizes the DNA of innocent family members. There is also apprehension that the search will subject an unequal number of Hispanics and African-Americans to these searches because minorities have traditionally been arrested at disproportionately higher rates than other racial groups. One study revealed that while the overall rate of false identification is slight, African-Americans have two times the risk of being incorrectly selected for further investigation (Rohlfs et al. 2013). Accordingly, this risk will cause African-Americans to suffer disproportionately from invasions of privacy and police questioning (Kirchner 2017b).

There is little legislative oversight for familial DNA investigations, and there are currently no national standards to offer assistance on how states should proceed in creating guidelines for these searches (Debus-Sherrill and Field 2017). Most controlling rules are created at an agency level (Debus-Sherrill and Field 2017). Invasion of privacy unease denotes a recurrent theme arising from the rules of those jurisdictions that permit familial DNA searches (Debus-Sherrill and Field 2017). Most of these states limit the quantity or kinds of searches that may be conducted (Murray et al. 2017). For instance, a jurisdiction may restrict investigators from using this technique unless it is being used to solve cold cases or for problems that affect public safety such as major crimes involving violence (Murray et al. 2017). States may also expressly define the breath of familial searches in their rules and procedures (Murray et al. 2017).

Colorado has launched a familial research project with the Denver Police Department whose software only extends searches to parents and siblings (DNA Forensics n.d.). New York gives its law enforcement agencies permission to conduct a partial DNA search of their DNA database (DNA Forensics n.d.). Florida DNA Lab technicians are allowed to give criminalists the names of convicted offenders who match a crime scene sample if 21 of 26 alleles are present. Men who have 21 similar alleles are usually brothers (DNA Forensics n.d.). Since 2012, Ohio has also authorized familial DNA searches, but it has created a 12-page protocol that must be followed (McBride 2016).

In September 2019, the United States Department of Justice issued rules on when police can utilize genetic genealogy to identify suspects in serious crimes. This is the first-ever policy concerning how their databases may be employed in the efforts by law enforcement to balance public safety and privacy concerns (Kaiser 2019). The policy goes on to state that "forensic genetic genealogy" should only be used for violent crimes, as well as to identify human remains. The police are also cautioned to exhaust traditional crime-solving techniques, including first using their criminal DNA databases (Kaiser 2019).

DNA Phenotyping

In the 1960s, 90% of murders were solved in the United States (Veneziano n.d.). Since that time, the percentage of resolved homicides has been reduced significantly (Veneziano n.d.). Today, only 66% of illegal killings are resolved (Veneziano n.d.). Nevertheless, scientific advances have facilitated the solving of murders that many years ago would have gone unsolved. A major scientific advance that has allowed this to occur is the ability to identify an individual based on a genetic profile garnered from DNA (Marano and Fridman 2019). The technique, however, cannot provide a match unless that person's genetic profile is stored in a database. Recent DNA innovations offer promise in situations where a person's DNA profile is not contained in a database by creating a digital mug shot premised upon human genetics (Curtis and Hereward 2018).

Forensic DNA phenotyping (FDP) is touted as offering great promise by using DNA left at a crime scene to predict a suspect's eye, hair, and skin color, gender, age, and ancestry (MacLean and Lamparello 2014). DNA phenotyping is an attempt at duplicating what our bodies do normally: converting genetic DNA into a person's physical appearance (Emmerich 2016). Parabon, a marketer of this procedure, claims that by ascertaining how genetic DNA characteristics translate into physical appearance, it is possible to "reverse-engineer DNA into a physical profile." The company further notes that this is done by examining genetic variants, known as genotypes, from the sample to predict the appearance of the suspect. Parabon further claims that proponents of this technique maintain that the phenotype prediction has statistical confidence equal to or greater than 90%. In this regard, FDP can predict a source's gender with 100% accuracy and hair, iris, and height with 70% accuracy (MacLean and Lamparello 2014). Because FDP by itself cannot identify a DNA donor, scholars consider it to be merely an "investigative tool" and not "an instrument to identify a specific person" (Schneider et al. 2019).

Merriam-Webster Dictionary notes that phenotype is derived from the Greek words *phainein*, meaning "to show" and *typos*, meaning to "type." Therefore, phenotyping is the ability to offer an amalgamation of a person's discernable characteristics or traits. FDP can be used to predict a DNA donor's physical appearance and thereby generate investigative leads, narrow a suspect pool, or assist an investigation when a DNA database does not provide a match for a DNA sample (Ciavaglia 2020). The foundation of this technique is the premise that DNA holds the

key to a person's physical characteristics, and it creates a biological blueprint that can forecast "physical feature developmental propensities" (Cookson 2015).

Two forms of forensic phenotyping exist: direct and indirect (Davey 2020). With direct phenotyping, statistical methods are used to infer external characteristics (Davey 2020). Direct phenotyping can reveal external characteristics such as eye and hair color, hair shape, skin color, height, age, and shape of one's skull (Koops and Schellekens 2008). Some claim that it may even be feasible to predict whether an individual has the predisposition to smoke and stutter or is left-handed (Koops and Schellekens 2008). In using indirect phenotyping methods, external features are ascertained from genetic variants associated with a person's ancestry (Koops and Schellekens 2008). Indirect phenotyping is comparable to the knowledge generated by a DNA search done by *23 and Me* or *Ancestry.com* (Brown 2020).

Every scientific advancement must be thoroughly tested before it can be deemed accurate and legally sound. Some scholars are critical of forensic phenotyping since it can only offer a general outline of the suspect's appearance (Kaelin 2016). There is also a lack of peer-reviewed studies supporting the science (Southall 2017). Critics also pose questions about the techniques' accuracy, racial bias, and infringement upon a person's privacy (Waseem 2018). The limitation of relying upon a limited number of genetic traits is that it places "actual people, innocent of wrongdoing, under criminal suspicion without any basis in fact or science" (Southall 2017).

The American Civil Liberties Union has weighed in on the issue and noted that DNA phenotyping is dangerous because a person's looks are not solely a result of DNA. A person can dye his hair or shave it off, grow a beard, lose or gain weight, become disfigured in an accident, get too much sun exposure, smoke, lighten or darken the skin, have plastic surgery, undergo gender reassignment, and engage in many other events that can affect the person's looks but which have no relation to facial DNA (Stanley 2016).

Racial Profiling

A major issue with FDP is that it can perpetuate racial prejudices among law enforcement agencies (Pollack 2015). Critics of FDP have argued that the technique will make law enforcement even more dependent on racial identifiers (Maclean 2013). Accordingly, these critics maintain that because law enforcement is considered to already be discriminatory toward minorities and because DNA phenotyping will increase racial profiling, the technique is thought to exacerbate racial issues in law enforcement investigations (Maclean 2013). This, others argue, can lead investigators to devote more time to investigating crimes dealing with certain minorities or increase the penchant for tunnel vision (Koops and Schellekens 2008). As an example, the DNA results may reveal a 60% probability that a suspect is of African descent and a 40% chance of being Caucasian. These statistics could cause a biased analysis of the conclusion by the investigator to only investigate African suspects (Koops and Schellekens 2008). Critics also contend that since there are a larger number of people of color in the DNA databases, discriminatory police practices will

make them more vulnerable to suspicion and surveillance (Roberts 2011). Also, would publishing DNA phenotype cause certain members of society to form a kind of guilt-by-association mentality of an ethnic group within communities (Kirchner 2017a)? As one author noted, "[g]iven the current global context of criminalization of minority population groups, technologies such as forensic DNA phenotyping become attractive political objects of surveillance or . . . instruments of innovation techno-security" (Queirò 2019).

The opposite argument is that the unease about racial profiling is unfounded (MacLean and Lamparello 2014). DNA analysis of a specimen left at a crime scene is colorblind and similar to a fingerprint (Maclean 2013). Phenotype findings, on the other hand, are premised upon objective evidence and not a value judgment (Maclean 2013).

Right to Privacy

The custom by law enforcement officials of uploading a suspect's DNA to a publicly accessible website fosters substantial concerns about whether the police have violated the right to privacy of both a suspect and his or her family (Guest 2019). After all, the Supreme Court in 1965, in the landmark decision *Griswold v. Connecticut*, 381 US. 479, determined that people have a constitutional right to privacy even though that right is not overtly set forth within the framework of the Constitution. One might argue that DNA phenotyping comes within the protections afforded against the disclosure of confidential medical information (Koops and Schellekens 2008). It is one thing to discover that a person is a male with blonde hair, but it is another to discover that a person has a predisposition for a genetic disease or a disease that may be linked with crime, like alcoholism or drug misuse (Koops and Schellekens 2008). This raises a few basic privacy questions: the right to not know and personal autonomy, and where do you draw the line as to what is relevant and should not be allowed? (Maclean 2013).

The Fourth Amendment prohibits unlawful searches and seizures. Abandonment of property, however, is an exception to the Fourth Amendment's warrant requirement (Kitchen 2016). Accordingly, DNA left at a crime scene enjoys no reasonable expectation of privacy by itself because such genetic material is considered to be abandoned property (Gusella 2013).

Legislative Response

Courts and legislatures have provided little guidance regarding the use of DNA to predict a suspect's physical characteristics. The judiciary's failure to discuss FDP is not surprising. The lack of peer-reviewed studies and failure of the technique to gain general acceptance in the scientific community render DNA phenotyping inadmissible under either *Frye* or *Daubert*. The failure of the legislature to provide guidance, however, is more remarkable.

According to the Federal Bureau of Investigation's website dealing with frequently asked questions on CODIS and NDIS, the Director of the Federal Bureau of Investigations has the authority to establish a DNA database containing (i) the DNA identification records of individuals convicted of or charged with certain offenses, (ii) analyses of DNA recovered from crime scenes, (iii) examination of DNA recovered from unidentified human remains, and (iv) inspection of DNA samples voluntarily provided by family members of missing persons. Surprisingly absent from these statutes is the use of DNA phenotyping for reasons of arresting an unknown suspect. One commentator, however, notes that the FBI has interpreted federal law to allow sex determinations and inferences concerning ancestry and family kinship for investigative reasons (Murphy 2013). While the federal government has yet to pass any statute addressing DNA phenotyping, some state legislatures have done so.

DNA phenotyping has been addressed by a small number of states (Maclean 2013). Investigators in Indiana, Rhode Island, and Wyoming are prohibited by law from using DNA phenotyping. Vermont's statute seems to impliedly permit DNA phenotyping as it allows for the collection of DNA, but the statute explicitly precludes analysis of DNA "for identify[ing] … any medical or genetic disorder." Michigan follows a similar approach to Vermont as it allows the police to collect DNA for identification purposes, but the sample may not be "analyzed for identification of any medical or genetic disorder." The same can be said for Louisiana, whose statute also seems to allow DNA to be used to predict a suspect's physical characteristics. Louisiana's statute defines a "DNA record" as "DNA identification information stored in [a database] for the purpose of generating investigative leads" including "the characteristics … of value in establishing the identity of individuals." Texas law provides that "[t]he information contained in the DNA database may not be collected, analyzed, or stored to obtain information about the human physical traits or predispositions for disease unless the purpose for obtaining the information is related to a purpose described by this section."

DNA phenotyping has generated much excitement in the forensics community because of its capacity to furnish new clues in a cold case. Whether it will transform forensic investigations and obtain general acceptance in the law enforcement community is yet to be determined; however, there has been an uptick in the number of articles dedicated to the topic in scholarly journals with a number of them predicting an optimistic outcome from the utilization of the technique in criminal investigations (MacLean and Lamparello 2014).

Conclusion

The use of DNA has revolutionized the criminal justice system and provides objective proof that can either convict or exonerate a suspect (FindLaw 2019b). DNA supplies the entire genetic code for an individual, and it has become an important aspect of forensics since the late 1980s, helping to solve thousands of

crimes. The results are very accurate, and the chances of one person's DNA matching another are minuscule.

It is now a common operating procedure for the police to obtain samples of blood, saliva, or skin cells found at a crime scene and compare them to DNA profiles contained in various databases. The application of DNA in criminal investigations, however, is not without controversy. As new forensic applications for DNA evidence are uncovered, some of these techniques raise complex ethical and legal issues.

Familial DNA and forensic phenotyping are two recent applications of genetic markers in criminal investigations. These techniques offer great promise, but they raise complex issues concerning racial profiling, invasion of privacy, and test accuracy. Little judicial or legislative assistance has been offered to tackle these concerns. It will be interesting to see how familial DNA and forensic phenotyping continue to develop and whether the investigative methods gain uniform acceptance in the relevant communities.

References

Arnaud C (2017) Thirty years of DNA forensics: how DNA has revolutionized criminal investigations. https://cen.acs.org/articles/95/i37/Thirty-years-DNA-forensics-DNA.html. Accessed 20 Sept 2020

Balsamo M, Cooper J, Flaccus G (2018) Earlier search for California serial killer led to the wrong man. https://apnews.com/article/91d6dabd083a45cbbdb2361eadbaec96. Accessed 23 Sept 2020

Brennan J (2018) What are some advantages and disadvantages of using DNA analysis to aid law enforcement in crime? https://sciencing.com/advantages-aid-law-enforcement-crime-6776368.html. Accessed 23 Sept 2020

Brown S (2020) Ancestry vs. 23 and me: which DNA testing kit is best to help you find your family origins? https://www.cnet.com/health/ancestry-vs-23-and-me-which-dna-testing-kit-is-best-to-help-you-find-your-family-origins/. Accessed 25 Sept 2020

Cesar C, Cerqueira S, Ramallo V, Hünemeier T, de Azevedo S, Quinto-Sanchez M, Paschetta C, Cintas C, Gonzalez M, Schuler-Faccini L, Bortolini M, Gonzalez-Jose R (2016) Predicting physical features and diseases by DNA analysis: current advances and future challenges. J Forensic Res 7(4):1–10. https://doi.org/10.4172/2157-7145.1000336

Chin MW et al Forensic DNA evidence and the law § 1:1, Westlaw (database updated May 2018)

Ciavaglia J (2020) How does DNA phenotyping work? https://www.buckscountycouriertimes.com/news/20200128/how-does-dna-phenotyping-work. Accessed 23 Sept 2020

Cookson C (2015) DNA: the next frontier in forensics. https://www.ft.com/content/012b2b9c-a742-11e4-8a71-00144feab7de. Accessed 23 Sept 2020

Cormier K, Calandro L, Reeder D (2005) Evolution of DNA evidence for crime solving - a judicial and legislative history. https://www.forensicmag.com/article/2005/01/evolution-dna-evidence-crime-solving-judicial-and-legislative-history. Accessed 22 Sept 2020

Cronen E (2008) DNA more effective than fingerprints in solving property crimes. http://webarchive.urban.org/publications/901179.html#:~:text=DNA%20More%20Effective%20Than%20Fingerprints%20in%20Solving%20Property%20Crimes,-Urban%20Institute&text=In%20crime%20scenes%20where%20biological,to%20lead%20to%20an%20arrest. Accessed 25 Sept 2020

Curtis C, Hereward J (2018) How accurately can scientist reconstruct a person's face from DNA? https://www.smithsonianmag.com/innovation/how-accurately-can-scientists-reconstruct-persons-face-from-dna-180968951. Accessed 23 Sept 2020

Davey R (2020) What is DNA phenotyping? https://www.news-medical.net/life-sciences/-What-is-DNA-Phenotyping.aspx. Accessed 23 Sept 2020
Debus-Sherrill S, Field M (2017) Understanding familial DNA searching: policies, procedures, and potential impact, summary overview, Office of Justice Programs. https://www.ncjrs.gov/pdffiles1/nij/grants/251043.pdf. Accessed 23 Sept 2020
DNA Forensics (n.d.) States using familial searches. http://www.dna.forensics.com/statesandfamilialsearches.aspx. Accessed 23 Sept 2020
Emmerich G (2016) The reality of DNA phenotyping. https://www.promegaconnections.com/the-reality-of-dna-phenotyping/. Accessed 23 Sept 2020
FindLaw (2019a) Familial DNA searches. https://criminal.findlaw.com/criminal-rights/familial-dna-searches.html. Accessed 23 Sept 2020
FindLaw (2019b) How DNA evidence works. https://criminal.findlaw.com/criminal-procedure/how-dna-evidence-works.html. Accessed 23 Sept 2020
Forensics Colleges (n.d.) 10 modern forensic technologies. https://www.forensicscolleges.com/blog/resources/10-modern-forensic-science-technologies. Accessed 20 Sept 2020
Gaille L (2018) 10 pivotal pros and cons of DNA fingerprinting. https://vittana.org/10-pivotal-pros-and-cons-of-dna-fingerprinting#:~:text=The%20primary%20disadvantage%20of%20DNA,positive%20or%20false%20negative%20results. Accessed 20 Sept 2020
Genetics Home Reference (2020a) Help me understand genetics. https://ghr.nlm.nih.gov/primer. Accessed 20 Sept 2020
Genetics Home Reference (2020b) What is DNA? https://ghr.nlm.nih.gov/primer/basics/dna. Accessed 20 Sept 2020
Guest C (2019) DNA and law enforcement: how the use of open source DNA databases violates privacy rights. Am Univ Law Rev 68(3):1015–1052
Gusella D (2013) No cilia left behind: analyzing the privacy rights in routinely shed DNA found at crime scenes. Boston Coll Law Rev 54(2):789–820
In Brief (n.d.) What are the advantages of the use of DNA in criminal cases? https://www.inbrief.co.uk/court-proceedings/dna-use-in-criminal-cases. Accessed 21 Sept 2020
Kaelin B (2016) DNA phenotyping: revealing the faces of killers. https://www.forensicsciencedegree.org/dna-phenotyping-revealing-the-faces-of-killers/. Accessed 23 Sept 2020
Kaiser J (2019) New federal rules limit police searches of family tree DNA databases. https://www.innocenceproject.org/virginia-enacts-familial-dna-testing-law. Accessed 23 Sept 2020
Kirchner L (2017a) Will 'DNA phenotyping' lead to racial profiling by police? https://psmag.com/news/will-dna-phenotyping-lead-to-racial-profiling-by-police. Accessed 23 Sept 2020
Kirchner L (2017b) The flaws of familial DNA matching. https://psmag.com/environment/the-flaws-of-familial-dna-matching-64736. Accessed 23 Sept 2020
Kitchen A (2016) Genetic privacy and latent crime scene DNA of nonsuspects: how the law can protect an individual's right to genetic privacy while respecting the government's important interest in combatting crime. Crim Law Bullet 52(2):346–370
Kolata G (1992) Say DNA analysis is reliable in identifying suspects. https://www.nytimes.com/1992/02/07/health/scientists-say-dna-analysis-is-reliable-in-identifying-suspects.html#:~:text=Scientists%20who%20analyzed%20data%20used,innocent%20person%20in%20a%20crime.&text=Though%20the%20technique%20has%20been,has%20not%20been%20proved%20reliable. Accessed 20 Sept 2020
Koops BJ, Schellekens M (2008) Forensic DNA phenotyping: regulatory issues. Columbia Sci Tech Law Rev 9:158–202
Liberty A (2015) Defending the black sheep of the forensic DNA family: the case for implementing familial DNA searches in Minnesota. Hamline Law Rev 38(3):467–517
Maclean C (2013) Creating a wanted poster from a drop of blood: using dna phenotyping to generate an artist's rendering of an offender based only on DNA shed at the crime scene. Hamline Law Rev 36(3):357–386
MacLean C, Lamparello A (2014) Forensic DNA phenotyping in criminal investigations and criminal courts: assessing and mitigating the dilemmas inherent in the science. Recent Adv DNA Gene Seq 8(2):104–112. https://doi.org/10.2174/2352092209666150212001256

Marano L, Fridman C (2019) DNA phenotyping: current applications in forensic science. Res Rep Forensic Med Sci 9:1–8. https://doi.org/10.2147/RRFMS.S164090
Matheson S (2016) DNA phenotyping: a snapshot of a criminal. Cell 166(5):1061–1064. https://doi.org/10.1016/j.cell.2016.08.016
McBride J (2016) Familial DNA searches: 5 fast facts you need to know. https://heavy.com/news/2016/12/familial-dna-searches-search-karina-vetrano-new-york-queens-father-analysis-testing-test-current-approaches-justin-christian-ohio/. Accessed 23 Sept 2020
Murphy E (2013) Legal and ethical issues in forensic DNA phenotyping. NYU School of Law, Pub Law & Legal Theory Res. Paper series working paper. No. 13-46
Murray A, Grant C, Beyer M, Wilz A, Williamson E, Maier J, Kesselring L, Wojtas K, Gibbs J, Karakkat K, Quarino L (2017) Familial DNA testing: current practices and recommendations for implementation. Investigative Sci J 9(4):1–10
Mustian J (2015) New Orleans filmmaker cleared in cold-case murder; false-positive highlights limitations of familial DNA searching. https://www.theadvocate.com/new_orleans/news/article_1b3a3f96-d574-59e0-9c6a-c3c7c0d2f166.html. Accessed 23 Sept 2020
National Human Genome Research Institute (n.d.-a) Deoxyribonucleic Acid (DNA). https://www.genome.gov/genetics-glossary/Deoxyribonucleic-Acid. Accessed 4 July 2020
National Human Genome Research Institute (n.d.-b) DNA fingerprinting. https://cen.acs.org/articles/95/i37/Thirty-years-DNA-forensics-DNA.html. Accessed 20 Sept 2020
Newman T (2018) What is DNA and how does it work? https://www.medicalnewstoday.com/articles/319818#what-is-dna. Accessed 11 Jan 2018
Norrgard K (2008) Forensics, DNA fingerprinting, and CODIS https://www.nature.com/scitable/topicpage/forensics-dna-fingerprinting-and-codis-736/. Accessed 22 Sept 2020
Panneerchelvam S, Norazmi MN (2003) Forensic DNA profiling and database. J Med Sci 10 (2):20–26
Pollack A (2015) Building a face, and a case, on DNA. https://www.nytimes.com/2015/02/24/science/building-face-and-a-case-on-dna.html. Accessed 23 Sept 2020
Queirò P (2019) The visibilities and invisibilities of race entangled with forensic DNA phenotyping technology. J Forensic Legal Med 68:101858. https://doi.org/10.1016/j.jflm.2019.08.002
Rainey J (2018) Familial DNA puts elusive killers behind bars but only 12 states use it. https://www.nbcnews.com/news/us-news/familial-dna-puts-elusive-killers-behind-bars-only-12-states-n869711. Accessed 23 Sept 2020
Reigle D (2019) Familial DNA testing, the golden state killer, and how it affects you. https://jolt.richmond.edu/2019/10/30/familial-dna-testing-the-golden-state-killer-and-how-it-affects-you/. Accessed 23 Sept 2020
Roberts D (2011) Collateral consequences, genetic surveillance, and the new biopolitics of race. Howard Law J Symp 54:567–586
Rohlfs R, Muphy E, Song Y, Stalkin M (2013) The influence of relatives on the efficiency and error rate of familial searching. PLoS One 9(1). https://doi.org/10.1371/annotation/2544c2e3-0065-4d14-ab1c-3d3fcb0fee84
Romano A (2018) DNA profiles from ancestry websites helped identify the golden state killer suspect. https://www.vox.com/2018/4/27/17290288/golden-state-killer-joseph-james-deangelo-dna-profile-match. Accessed 23 Sept 2020
Rosenberg E (2017) Family DNA searches seen as crime-solving tool and intrusion on rights. https://www.nytimes.com/2017/01/27/nyregion/familial-dna-searching-karina-vetrano.html. Accessed 23 Sept 2020
Rothstein M (2014) Privacy and technology in the twenty-first century. U Louisville Law Rev 52: 333–344
Schneider PM, Prainsack B, Kayser M (2019) The use of forensic DNA phenotyping in predicting appearance and biogeographic ancestry. Dtsch Arztebl Int 116(51–52):873–880
Science Learning Hub (2005) DNA profiling. https://www.sciencelearn.org.nz. Accessed 20 Sept 2020

Smith G, Gordon J (1997) The admission of DNA evidence in state and federal courts. Fordham Law Rev 65(6):2465–2488

Southall A (2017) Using DNA to sketch what victims look like: some call it science fiction. https://www.nytimes.com/2017/10/19/nyregion/dna-phenotyping-new-york-police.html. Accessed 23 Sept 2020

Stanley J (2016) Forensic DNA phenotyping. https://www.aclu.org/blog/privacy-technology/medical-and-genetic-privacy/forensic-dna-phenotyping. Accessed 23 Sept 2020

Strutin K (2015) DNA without warrant: decoding privacy, probable cause, and personhood. Richmond J Law Pub Interest 18(3):319–368

Thompson W, Mueller L, Krane D (2012) Forensic DNA statistics: still controversial in some cases. The Champion 12–23. https://www.nacdl.org. Accessed 23 Sept 2020

Veneziano S n.d. Painting a mug shot with genetics: a spotlight on DNA phenotyping. https://www.innocenceproject.org/virginia-enacts-familial-dna-testing-law. Accessed 23 Sept 2020

Waseem F (2018) How a Reston-based technology company is helping crack cold cases. https://www.restonnow.com/2018/02/14/how-a-reston-based-technology-company-is-helping-crack-cold-cases/. Accessed 23 Sept 2020

A Glimpse of Famous Cases in History Solved by DNA Typing

15

Hirak Ranjan Dash, Kamayani Vajpayee, and Radhika Agarwal

Contents

H. R. Dash (✉)
School of Forensic Science, National Forensic Sciences University, Delhi Campus, New Delhi, India

K. Vajpayee
DNA Fingerprinting Unit, Forensic Science Laboratory, Bhopal, Madhya Pradesh, India

School of Arts and Sciences, Ahmedabad University, Ahmedabad, India

Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, Gujarat, India

R. Agarwal
DNA Fingerprinting Unit, Forensic Science Laboratory, Bhopal, Madhya Pradesh, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_14

Abstract

DNA fingerprinting technique is considered to be the most irrefutable evidence in the criminal justice system. The technique has been advanced from the RFLP-based analysis to the estimation of alleles of PCR-amplified STR markers by capillary electrophoresis. The present-day forensic DNA practice is witnessing a paradigm shift from CE-based analysis to NGS-based sequence determination of the alleles. Despite all the advancements, the technology has been used successfully in solving many criminal and civil cases of paternity dispute, identification, murder, and sexual assault nature. Few highlighted cases solved by this technique have been discussed in this chapter. Besides, the technology has also been used in solving many cold cases in history as well as exonerating innocents. Though NGS technology is in its preliminary stage of use in forensic DNA analysis, its importance in solving complicated criminal cases has also been discussed.

Keywords

DNA fingerprinting · Criminal cases · History · Innocence project · Cold cases

Introduction

The advent of the DNA fingerprinting technique has revolutionized the field of criminal justice system. When the technique is performed under proper guidelines and quality control measures, it is highly reliable in convicting a criminal as well as exonerating innocent individuals. The fundamental of this technique relies on the pure genetics approach. The human genome consists of around 3 billion base pairs of nucleotides which has enough information for any individual to be characterized. However, such genetic information is coded by only lee than 10% of the genome. Rest 90% of the genome is called junk DNA whose function is unknown to date (Mattick and Dinger 2013). Such noncoding DNA sequences are mostly comprised of repeat sequences. These highly polymorphic noncoding regions are assigned as the mini- or microsatellites which are localized at specific sites of a chromosome, which are spread throughout the genome. The present-day DNA analysis explores such microsatellites or short tandem repeat (STR) markers for genetic individualization (Kim et al. 2018).

In the initial days, DNA fingerprinting examination was carried out by restriction fragment length polymorphism (RFLP) technique discovered by Sir Alec Jeffreys. In this technique, isolated DNA is subjected to restriction enzymes for the generation of various DNA fragments based on the availability of specific restriction sites. The digested DNA fragments are separated by agarose gel electrophoresis to generate

distinguished bands for each DNA fragment. Such DNA fragments are further transferred to a membrane via the Southern blot procedure. A labelled DNA probe is then hybridized to the membrane and the length of the DNA fragments complementary to the probes are determined. Each length fragment is considered as an allele and is used for further analysis for a match or mismatch between the questioned and reference samples (Rankin et al. 1996). However, such a technique was highly labor-consuming and required a huge amount of high-quality DNA which is rare in real-time forensic cases (Yang et al. 2013). Hence, the technique was further evolved to polymerase chain reaction (PCR) based amplification of STR markers and their determination by capillary electrophoresis (CE) technique.

Many multiplex kits have been developed over the years to analyze STR markers on autosomal, Y and X chromosomes (Kayser 2017). The major advantage of the STR technique over the earlier RFLP-based technique is that they are relatively easy to identify and collate, regardless of the analysis platform used. Amplification of STR markers using PCR allows to utilization of minimum pieces of evidence, most commonly encountered in the crime scene. Besides, multiplexing of STR markers allows the identification of an individual with high discrimination power (Gill 2002). Based on the results obtained from the STR analysis over time, many DNA databases have been established throughout the world to solve criminal cases by searching the crime scene DNA in the database to get a hit (Jakovski et al. 2017). DNA technology is evolving continuously by the introduction of mtDNA analysis, the use of geographic, ancestral, and phenotypic single nucleotide polymorphism (SNP) markers, and the use of next-generation sequencing (NGS) (Sobrino and Carracedo 2005). The technology was initially invented to be used in human identification. Many success stories of this technology in solving criminal cases are found in history. A few of such interesting cases have been discussed in this chapter.

Famous Cases in History Solved by Human DNA Analysis

Colin Pitchfork Case

The first case in history solved by DNA analysis is not less than a fairy tale. A 15-year-old Lynda Mann went to visit a friend on 21 November 1983 but did not return home after that. The next day, her body was found laying on a nearby footpath. Autopsy analysis confirmed sexual assault with the victim and death occurred due to strangulation. As a result of sexual assault, semen samples could be recovered from the victim's body, but due to the unavailability of any investigative leads, the case becomes cold. Similarly, on 31 July 1986 another girl, Dawn Ashworth went missing in a similar fashion and her body was found in a forest area 2 days later. Sexual assault was also confirmed with the victim besides strangulation as the cause of death. A semen sample was also collected from Dawn's dead body. As both the cases were similar, the investigation process continued simultaneously for both the cases. Later, a 17-year-old Richard Burkland who was familiar with Dawn's dead body confessed to both the crimes during the interrogation of the investigating agency.

As the DNA fingerprinting technique was recently discovered by that time by Alec Jeffreys of the University of Leicester, the authorities thought of utilizing that technique to confirm the identity of the culprit. Surprisingly, the DNA profile of Richard did not match with the DNA profile obtained from the semen samples of both cases (Butler 2010). The investigating agencies in collaboration with the University of Leicester collected blood and saliva samples from around 5000 local men to find a DNA match with the semen samples. But none of the samples resulted in a DNA match with the semen samples. Later, a man Ian Kelly was overheard in a bar regarding his acceptance of £200 to provide his blood sample for his friend Colin Pitchfork, a local baker. In connection with that, Colin Pitchfork was arrested on 19 September 1987 and when his original DNA sample was analyzed, a perfect match was obtained with the semen samples obtained from the earlier two sexual assault and murder cases. Subsequently, he admitted to the crime and was awarded life imprisonment. Thus, Colin Pitchfork becomes the first person to be convicted based on DNA technology, and hence, his name holds weight in Forensic DNA technology (Saad 2005).

Tommie Lee Andrews Case

A man burglarized and assaulted a woman sexually at gunpoint after entering her home in Florida on the night of 21 February 1987. Investigating agency retrieved fingerprints from the victim's home but could not find a match in the database. When Tommie Lee Andrews was arrested in a similar nature of series of cases, the fingerprint could be matched but the accused continuously denied his involvement in the sexual assault cases. Though the legal case against Tommie Lee Andrews was much stronger, Jeffrey Ashton, an assistant attorney, Florida was interested in proving sexual assault through the evolving technique, i.e., DNA fingerprinting. During the trial, the accused tried to escape a DNA test and claimed against the laboratory. Further, a DNA test was conducted and the DNA profile retrieved from the semen sample collected from the home of Florida's victim was found to match with the DNA profile obtained from the reference sample of Tommie Lee Andrews. The sexual assault case against the accused was hence proved and he was awarded 20 years of imprisonment for sexual assault, aggravated burglary, and burglary (Pearsall 1989). At that time, DNA database was not available in any part of the world. Otherwise, the case could have solved much earlier by searching the crime scene DNA profile in the database. Finally, Tommie Lee Andrews has become the very first person to be convicted by DNA evidence in the US court.

Joseph Castro Case

People of the State of New York v. Joseph CASTRO is a landmark case in the history of the use of DNA testing in criminal investigations. It was for the very first time where the admissibility of DNA testing was questioned in the US court.

Castro was accused of murdering his neighbor along with her child. The bloodstain recovered from his watch was taken for DNA examination by Lifecodes Corp in 1987. They reported that the DNA analyzed from the bloodstain matched with the victim, Vilma Ponce. However, Castro defended by swearing that the bloodstain was his. A rigorous 12 weeks of pretrial was held by the New York Supreme Court on the admissibility of DNA evidence. Judge Gerald Sheindlin in 1989 raised three-pronged tests to determine the admissibility of the DNA evidence. Out of the three-pronged, the Supreme court held that the first two prongs are generally accepted by the scientific community whereas, the third one which talked about the scientific and validated methods being used by the research laboratory was not accepted. The court's ruling states that DNA identification-based exclusion is more admissible than that of inclusion. Besides, the court also stated that DNA tests could be used to show that blood on Castro's watch does not belong to him, but tests could not be used to show that the blood was that of his victims. Based on pretrial and prongs the DNA evidence was held inadmissible. However, later, Castro confessed to the murders.

O. J. Simpson Murder Case

The most infamous case in the history of forensic DNA analysis involves the case of murder suspect O.J. Simpson, a famous sports personality in the USA. On 12 June 1994, both Simpson and his ex-wife Nicole Brown attended their daughter's dance presentation at her school. Later, Brown and her family went to a restaurant without Simpson followed by having ice cream and returned to her house at Brentwood. At around 9.00 pm Ron Goldman, a waiter of the restaurant went to Brown's home to drop the eyeglasses that were left by Brown's mother in the restaurant. When Brown's dog was found barking by the neighbors, they visited the home to find the dead body of Brown lying outside her home. Later police reached the crime scene to discover Goldman's dead body next to that of Brown's. Brown was found murdered with multiple stabbings in her head and neck with some defensive wounds present on her hands. Goldman was found killed by multiple stabbing in his body and neck with relatively few defensive wounds. After through crime scene examination and prima facia, O. J. Simpson was suspected to be involved in the brutal murder and was prosecuted subsequently.

With no eye witness to the case, the investigating agency tried to establish the suspect by a recently introduced DNA fingerprinting examination. The DNA examination showed the presence of Simpson's DNA on blood drops next to the victim's body as well as towards the back gate at Bundy, with an error rate of 1 in 9.7 billion. DNA of Simpson, Goldman, and Brown was found inside Simpson's vehicle, outside his door, and on the bloody gloves found behind Simpson's home. Additionally, the DNA of Simpson and Brown was found on the bloody pair of socks found in Simpson's bedroom at an error rate of 1 in 6.8 billion. Besides DNA evidence, hair and fiber examination also connected between Simpson and the crime scene. Shoe print analysis also suggested a single assailant wearing the shoe of a rare kind often used by Simpson with the same shoe number (Geis and Bienen 1998).

Despite having a huge bundle of evidences against O. J. Simpson, he was finally acquitted in this infamous murder case. The DNA evidence gathered against Simpson was coherent and complete with factual support from the prosecution. But the prosecution failed to make the judge understand the complex nature of the DNA test and the jury was found confused about the test (Thompson 1996). Most of the DNA samples in Simpson's case contained mixed sources. Thus, the defense argument of a high probability of a random man included in the mixed stain was accepted by the jury. Many valuable lessons were learned from this case regarding the limitations of DNA evidence and the importance of collection, preservation, and handling of biological evidences before their DNA examination.

Craig Harman Case

Michael Little was driving his truck in Surrey on 21 March 2003. While traveling through an underpass, he was hit by a brick from the overpass and it caused severe fatal injury to his chest. However, the assailant could not be identified in this case. During the investigation, two DNA profiles could be detected from the source of the brick, one of Little and the second was unknown. Before this incident, a car was burglarized in the same town. In that case, the suspect was not able to start the car and eventually left his bloodstains in the car. When the DNA profile was generated from the bloodstains found in the car, it was a complete match with the additional profile found on the brick that eventually killed Little. The DNA profile was searched in the database but no hit could be generated. Further, the DNA profile was identified to be of Caucasian origin. An investigative lead was also obtained regarding the accused being below 35 years of age. To identify the culprit, the investigating agency conducted a DNA examination on more than 350 individuals from the nearby village, but still, no match was found.

Later, the law enforcement agency decided to perform a familial searching to find a clue in this case. When such DNA examination was conducted on white males under 35 years of age, a similar DNA profile was obtained in 25 individuals. In one person, a match was found in 16 out of 20 DNA markers. He told the investigating team that, his brother lives in the same location of the crime scene. With this lead, police found his 19-year-old brother Craig Harman. Further DNA examination revealed the same origin of the bloodstain on brick and car and that of Craig Harman. He admitted to the crime and was sentenced to 6 years of imprisonment after pleading guilty to manslaughter. This case is regarded as the first case in history to be solved by familial search. Since then, familial searching of DNA databases has gained momentum in many countries including the USA, the Netherlands, Australia, and New Zealand to solve a blind case (Maguire et al. 2014).

Identification of 9/11 Victims

Before the 9/11 attack on the World Trade Center, DNA technology was being used to identify mass causalities limiting to not more than 500 individuals (Leclair et al.

2004). The death in such an unprecedented number and condition of mortal remains stood a huge challenge before the investigating team to identify each victim. In a parallel line, the investigating team worked long shifts 7 days a week to generate a DNA profile from each tissue received along with the known reference samples, i.e., tissue from biopsy samples, Pap smears, extracted teeth, and saliva from toothbrushes (Reif 2002). Besides, a DNA profile was also generated from the blood samples of the surviving families for potential matches. In the Pennsylvania site, all 40 passengers were identified based on DNA examination along with four unknown DNA profiles presumed to be originating from the terrorists. Similarly, in the Pentagon site, 179 out of 185 people were identified with their family reference samples. The rest five samples were thought to be originating from the terrorists (Reif 2002). Both these DNA-based identification processes were completed by 11 December 2001 and 16 November 2001, respectively.

The major challenge faced by the DNA analysts was the condition of the samples. In certain instances, a complete body was discovered, whereas, in another instance, tiny fragments of charred bones were found. The fire generated a temperature of about 1000 °C, which remained in the site for nearly 3 months. Tissue samples were also retrieved after many days leading to compromised samples (Biesecker et al. 2005). This leads to the generation of adversities to the existing DNA-based identification technologies. Initially, a 13 locus Combined DNA Index System (CODIS) panel was used for identification. As most of the samples were fragmented and degraded due to exposure to high temperature, a low yield was obtained in most of the samples. Upon recommendation from the Kinship and Data Analysis Panel (KADAP), mtDNA analysis was tried due to their high copy number and successful utilities in sample limiting conditions. As mtDNA analysis alone cannot generate a conclusive identification, the conjugation of mtDNA analysis along with miniSTRs were also tried (Butler et al. 2003). However, for such a huge number of samples, mtDNA analysis was found to be labor-intensive as semi-automated techniques are involved in this process. Finally, the single nucleotide polymorphism (SNP) markers were found to be most suitable for this identification due to their small amplicon size (Biesecker et al. 2005). Such a huge effort generated around 52,000 STR, 44,000 mtDNA, and 17,000 SNP profiles from 19,913 putative victim tissue fragments and 2749 missing individuals. This case is considered to be one of the mass-scale identification of human remains of worse quality through DNA fingerprinting analysis.

Famous Cases in History Solved by Nonhuman DNA Analysis

Snowball Case

Shirley Ann Duguay, a Canadian woman was reported missing from her home on 3 October 1994. Within few days, her abandoned car was found containing her bloodstains. After 3 weeks, a leather jacket was found in the woods 8 km from the victim's home containing the victim's blood. Keen forensic observation led to the finding of few domestic cat hairs on the jacket. On 6 May 1995, the victim's body

was found in a narrow grave in Prince Edward Island (Menotti-Raymond et al. 1997a).

Out of 27 hairs found on the jacket, DNA was extracted from one hair containing the root. DNA typing was carried out at ten feline dinucleotide repeat STR loci that were discovered earlier for optimum use in forensic analysis (Menotti-Raymond et al. 1997b). The suspect in this case was the deceased's estranged husband, Douglas Beamish who was living with his parents and a white American shorthair pet cat named Snowball. Similarly, the DNA profile was generated from the blood sample of Snowball and genotyped. A statistically significant match was obtained between these two DNA profiles with an error rate of 2.2×10^{-8} and 6.9×10^{-7} for the island and the US canine population respectively (Menotti-Raymond et al. 1997a). Based on this DNA result, the Supreme Court of Prince Edward Island convicted the defendant of second-degree murder and he was awarded a life sentence for this. This is the first registered case in history to be solved by nonhuman DNA analysis and promises its huge implications in forensic science.

The Bogan Case

On 3 May 1992, a women's dead body was found in a deserted place in Phoenix, Maricopa County, Arizona. The crime scene investigation suggested strangulation with the deceased's T-shirt as the cause of death. A syringe and pieces of clothing were also found around the body. Besides, a pager was also found nearby which was collected. A fingerprint search of the victim identified her as 30-year-old Denise Johnson. As she was a drug addict and used to sell drugs on the highways, the investigators could not ascertain a prime suspect in the case. However, on further investigations, the pager owner was detected as Mark Bogan, a local truck driver. When he was questioned by the investigating agency, he confessed to meet the deceased but had never been to the interior place where the body was found. Rather, he said that he dropped Denise on the highway after giving her a lift. Further, the autopsy report showed high cocaine content in the deceased body. There was no point to consider Bogan as a suspect and the case remained with no leads and no suspects.

Later the case was transferred to a Homicide Detective Charles Norton. When Charles visited the crime scene, he came across a fresh abrasion on one of the low branches of a Palo Verde tree. It was photographed and he collected few beans off the tree for further analysis. When Mark Bogan's truck was searched thoroughly, no trace of evidence could be detected in the form of blood, semen, saliva, fingerprints, or any others. However, two bean pods of the Palo Verde tree were found in the back of the truck. Randomly amplified polymorphic DNA (RAPD) technique was employed to match two plant DNAs in the University of Arizona. Additionally, randomly selected Palo Verde trees were also included in the analysis. The result showed a distinct profile for all the trees and a match in the DNA profile was obtained between the pod taken from the tree at the crime scene and bean pods found on the back of Bogan's truck. Later he admitted to the crime and was found

guilty of murder in the first degree and was awarded life imprisonment. This case is considered to be the first case in history to be solved by the analysis of plant DNA (Sensabaugh and Kaye 1998).

Cold Cases Solved by DNA Analysis

With the advent of DNA technology, many organizations have started funding the examination of "cold cases" by this technology. In such programs, the investigating agencies are entitled to identify, review, reinvestigate, and reanalyze the mysterious cold cases through DNA technology. Most of these cases undergoing reexamination date back to the earlier days, when forensic DNA technology was not discovered. Reexamination of cold cases through DNA analysis has solved many important cases of homicides, sexual assaults, and other violent offenses. Before the discovery of STR-based technique, a substantially higher quantity of DNA with good quality was required for forensic DNA analysis. The development of miniSTRs, Y-STRs, and mtDNA analysis over time have also revolutionized the forensic DNA analysis of sexual assault and missing person cases (Marks 2009). The establishment of criminal databases and periodical searching of DNA profiles from the previously unsolved cases have resulted in the identification of culprits in many cases (Panneerchelvam and Norazmi 2003). In this regard, the crime scene samples, which were previously considered to be unsuitable for such tests, are also being reexamined with the advanced technology to obtain a strong conclusion. To date, many such cases have been solved over time using advanced DNA technology. A few such cases have been discussed below.

The Boston Strangler Case

Mary Sullivan, a 19-year-old unmarried girl moved to Boston from Cape Cod and started living in a rented accommodation in the bustling Beacon Hill neighborhood. After few days of her arrival, she was found dead in January 1964. The autopsy suggested sexual assault with her and the cause of death was strangulation. Albert DeSalvo, famously known as the Boston Strangler, confessed to the killing of 11 women including Mary Sullivan. Further, during the court proceedings, Albert recanted the crime leaving behind the investigating agencies to think that the real assailant had eluded in this particular case. Though Albert was convicted of life imprisonment in other sexual assault cases, his association with any of the Boston strangulation cases could not be established with the available scientific techniques. Later, Albert DeSalvo was murdered by fellow jail inmates in 1973. But after decades of his death, the mysterious cases of strangulation and sexual assault remained open regarding its actual culprit.

When the city of Boston received funds under the cold case program during 2009–2012, the case was reexamined by DNA technology. In this case, DNA was analyzed from the seminal fluid found on Sullivan's body and a blanket of the crime

scene. The Y chromosomal STRs were analyzed in the sample along with the reference sample from DeSalvo's nephew. The Y chromosome STRs from the crime scene matched with the DNA profile obtained from DeSalvo's nephew. Baring mutation, Y-STRs are found to be the same in the patrilineal lineage which includes fathers, sons, brothers, paternal uncles, nephews, as well as third and fourth cousins (Ballantyne et al. 2012). Thus, the Y-STR match between the crime scene and DeSalvo's nephew excluded 99.9% of the male population to be the possible contributor of male DNA. To individualize a DNA profile, the investigating authorities exhumed DeSalvo's body and conducted a confirmatory autosomal STR analysis using femur bone and three teeth samples (Foran and Starrs 2004). Based on this analysis, the chance of a random white man other than DeSalvo contributing the DNA profile is one in 220 billion, thus obtaining a strong conclusion regarding the assailant in Mary Sullivan's sexual assault and murder case.

Murder and Sexual Assault of Girl in 1988

A young babysitter of 22 years old was found murdered by suffocation in a pillow in the living room of her working place in northern Italy in 1988. Crime scene investigation found struggle signs as well as the possibility of sexual assault. However, due to the limited availability of analytical techniques, further biological analyses could not be carried out at that moment. The case was reopened in 2007 and 2010 to analyze the DNA evidences of bloodstains found on the crime scene as well as on the clothes of the deceased. The male origin of the bloodstains as revealed by DNA analysis increased the curiosity of the investigating agency to reach out to the possible culprit. Further investigation confirms the involvement of a member of the sect, who was 14 years old at the time of the incidence by a DNA match (Caglia et al. 2011). Finally, in 2011, the respondent, who is 37 years now was convicted for the crime committed in 1988. This highlights the beauty of DNA technology, which helps in reaching out of a person even after decades of committing a crime.

Arrest of Nebraska Sex Offender in 1983 Slaying of UNO Student

In mid-August, 1983, the naked body of Firozeh Dehghanpour, a 20-year-old student of the University of Nebraska, Omaha, was found by some fishermen under a bridge. An autopsy examination of the dead body confirmed the death due to severe bleeding originating from cuts of the throat and abdominal stabs. However, the report could not confirm any sexual assault of the victim. Blood and hairs found on the bridge railing and crime scene were examined with the available technology. Though the hairs could be attributed to Dehghanpour's origin, some of the bloodstains were found to be exogenous. Further investigation with around 100 UNO students could not reveal the identity of the culprit. The authorities also denied any possible connection with the unrest in Iran. Later, a UNO business student, also a friend of Dehghanpour contacted Omaha and Council Bluffs police departments to

reinvestigate the case with the advanced DNA technology. As the biological samples were collected earlier from the crime scene and preserved, a DNA examination was carried out. When the obtained DNA profile was hit in the CODIS database, it showed a match with Christensen, who is listed on the Nebraska Sex Offender Registry as a lifetime offender. This case adds another feather to the glory of forensic DNA technology and its use in solving previously unsolved cold cases.

DNA Examination of Unsolved Female Homicide Cases During 1990–1999

In the pre-DNA era, vaginal, oral, and anal swabs and smears were being collected from female homicides and examined for the presence or absence of sperms. After examination, the samples were stored in the Office of the Chief Medical Examiner (OCME) archives. After the discovery of successful DNA examination of postcoital slides stained with cytological techniques (Dimo-Simonin et al. 1997), a total of 376 cases were examined with DNA technology to find an investigative lead. Out of six unsolved cases, DNA profile was generated in as many numbers of samples. In two cases, the CODIS result showed two hits with the repeated offenders, thus solving the case outrightly. In another two cases, mixed DNA profile could be detected, hence could not be submitted in CODIS for a probable matching. In rest two cases, reference samples of the suspect(s) were analyzed for matching. However, in both cases, the suspects were exonerated and a CODIS entry was made in such cases (Clark et al. 2019). The stability and quality of sperm DNA is always a question in sexual assault cases as DNA fragmentation and degradation occur over time due to oxidative stress and apoptotic activity (John Aitken and Koppers 2011). Still generation of conclusive DNA results in cold sexual assault cases has opened a huge window for the forensic DNA practitioners as well as the investigating agencies to have closer look at the previously unsolved cases.

Jeffrey Gafoor Case

A 20-year-old prostitute, Lynette Deborah White was found stabbed multiple times on Valentine's Day in Cardiff, Wales, in 1988. Multiple stab wounds were found on her chest and breasts, and other injuries were present on her face, stomach, arms, wrists, and inner thighs besides some defensive wounds on her hands. Further pathological examination revealed the presence of azoospermic semen in White's vagina and underwear. Some of the bloodstains found on White's clothing were from a foreign male source of AB blood type. A white male was found present in the locality at the time of incidence, but the investigating agency could not trace him. In late 1988, five men were charged with murder and after one of the longest court hearings, three men, Stephen Miller, Tony Paris, and Yusef Abdullahi were convicted in this case and were awarded life imprisonment. However, none of the

scientific evidence was found against these three individuals. The court of appeal quashed the convictions in December 1992.

The case was reopened in 2000, with a fresh set of forensic evidence found on the cellophane wrapper from a cigarette packet. As a DNA test was in use at that time, a DNA profile was generated and searched in the United Kingdom National DNA Database, but no match could be found. After the generation of the second-generation multiplex plus (SGM plus) test in 2002, the investigating agency was able to generate a reliable crime scene DNA profile. However, no match in the database could be generated for this profile also. Later, in 2003, a familial searching was conducted and a match was generated with a 14-year-old boy with a similar genetic profile. Besides, a rare genetic variant was also found in the family members of the suspect. As the 14-year-old boy was not even born at the time of the murder, his relatives were searched, and finally, police arrested his 38-year-old paternal uncle, Jeffrey Gafoor. Further, Gafoor confessed to the crime and he was awarded life imprisonment for the crime he committed in 1988. Later, in 2009, 13 police officers involved in this case were summoned with conspiracy and accusing 3 innocent persons in the murder charge. This case opens a new arena in the field of familial DNA analysis in cold cases.

Exoneration of Convicted Individuals Through DNA Analysis

Gary Dotson's Case

On 9 July 1977, 16-year-old Cathleen Crowell with dirt-stained clothes reported to a police officer near a shopping mall in Chicago that when she was approaching mall parking after work, 3 men came from behind in a car and grabbed her to the back seat. Two of them held her while the third person tore her clothes, assaulted her sexually, and scratched several letters on her stomach using a broken beer bottle. After her complaint, she was rushed to the hospital and a routine sexual assault examination was conducted to collect the semen samples. However, due to the unavailability of DNA technique, a matching could not be conducted. When police tried to prepare a sketch of the assailant using the information provided by Crowell, it was a perfect match with Gary Dotson except for his moustache. During trial, Crowell identified Dotson as her rapist and the court punished Dotson for sexual assault overlooking the testimonies provided by Dotson's friends.

In 1985, Cathleen Crowell shared with her New Hampshire pastor that she has sent an innocent man into jail. Fearing of becoming pregnant after having consensual sex with her boyfriend, she lied about an unknown person to be her sexual assailant. There was huge media attention to this matter and Dotson was forced to be released on parole. Later he was again sent to jail on charges of domestic violence. However, a DNA test was conducted which showed a mismatch between the DNA profile obtained from the seminal stains and the reference sample of Gary Dotson. Rather a conclusive match in DNA profile was obtained with the reference sample of David

Bierne, the earlier boyfriend of Crowell. Subsequently, based on the DNA test, Dotson was released from the prison becoming the first person in history to be exonerated by the DNA evidence.

Kirk Bloodsworth's Case

Kirk Noble Bloodsworth was sentenced to death in 1984 in a rape and murder case of a 9-year-old girl in Rosedale, Maryland. The evidence raised against him was the testimony of five eyewitnesses and his (probable) shoe marks on the victim's body. Kirk, maintaining his innocence, used to write letters to several innocence projects, nongovernmental agencies, and to others who would trust him for his innocence. Later in 1992, Kirk came across an article about how DNA fingerprinting aided in exonerating a convict of murder. He then appealed for DNA testing which was accepted and the testing proved his innocence. In 1993, Kirk was released. After a decade, a forensic biologist while revisiting the case found bloodstains on a piece of cloth that was used to wrap the victim. DNA testing on the bloodstain was conducted and got matched with Ruffner, who was declared convict of the rape and murder case.

Post-Conviction DNA Testing

With the rapid increase in the use of DNA analysis in forensic cases, the quality of criminal justice decision-making has increased tremendously. Throughout the world, DNA testing of already convicted individuals before the commencement of DNA test is being carried out to receive surprising outcomes. As per the report of the US National Institute of Justice, 28 post-conviction exonerations have been reported by 1996 (Johnson and Williams 2004). Many innocence projects have been launched to date throughout the globe to help the wrongly accused individuals due to the unavailability of DNA technology. A list of such projects is given in Table 1. There were many success stories of the post-conviction testing and one of the highlighted cases have been described below. Since 1989, more than 10,000 cases have been identified where the prime suspects were wrongly accused due to the unavailability of DNA testing. After the reopening of such cases, 25% of people were found wrongly accused and were excluded in the DNA result. However, due to the unavailability of evidence, 29% of such cases were closed.

Michael Shirley's Case

An 18-year-old Michael Shirley was arrested for the murder of Linda Cook in 1987. Cook was killed by stamping on her head and neck besides sexual assault. The semen swabbed from Cook's vagina, vulva, and anus showed a similar blood group as that of Shirley's. A distinctive shoe mark print present over Cook's belly also matched with a shoe in possession with Shirley. Shirley was also having some scratches on his body and bloodstains were found on his trouser, which was advocated

Table 1 List of Innocence projects operating throughout the world

Sl. No.	Name of the Project	Web-link
1	The innocence project	https://innocenceproject.org/
2	California innocence project	https://californiainnocenceproject.org/
3	The innocence project of Texas	https://innocencetexas.org/
4	Midwest innocence project	https://themip.org/
5	Great north innocence project	https://www.greatnorthinnocenceproject.org/
6	Illinois innocence project	https://www.uis.edu/illinoisinnocenceproject/
7	Innocence project London	https://www.gre.ac.uk/las/law-criminology/innocence-project
8	Innocence project New Orleans (IPNO)	https://ip-no.org/
9	Georgia innocence project	https://www.georgiainnocenceproject.org/
10	Mid-Atlantic innocence project	https://exonerate.org/
11	Innocence project of Florida	https://www.floridainnocence.org/
12	The New England innocence project (NEIP)	https://www.newenglandinnocence.org/
13	Wisconsin innocence project	https://law.wisc.edu/fjr/clinicals/ip/
14	Ohio innocence project	https://law.uc.edu/real-world-learning/centers/ohio-innocence-project-at-cincinnati-law.html
15	Washington innocence project	https://wainnocenceproject.org/

to be originated due to the murder he committed. DNA profiling test could not be conducted at that time due to the availability of fewer semen materials, which was insufficient to conduct RFLP-based DNA testing (Johnson and Kotowski 1996).

When the forensic DNA analysis saw technological advancements in the form of PCR amplification followed by detection of smaller and degraded samples, the stabs taken from Cook's body and stored since 1986 were subjected to this examination. A mixed profile could be obtained from these swabs; however, the lack of reference samples of the suspect and the victim provided an inconclusive result (Johnson and Williams 2004). In 2001, the oral swabs collected from both of them were examined and the result showed the presence of a foreign DNA profile that did not match with that of Shirley's. Based on this result, Shirley appealed against his sentence and finally, his conviction was quashed in July 2003. This case highlights the necessity to make suitable provisions to allow those wrongly convicted to utilize DNA evidence.

Promise of NGS Technology in Solving Forensic Cases

The currently considered gold standard, the capillary electrophoresis-based DNA analysis detects length variation in short tandem repeat (STR) markers. However, this technique has some limitations including the inability to analyze multiple

genetic polymorphisms in a single reaction, low-resolution of the currently used markers, loss of useful genetic information in degraded samples, and low resolution in mtDNA and mixture analysis (Yang et al. 2014). In this regard, next-generation sequencing (NGS) has been introduced as a boon to forensic analysts to perform such tests more robustly. Though the NGS technology is in its preliminary stage of use in most forensic DNA laboratories, it has shown a huge promise in solving complicated cases, that are previously unsolved by CE-based techniques.

The first case to be solved by the NGS technique is reported from the Netherlands. In a case of sexual assault, the accused was initially acquitted based on an inconclusive DNA result obtained from the CE technique. When the prosecutor was granted an appeal, the exhibit containing remaining biological samples was reexamined using the massively parallel sequencing (MPS) technique. Such a test conducted at the Leiden University Medical Center resulted in highly informative DNA profiles and a proper match could be obtained in this case (de Knijf 2019). With this development, more difficult criminal cases are expected to be solved by the NGS technique in near future.

Conclusive Remarks

Over the years, DNA fingerprinting technology has been proved to be useful as the most irrefutable evidence in the criminal justice system. The technique has been tested time and again in the courts and in this process, the technique has only evolved. From the early days of restriction fragment length polymorphism (RFLP) technique and PCR-based STR technique, the technology has further evolved to NGS-based analysis nowadays. Whenever the world faces any problem, maybe the cases of sexual assault, identity, mass disaster, military identification, child trafficking, DNA technology has come as a rescue to solve every case. Looking into the glorious history of the technique in solving criminal cases and the availability of advanced DNA techniques in hand like NGS, the technology will continue to help mankind and the justice delivery system beyond doubt.

References

Ballantyne KN, Keer V, Wollstein A, Choi Y, Zuniga SB, Ralf A, Vermeulen M, de Knijff P, Kayser M (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6:208–2018. https://doi.org/10.1016/j.fsigen.2011.04.017

Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Conneally PM, Duceman B, Eisenberg A, Forman L, Kidd KK, Leclair B, Niezgoda S, Parsons TJ, Pugh E, Shaler R, Sherry ST, Sozer A, Walsh A (2005) DNA identifications after the 9/11 world trade center attack. Science 310:1122–1123

Butler JM (2010) Fundamentals of forensic DNA typing. Academic, London

Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. Journal of Forensic Science 48:1054–1064

Caglia A, Stefanoni P, La Rosa A (2011) Cold cases: new technologies for DNA analysis allow the reopening and solution of unsolved cases. Forensic Science International: Genetics Supplement Series 3:e230–e231. https://doi.org/10.1016/j.fsigss.2011.09.001

Clark M, Gill J, Sasinouski K, McGuire A (2019) Cold case homicides: DNA testing of retained autopsy sexual assault smears. J Forensic Sci. https://doi.org/10.1111/1556-4029.14023

de Knijf P (2019) From next generation sequencing to now generation sequencing in forensics. Forensic Sci Int Genet 38:175–180

Dimo-Simonin N, Grange F, Brandt-Casadevall C (1997) PCR-based forensic testing of DNA from stained cytological smears. Journal of Forensic Science 42:506–509

Foran DR, Starrs JE (2004) In search of the Boston strangler: genetic evidence from the exhumation of Mary Sullivan. Med Sci Law 44:47–54. https://doi.org/10.1258/rsmmsl.44.1.47

Geis G, Bienen LB (1998) Crimes of the century: from Leopold and Loeb to O.J. Simpson. Northeastern University Press. p 174. ISBN 978-1-55553-360-1

Gill P (2002) Role of short tandem repeat DNA in forensic casework in the UK—past, present, and future perspectives. BioTechniques 32:366–385

Jakovski Z, Ajanovska RJ, Stankov A, Poposka V, Bitoljanu N, Belakaposka V (2017) The power of forensic DNA data bases in solving crime cases. Forensic Sci Int Genet Suppl Ser 6:e275–e276. https://doi.org/10.1016/j.fsigss.2017.09.085

John Aitken R, Koppers AJ (2011) Apoptosis and DNA damage in human spermatozoa. Asian J Androl 13:36–42. https://doi.org/10.1038/aja.2010.68

Johnson ED, Kotowski TM (1996) Chemiluminescent detection of RFLP patterns in forensic DNA analysis. Journal of Forensic Science 41:569–578

Johnson P, Williams R (2004) Post-conviction DNA testing: the UK's first 'exoneration' case? Sci Justice 44:77–82. https://doi.org/10.1016/S1355-0306(04)71692-5

Kayser M (2017) Forensic use of Y-chromosome DNA: a general overview. Hum Genet 136:621–635. https://doi.org/10.1007/s00439-017-1776-9

Kim YT, Heo HY, Seo TS (2018) Advanced short tandem repeat genotyping for forensic human identification. In: Chang HN (ed) Emerging areas in bioengineering. Wiley-VCH Verlag GmbH & Co. KGaA. ISBN: 9783527340880

Leclair B, Frégeau CJ, Bowen KL, Fourney RM (2004) Enhanced kinship analysis and STR-based DNA typing for human identification in mass fatality incidents: the Swissair flight 111 disaster. Journal of Forensic Science 49:939–953

Maguire CN, McCallum LA, Storey C, Whitaker JP (2014) Familial searching: A specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives—The UK experience. Forensic Sci Int Genet 8:1–9. https://doi.org/10.1016/j.fsigen.2013.07.004

Marks K (2009) New DNA Technology for Cold Cases. Law and Order 57(36–38):40–41,43

Mattick JS, Dinger ME (2013) The extent of functionality in the human genome. HUGO J 7. https://doi.org/10.1186/1877-6566-7-2

Menotti-Raymond MA, David VA, O'Brien SJ (1997a) Pet cat hair implicates murder suspect. Nature 386:774

Menotti-Raymond M, David VA, Stephens JC, Lyons LA, O'Brien SJ (1997b) Genetic individualization of domestic cats using feline STR loci for forensic applications. Journal of Forensic Science 42:1039–1051

Panneerchelvam S, Norazmi MN (2003) Forensic DNA profiling and database. Malaysian Journal of Medical Sciences 10:20–26

Pearsall A (1989) DNA printing: the unexamined "witness" in criminal trials. California Law Review 77:665–703. https://doi.org/10.2307/3480565

Rankin DR, Narveson SD, Birkby WH, Lai J (1996) Restriction fragment length polymorphism (RFLP) analysis on DNA from human compact bone. Journal of Forensic Science 41:40–46

Reif W (2002) September 11: how they identified the victims. The Lancet 360:807–808. https://doi.org/10.1016/S0140-6736(02)09931-2

Saad R (2005) Discovery, development, and current applications of DNA identity testing. Proc (Baylor Univ Med Cent) 18:130–133. https://doi.org/10.1080/08998280.2005.11928051
Sensabaugh G, Kaye DH (1998) Non-human DNA evidence. Jurimetrics 39:1–16
Sobrino B, Carracedo A (2005) SNP typing in forensic genetics: a review. Methods Mol Biol 297: 107–126
Thompson WC (1996) DNA evidence in the O.J. Simpson trial. University of Colorado Law Review 67:827–857
Yang W, Kang X, Yang Q, Lin Y, Fang M (2013) Review on the development of genotyping methods for assessing farm animal diversity. Journal of Animal Science and Biotechnology 4. https://doi.org/10.1186/2049-1891-4-2
Yang Y, Xie B, Yan J (2014) Application of next-generation sequencing Technology in Forensic Science. Genomics Proteomics Bioinformatics 12:190–197

Role of the Molecular Anthropologist in the Forensic Context

16

Elena Pilli

Contents

Abstract

The greatest strength of forensic science is the scientific multidisciplinary approach used to support investigation activity in solving and explaining crimes. There are diverse scientific disciplines, including anthropology, involved in the forensic process. The crucial purpose of the forensic anthropology discipline is the identification of human remains when nearly or completely skeletonized remains are found and a standard soft tissue autopsy can no longer be performed. Various circumstances, such as those involving war crime victims, mass disaster events, homicides, or missing persons, can require establishing the identity of the deceased individual. In all these situations in which challenging samples such as skeletal remains are found, the joint activity of physical and molecular anthropologists can be effective in supporting the identification process. Forensic molecular anthropologists are able to exploit the expertise achieved from the

E. Pilli (✉)
Department of Biology, University of Florence, Laboratory of Anthropology - Forensic Molecular Anthropology Unit, Florence, Italy
e-mail: elena.pilli@unifi.it

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_61

analysis of highly degraded samples, so they can be an important allies of physical anthropologists in challenging situations. This is highly relevant in cases where a DNA profile from aged/compromised bones must be obtained to compare it with that of relatives and where other information, such as age at death, sex, ancestry origin, and so on, has to be provided to confirm the hypotheses of anthropological examination and/or when antemortem data are limited to physical description and/or photographs.

Keywords

Forensic molecular anthropologists · Challenging samples · Skeletal remains · Identification process · DNA

Introduction

Criminalistics or, in its broadest sense, forensic science, generally described as the application of scientific principles/methods to recognition, documentation, collection, and analysis of physical evidence from a crime scene, was defined by Paul L. Kirk (1902–1970), one of the most noted criminalists of the twentieth century, as "an occupation that has all of the responsibilities of medicine, the intricacy of the law, and the universality of sciences." Criminalistics includes numerous diverse disciplines such as chemistry, medicine, biology, physics, psychology, and many more. The scientific universal-multidisciplinary approach applied to a forensic context can be considered the greatest strength of forensic science in resolving real crime cases. The current practice of using scientific disciplines in apprehending suspects, adjudicating cases, and applying judgments is unequalled among the various tools available to law enforcement. Although forensic science has a long history dating back hundreds of years, attention to criminalistics has never been as great as in the last decade. It is possible that the advent of new technologies, such as DNA testing, the analysis of scientific evidence in processes of great media impact, and the exponentially increased interest in entertainment programs and the proliferation of TV series, has contributed to increasing the curiosity about and knowledge of forensic science by the public. However, despite popular perceptions, the discipline of forensic science is remarkably complex, characterized by accurate scientific measurements and rigorous analytical and deductive reasoning and supported by continuous research activity and careful validation procedures. The field of forensic science is constantly evolving; therefore, new methods and techniques are continuously being developed and validated for forensic purposes. Among the diverse scientific disciplines that support forensic activity is anthropology. Forensic anthropology (FA), defined by the American Board of Forensic Anthropology (ABFA) (The American Board of Forensic Anthropology, http://theabfa.org/ accessed 21 December 2020) as "the application of the science of physical or biological anthropology to the legal process," is a traditional discipline that aids coroners, police, and lawyers in gathering, examining, and interpreting bone and skeletal

fragment evidence in criminal investigations, civil cases, missing persons, disaster victim identification (DVI), and humanitarian and human rights abuse cases (Ubelaker et al. 2019a). Forensic anthropologists have played key roles in the historical development of forensic science applications, and their knowledge has been used in DVI for over a century. However, it was not until 1970 that Thomas Dale Steward (American anthropologist) underlined the importance of including anthropology in the identification process. Also, forensic anthropology has changed radically over the years. Some of these changes, fueled by new technology, have led forensic anthropologists to become involved in the detection, recovery, and analysis of human remains, fragment recognition, differentiation of human from non-human remains, estimation of species, ancestry, age, gender, stature, reconstruction of face based on medical imaging technology of skeleton, and personal identification of human skeletal remains (Lei et al. 2019).

The Identification Process

The identification process of victims from different contexts such as war grave recovery and cases of political, ethnic, religious violence, and mass disaster and the dignified management of their remains are moral duties in all contexts and the ultimate aim of forensic practitioners for both legal aspects and social and religious matters. The determination of the identity of unknown decedents is considered a fundamental step in forensic activities, for example, to aid criminal proceedings in convicting the guilty and exonerating the innocent and to facilitate settlement of inheritance and preserve the right of a partner to remarry, and it is considered an essential mark of respect for the deceased and the surviving family members and friends. The identification process has emerged as a multi-step process and is a very complex activity. This activity is heavily conditioned by the varying conditions in which bodies are discovered, depending on the cause of death and any subsequent postmortem trauma the bodies experience. When deaths occur as a result of armed conflict or mass disaster, the identification process can be more challenging. A multitude of complications can arise, such as deposition of bodies in clandestine mass graves, limited contextual information, a long interim between death and recovery/identification of the victim, and also partial or incomplete antemortem data. In addition, the complexity of identification inevitably increases as the number of individuals to be identified can range from one to thousands.

Forensic anthropology is the science that primarily deals with the examination and identification of human skeletal remains in the forensic context. However, forensic anthropologists do not practice in isolation, but work as part of a team of forensic practitioners, which typically includes odontologists, radiologists, pathologists, fingerprint examiners, molecular anthropologists/biologists, mortuary technicians, and photographers. The contributions of forensic anthropologists to identification are especially needed and valuable in the analysis of extensively decomposed and skeletonized human remains when a standard soft tissue autopsy can no longer be performed. By establishing the biological profile of human

remains – intended not only as the identification of all the biological characteristics of the deceased (e.g., sex, age at death, stature, and ancestry) but also personal identification, trauma and taphonomic analysis, the postmortem interval (PMI) estimation, and the application of anthropological knowledge to the investigation of mass disasters – forensic anthropologists contribute to positively identifying human remains through the wealth of supplemental information provided and/or directly through the careful examination and assessment of several anatomical features of the skeleton and comparison with known antemortem information (Ubelaker et al. 2019b). In addition, their assistance at the scene will help to prevent the collection of non-human remains and non-osseous items and to select the most appropriate skeletal elements for DNA analysis, reducing the number of samples for subsequent analyses and the generation of unessential data. In all challenging situations and when the information from the anthropological examination is limited, DNA analysis is required, and molecular experts (anthropologists/biologists) can be important allies of physical anthropologists by supporting the identification process.

Bones and teeth are more durable biological material than other remains, and therefore, in many forensic cases, they represent a potential source of genetic material remaining after exposure to environmental conditions and traumatic events and after a significant time has passed since the death of the individual. The ability to recover and analyze DNA from highly degraded skeletal remains represents one of the most significant challenges for molecular anthropologists and forensic science. The analysis used for identification of human remains has some commonalities with the methodology employed for analysis of crime scene evidence and kinship testing. However, there are diverse aspects of the analytic process that are specific to remains identification, especially when fragmented and highly degraded bone samples are to be examined. In all these cases, professional figures (forensic molecular anthropologists) who are experts at handling/analyzing highly degraded samples are essential allies of physical anthropologists and forensic experts in the identification process.

In the classical approach, which is the most widely used approach, DNA is used to support or refute a hypothesis of identity generated, for example, though visual recognition, eye-witness reports, and any other relevant data. In contrast, in the opposite approach, DNA is central in the identification process, and a database of postmortem data generated from the human remains is compared with a database of antemortem data (typically from relatives). When a potential match is generated by cross-checking the two sets of data, a numerical quantification to explain the strength of a given match is provided. In order to reduce the potential for misidentification, further checks are carried out, for example, ensuring that the biological profiles of the human remains are sufficiently similar to those of the identified person (Puerto et al. 2014). This inevitability necessitates laborious collection of further antemortem information for the missing persons, such as stature, age, sex, and possible pathologies, from the relatives' interview.

DNA from Skeletal Remains

DNA decay begins immediately after cells die when enzymes from the organism and then bacteria and fungi present both within and outside the body start breaking down cellular structures and DNA molecules. However, the rigid structure and chemical composition of bones and teeth provide protection against DNA degradation. The physical barrier against bacteria and fungi that the hard tissues afford and the chemically bound of DNA molecules to the hydroxyapatite that stabilizes the DNA provide some protection against DNA degradation. Therefore, analyzable DNA can persist in the skeletal elements much longer than in the soft tissues of the body. When working with bones and skeletal fragments, the main potential issues are low amount of starting molecules, degradation of DNA, and the presence of polymerase chain reaction (PCR) inhibitors.

DNA Damage

DNA damage is a complex multifactorial process (e.g., due to chemical and physical factors), which gives rise to very challenging samples from which to gain a genetic profile for identification. Factors of DNA damage include oxidation, hydrolysis, pyrimidine dimers, cytosine deamination, and DNA–DNA and DNA–protein cross linkages. As previously described (Pilli et al. 2021), the most common factors than others in forensic samples are as follows:

1. **Hydrolysis:** The main cause of DNA damage is the breaking of double helix into small fragments. The most susceptible bond to damage in DNA is the N-glycosyl bond. The cleavage of this bond results in the loss of a base leaving an apurinic/apyrimidinic (AP) site that forms a nick. Once a nucleotide is released, the AP can undergo a chemical rearrangement that promotes occurrence of strand breakage. If the average DNA fragment length is reduced to below 200–250 bp, a significant loss of genetic information can occur due to the lack of suitable template DNA for amplification. As a result, partial or no DNA profiles can be obtained.
2. **Oxidation:** Many other lesions are mediated by free radicals. Reactive oxygen species, such as superoxide and hydrogen peroxide, are created by ionizing radiation or microbial metabolism of anaerobic bacteria that colonize postmortem (PM) tissue. Oxidative attack on the DNA bases breaks carbon–carbon double bonds of both pyrimidines and imidazole ring of purines, leading to fragmentation. Damage includes modifications of sugar residues, conversion of cytosine and thymine to hydantoins, removal of bases, and cross linkages. This lesion leads to replication block during PCR: standard Taq DNA polymerase cannot bypass these lesions.
3. **Pyrimidine dimers:** When DNA is exposed to ultraviolet light, pyrimidine dimers (pairs of adjacent thymine (T) or cytosine (C) bases) are created.

Photochemical exposure induces the formation of covalent bonds between two thymines when they are close on the double helix. These linkages cause DNA polymerase to stall and arrest replication during PCR. Other photo-lesions are formed on UV exposure, such as purine and pyrimidine oxidation products.

4. **Environmental DNA damage:** Many molecular taphonomic processes influence the preservation of biological material at macroscopic (sample) and microscopic (proteins and DNA) levels depending mainly on environmental and time factors. While degradative processes accumulate with time, the environmental conditions, such as temperature, humidity, pH, and soil composition, can vary the rate and aggressiveness of the degradation. Diverse environmental factors can act to generate differential preservation in different skeletons, different bones within the same skeleton, and even variations in DNA quality and quantity across the same bone. Therefore, the depositional environment plays a greater role in molecule degrading than the absolute age of the DNA sample.

   Environmental Conditions:

   a. *Temperature*

      All degradation chemical reactions are heavily influenced by temperature; as a result, the increase of 10 °C in the temperature can accelerate the reaction rate by two- to threefold. Additionally, warmer temperature promotes micro-organism growth, which contributes to biological decomposition. The microbes digest the protein component of bone, making the DNA more prone to damage. In addition, they produce enzymes that fragment DNA molecules. However, although some publications suggest that in some cases, mild heating could facilitate an increase in DNA yield from hard tissues, cooler temperature is preferable to preserve the DNA integrity.

   b. *Humidity*

      The presence of humidity in the depositional environment can influence biological decomposition. Generally, water molecules participate in hydrolytic reactions that act to fragment and modify DNA molecules; therefore, the more the humidity in the depositional environment, the greater the likelihood of bone degradation, which, in turn, contributes to DNA loss. On the contrary, a burial in peat bog (a particular water environment) may be beneficial to DNA preservation due to the low presence of oxygen and high salt concentration that could slow DNA degradation by reducing microbial activity.

   c. *pH*

      Also the pH of the depositional environment influences the biological decomposition, the rate of microbial decomposition and, consequently, affects the degree of DNA damage since the rate of acid catalyzed depurination leading to strand breaks is pH dependent. Bones and teeth reach a chemical equilibrium with the depositional environment via mineral leaching and the uptake of different solutes from the soil. This process can lead to bone degradation that can impact the rate of DNA damage.

d. *Soil Composition*

The chemical composition of the soil can interfere with the genetic analyses conducted on skeletal remains due to the presence of solutes, such as tannins and humic acids that may be coextracted with the DNA and inhibit subsequent analytical steps.

5. **Bone type**: In addition to the many environmental factors, intrinsic factors, such as bone type and density, can play a role in the process of DNA decomposition as well. Bone size and construction can impact skeletal DNA preservation. Larger bones tend to survive better and are, therefore, most available for genetic analysis. Moreover, until recently, the dense cortical portions of lower limb bones and the tougher tissues of the teeth were considered to be more reliable skeletal elements in generating DNA profiles than those spongy. To date, however, several studies demonstrated that petrous bone represents the better skeletal element in terms of DNA amount and preservation than femur and tooth for human identification purposes even from ancient samples, probably due to the high density of this skeletal element associated with resistance to damage and reduced bacteria-mediated DNA decay and other PM DNA decay. Recently, Sirak et al. (2020) identified human auditory ossicles as another optimal skeletal element for ancient DNA (aDNA) analysis.

It is important to note that none of these environmental and intrinsic factors operate in isolation, and they may work together or in opposition regarding DNA degradation: the death will, therefore, be affected by a complex interaction between these several variables. As a result of all these different chemical and physical factors, the remaining DNA fragments that can be extracted from this biological material are short in length, and their integrity is partly compromised.

Contaminating DNA

Another important issue associated with the analysis of degraded samples is the contamination. Exogenous DNA molecules from the environment or from people who handle the samples can easily outcompete the small amount of endogenous DNA. Contamination even with low amounts of high-quality DNA from working staff during discovery of skeletal human remains, exhumation, and laboratory personnel activities or reference samples (such as those of the victim, the suspect, and the relatives are typically available as blood stains or buccal swabs and contain large amount of high-quality DNA) can occur, and it can make it difficult to gain reliable results. Contaminating human DNA can be found in people's dead skin cells, hair, saliva, sweet, and blood. In addition, laboratory consumables and reagents can be contaminated by human DNA during production in manufacturing facility. Additional DNA contamination, such as cross-contamination among samples or DNA extracts could occur during experimental processing if precautions are not in

place. Therefore, as proposed by ancient approach, when working with these types of samples, good practices, targeted to reduce/minimize and control contamination before and after the arrival of such specimens in the laboratory, should be mandatory. On the basis that there should be laboratories dedicated to this type of samples (i.e., dedicated to skeletal elements) and that all DNA extractions and PCR involving the samples should be carried out in a laboratory physically separated from the laboratory in which PCR cycling and post-PCR analyses were performed, some of the precautions include the following:

1. The use of protective clothing such as disposable laboratory coats (multiple gloves changes during samples handling) during excavation and when handling specimens in the laboratory.
2. Laboratory cleaning procedures with bleach and UV irradiation of hoods and laboratory bench surfaces before and after usage.
3. Processing the question samples prior to known samples and physically isolated pre-PCR facility. In addition, some laboratories control movement of laboratory personnel between spaces (a technician is not permitted on the same day to return to pre-PCR area if he/she has already entered the post-PCR area).
4. Reagent blanks and negative controls are run to monitor level of exogenous DNA in reagents, laboratory, environment, and instruments.
5. Decontamination of reagents/tools (bleach and/or UV irradiation) and specimens by removing their surface should be routine.

Two are the key points to keep in mind: 1. protection of the samples from DNA contamination and 2. prevention of further endogenous DNA degradation. Therefore, as proposed for aDNA analysis, it is good practice that specimens are not washed with water (water contains contaminating bacterial DNA and can deeply penetrate into the samples and cause unnecessary hydrolytic damage to the endogenous DNA) and stored correctly (samples should either be completely dried to avoid further contamination with microbial DNA or stored in a cold, dry place as soon as possible).

PCR Inhibitors

Inhibitors are exogenous molecules to the samples of interest, such as humic acids, fulvic acids, tannins, and microbial DNA, or present within the sample itself such as heme in blood, melanin in hair, and fatty acids, collagen, and Ca^{2+} in bone, and are mostly co-extracted with the target DNA. The presence of inhibitory molecules in a sample may offer a challenge for PCR amplification. These factors are most common causes of amplification failure even if adequate amounts of DNA are present. Inhibitors may negatively affect cell lysis during extraction or reduce polymerase activity. Given the importance of removing PCR inhibitors from DNA extracts, it is not surprising that different strategies were developed to eliminate or overcome PCR inhibitory effects. These include dilution of the extracts, the addition of more Taq

DNA polymerase, and the use of different DNA polymerases and sample cleanup devices. In addition, PCR enhancers such as bovine serum albumin, betaine, and PCRBoost™ were tested to minimize or prevent inhibition effects. The presence of inhibitors can be detected via quantitative real-time PCR (method of choice for quantifying the amount of DNA in a sample) using an internal PCR control (IPC) in the reaction. As the IPC is co-amplified with DNA samples, a reduction in the reaction efficiency and a delay in amplification may be due to the presence of inhibitors in a particular sample.

Role of Forensic Molecular Anthropology

Just as forensic anthropologist, forensic molecular anthropologists (FMA) use the science/expertise, methodology, and technology of molecular anthropology – the ancient DNA approach – to handle highly degraded bone samples and produce a DNA profile or any other genetic information useful for forensic identification purposes. FMA studies skeletal element DNA in order to genetically identify individuals; provide any further genetic information about the remains, such as phenotypic characteristics and ancestry origin; and reassociate disparate skeletal elements when anthropological studies are not permitted. FMA activity is not only applied to single cases, its expertise can also be useful in the study of war crime victims, mass disasters, and unidentified person cases. More significant and extremely valuable is the contribution of forensic molecular anthropology when severe fragmentation, decomposition, burning, or commingling has occurred, and at crime scenes where human remains have to be retrieved and properly collected. Unlike fingerprints and dental features, DNA-based identity testing is not restricted to any particular body part. Therefore, the ability to generate a DNA profile permits reassociating separated remains typical of mass graves or disasters such as plane crashes or explosions.

The Choice of Skeletal Elements and Their Sampling

The choice of which skeletal elements to sample depends on the aim of the study and obviously what is found in the burial site. However, not all skeletal elements are equally effective at preserving DNA. Goodwin (Goodwin 2017) reported that "data are available from a large number of cases and provide a hierarchy of preference when choosing which element(s) to use for DNA analysis." The selection of specimens should be based on their good preservation and minimal diagenetic alteration. In addition, as DNA analysis is destructive, it is a good practice, also for forensic purposes, not to sample parts of the skeleton that could be informative for morphological and pathological studies, unless a DNA analysis of diseases is required. As recommended by the International Society for Forensic Genetics DNA Commission (Prinz et al. 2007), dense cortical bones such as tibia and femur should always be the first choice of sampling. Due to their density, which provides a

protective crystal matrix endogenous DNA, the tibia and femur are less vulnerable to contaminating DNA (Campos et al. 2012). Recently, a team of molecular anthropologists (ancient DNA experts) has demonstrated that human petrous bone, located at the base of the skull between the sphenoid and occipital bones, has yielded more endogenous DNA than tooth cementum (Hansen et al. 2017) and other bones, presumably due to the very high density of petrous bones associated with reduced bacteria-mediated DNA decay and other postmortem DNA decay. Kulstein et al. (2017) and Pilli et al. (2018) demonstrated the possibility of producing an identification profile from petrous bone samples for forensic purposes, even starting with highly degraded samples, and Gonzalez et al. (2020) showed that DNA extracted from the petrous bones has more long fragments likely to be amplified than teeth. This result implies that it yields complete STR (short tandem repeat) profiles more often than teeth, and petrous bones can also be considered the ideal substrate for forensic genetic investigation.

DNA analysis involves invasive and destructive sampling procedures that are often incompatible with anthropological, anatomical, and bio-archeological analyses requiring intact skeletal remains and totally unacceptable when handling bones that are valuable for anthropological or religious reasons. Therefore, in order to preserve the original structure of the skull for future morphological/anthropological studies, molecular anthropologists focused their efforts on removal of petrous bone without changing the integrity of the cranium by cutting the connection between petrous and temporal bones using a surgery blade. Subsequently, they focused on developing a minimally invasive method that reduces the loss of genetic data associated with the use of other skeletal elements and enables the combined craniometric and genetic study of individuals with archeological, cultural, and evolutionary value. In addition, Pluta et al. (2016) have demonstrated the possibility of obtaining an STR profile from DNA isolated from human bones using a technique that is non-destructive to the bone surface.

Although teeth also provide a valuable source of DNA, little is known about which region is best in terms of DNA yield. Different methods are described in the literature to collect material from teeth for DNA analysis. These procedures include grinding the whole tooth or the root, sectioning the tooth, accessing the pulp cavity and dentine via the crown, and incubating the whole tooth overnight, after external surface decontamination in an extraction solution. In order to obtain the maximum a DNA yield possible, tooth destruction is recommended. However, the disadvantage of this destructive activity is associated with the impossibility of carrying out further analysis, which may be required. In these cases, preserving the specimens for potential subsequent morphological/radiographical or epigenetic dental trait studies is essential. Therefore, as proposed for ancient and museum remains, tooth sampling can be carried out by accessing the pulp cavity via the apical end of the root using a dental micro-drill at low speed, and the inside of the tooth can be drilled until the needed amount of powder is produced for extraction. In this way, the teeth could be replaced in the jaw socket after sampling.

DNA Analysis for Human Identification

Molecular analysis has become a common practice for identifying human remains, providing information for kinship relationships, and helping to restore family links. However, the ability to generate DNA profiles from skeletal remains is a challenge in numerous forensic and archeological contexts due to the previously discussed issues that reduce the possibility of obtaining reliable results.

In addition to choosing of the "best" skeletal element in terms of DNA yield, the choice of DNA extraction method represents a crucial tool in the identification of human remains. Extraction of DNA from soft tissue is relatively straightforward, and several protocols can be applied. In contract, the hard tissues of bones present a greater technical challenge associated with their mechanical disruption prior to DNA extraction. Also, extraction of DNA from bone and tooth samples often requires dedicating DNA extraction methods different from those used for other types of biological samples. Various techniques are utilized by laboratories to extract and purify DNA from hard tissue, but all of them focus on maximizing DNA yield, minimizing any additional DNA damage, and removing any inhibitors that could be co-extracted with the sample DNA and interfere with subsequent analyses. An advantage that hard tissues present, which soft tissues do not, is that the surfaces can be cleaned using both chemical (typically bleach) and physical methods, that is, by the mechanical removal of the surface of the bone using, for example, a rotary sanding tool. This step consists of removing external contaminating DNA from the bone and teeth that would contribute to generating unreliable results. A commonly employed final step in cleaning is the exposure of skeletal elements to ultraviolet radiation. After surface decontamination, the hard tissues have to be pulverized prior to incubation in EDTA (Ethylenediaminetetraacetic acid) and/or an extraction buffer and proteinase K in order to dissolve the organic and inorganic portions of the tissue. Several approaches are available for the physical disruption of bone/tooth material, but in order to carry out a precise and accurate sampling without destroying the bone structure (especially the petrous bone) and producing a fine powder, the most commonly used with degraded bone samples is drilling the sample with a dental drill with disposable tips. Bone/tooth pulverization allows us to obtain a greater contact surface with the different chemicals employed in the DNA extraction process, encouraging the release of a greater amount of DNA from the hydroxyapatite mineral matrix. The amount of bone/tooth powder used for DNA extraction varies from laboratory to laboratory, as well as the methods used for DNA extraction. Obviously, if the bone is aged or has been subjected to extreme environmental conditions, more care has to be taken with the selection of the extraction method. Demineralization is a standard step when handling forensic bone samples: bone powder is incubated with EDTA, generally overnight but also for days, in order to break down calcium phosphate, a major component of the mineral portion of the bone, and leave decalcified material for subsequent extraction. This method, not used by molecular anthropologists, can be substituted for by total demineralization

with the addition of a detergent such as SDS (sodium dodecyl sulfate), SLS (sodium lauryl sulfate), and Tween 20 and proteinase K to the EDTA in order to completely break down the hard material and increase the yield of DNA (Dabney and Meyer 2019; Saragoni et al. 2020). In their work, Mckinnon and Higgins (2021) developed real-time PCR assays in order to investigate the effects of modified demineralization protocols on DNA yield from burned remains and demonstrated that retaining supernatant post-demineralization resulted in significantly greater DNA yields compared with discarding it, as proposed by the total demineralization protocol.

After demineralization, various methods proposed by several forensic laboratories can be employed, but not all are equally effective for forensic investigations because degradation of DNA molecules impedes the ability of forensic examiners to obtain adequate DNA profiles from burned and highly degraded samples. Current ancient DNA extraction methods have proven highly effective at starting from 0.05 g of powder (most published protocols for forensic analysis call for as much as 2.5 g to as little as 0.2 g of starting bone powder) and recovering short fragments, even as sort as 50 bp, via the use of commercial silica columns from archeological specimens (Dabney et al. 2013).

Comparing ancient and forensic DNA extraction techniques, Emery et al. (2020) demonstrated that the Dabney extraction method is a viable alternative for generating full and/or partial STR profile from burned human remains resulting from temperatures exceeding 550 °C. Therefore, their results suggest adopting aDNA extraction methods as an alternative to current forensic practices to improve DNA yields from challenging human remains.

After extraction, DNA is quantified using real-time PCR, and autosomal STR typing is performed using several commercial kits to determine individual genotypes of skeletal remains. To integrate autosomal STR analysis, Y chromosome and mitochondrial DNA (mtDNA) analysis can be used. Neither of these two markers alone can identify a person because multiple individuals in any given population can have the same Y chromosome/mtDNA, but both can either refuse matches or increase the significance of a match (Leclair et al. 2007).

Recently, new DNA profiling systems based on next-generation sequencing (NGS) have become available for forensic purposes. NGS technologies, widely used in molecular anthropology with degraded bone samples, permit a large number of genetic markers (including length and sequence polymorphism) to be characterized for each sample, thereby allowing us to obtain very highly discriminating results. However, the use of NGS technology is currently limited in forensic laboratories due to the high costs of machines and kits and the long running time that is currently still required by NGS systems: several days are needed from DNA extraction to data analysis.

Obtaining a DNA profile to identify human remains in itself is of limited use if antemortem data are not available for comparison. In order to provide the strongest evidence for identification, a direct reference sample has to be accessible for comparison with the missing persons' profile. In the USA, military personnel provide a blood sample that is stored as a reference and analyzed if the individual goes missing in action and needs to be identified. Moreover, a direct reference profile

could be obtained from biopsy and other archived medical samples belonging to the victim, if they can be located and permission for their use obtained, or from personal possessions, such as hairbrushes, toothbrushes, and razors, although reliably establishing the identity of the personal possessions can be challenging.

In most circumstances, such as armed conflicts and, other situations of violence, and in all cases in which no direct reference samples are available, antemortem data can only be obtained through relatives donating biological samples to the authorities carrying out the identification. Parents, children (better if the spouse is also available), and full siblings are typically the most valuable relatives.

Molecular Information When No Reference Data from Relatives Are Available

Currently, DNA profiling methods only compare a victim's DNA with DNA from direct reference sample or relatives. Where there is no hypothesis of identity, it is useful for forensic experts to be able to predict what the person of interest looks like by analyzing the DNA from their bones. In fact, in cases in which the primary methods of identification, such as fingerprint, DNA (STR typing), and dental records, are not applicable due to body decomposition and to the lack of antemortem data, as in the case of the deaths of migrants in the Mediterranean Sea or disasters in developing countries, other means of identification need to be used to aid the establishment of the identity of unknown decedents. As proposed by Caplova et al. (2018) in their review, alternative methods of identification based on the use of physical appearance of the deceased have to be investigated when the available antemortem information for comparisons consists only of physical description and photographs. Many cases worldwide require alternative techniques of personal identification, especially with the recent free movement and migration, in which the source of antemortem data may be limited to photographs. The physical appearance of a person (the human face and body) displays numerous individualizing features, such as age, sex, and ancestry. In such a context, the analysis of DNA extracted from bone specimens can be very helpful in providing additional information to support/confirm forensic anthropology data in the study of human remains and in creating the biological profile in order to facilitate the identification process. A biological profile is constructed based on a number of important physical factors: sex identification, age at death, pigmentation, height, biogeographical ancestry, and craniofacial features.

Sex identification: Sex estimation is one of the primary steps for constructing the biological profile of skeletal remains, leading to their identification in a forensic context. Traditionally, sex determination of human skeletal material is performed by assessing sexually dimorphic traits mainly of the pelvis and skull (Dayal et al. 2008). Nevertheless, many forensic cases do not have access to a complete skeleton. For this reason, many studies have aimed at developing standardized methods for sex estimation from other postcranial bones, such as the vertebral

column (Garoufi et al. 2020) and humerus (López-Lázaro et al. 2020). Due to the fact that morphological methods have a relatively high success rate in the case of remains where the pelvis is well preserved, when the adult skeletons and bones are in good condition of preservation and, the morphometric variability in the population to which they belong is known, only in some cases skeletal sexing may result in unreliable results. Advances in the field of molecular genetics have provided sensitive methods for sex determination that have become a valuable tool in forensic casework. In most commercially available multiplex PCR kits for human identification, the amelogenin system is included with amplicon sizes of 106 and 112 bp for X and Y chromosomes, respectively. However, huge discrepancies have been observed with amelogenin-based sex determination, mostly due to X and Y deletion in the population and mutation in primer binding sites, and many alternative markers/techniques have been evaluated (Dash et al. 2020). In addition, the amelogenin assay is not entirely effective in the analysis of highly degraded samples, resulting in amplification failures and incorrect sex assignments. To overcome these problems, alternative molecular-genetic assays were developed (Fazi et al. 2014; Boonyarit et al. 2014). However, it was observed that, with severely degraded DNA, such as that which is present in heavily decomposed remains, these alternative methods have also failed to produce useable results. A new robust tool based on real-time PCR amplification of short intergenic sequences ($\leq$50 bp) was developed for genetic sex determination in a forensic context (Madel et al. 2016). In addition, with the development of NGS technology, a new method that identifies sex by considering the number of reads in shotgun DNA sequencing data that align to the X and Y chromosomes has been proposed by Skoglund et al. (2013). However, this approach relies on at least 100,000 sequences mapping to the human genome for accurate assignment, a prohibitive condition for many badly preserved remains. Therefore, an alternative approach was developed by Mittnik et al. (2016) that takes into account the ratio of sequence alignments to chromosome X compared with the autosomes and which gives accurate results with as little as several thousands of reads mapping to the human genome.

Age at death: Evaluation of chronological age of the donor of an unknown sample plays an important role in forensic investigations in order to identify an individual (the applicability of law depends on the age of the person; sometimes, the identity and age of individuals are unclear such as in the migration cases) or human skeletal remains. For forensic anthropology, age estimation involves morphological analyses of skeletal features. The literature provides many articles that apply known macroscopic and microscopic techniques, divided into dental and skeletal ones, aimed at determining a range of age for the subject. As long as a body is still developing, age estimation can be more accurate and provide smaller error rates, as is commonly the case with sub-adult individuals. In such cases, dental development methods as well as skeletal growth methods can furnish accurate age ranges and, thus, more precise information concerning the age at death of the subject. However, once skeletal and dental development reached full maturity, age estimation is much more complicated because the

only parameters that can be used are based on the physiological degeneration observed in skeletal and dental structures with age. In particular, the articular degeneration of specific joints, which are not modified by pathological and occupational factors (auricular surface, pubic symphysis, ribs, as principal examples), is usually considered as well as dental wear or periodontal alterations (Zelic et al. 2020). Microscopic methods have also been developed for cases where macroscopy proves to be useless. It is, however, necessary to highlight that there are at present no reliable methods for the estimation of age at death of the elderly, and aging dead adults is still riddled with limits, significant errors, and population-specific references, which are limits that are necessary to be known and well explained when dealing with investigating authorities. Some chemical methods, such as combined aspartic acid racemization, provide precise age estimation on dental specimens. Recently, the potential of FTIR-ATR and chemometrics was positively evaluated to estimate age at death in human bone (Pedrosa et al. 2020). Also, biological approaches were investigated and several proteins and DNA epigenetic markers that predictably change over time were discovered; for more details, refer to "Forensic Analysis of Externally Visible Characteristics: Phenotyping" (Podini and Gettings 2021). The main advantage of these methods is that they can be applied to any tissue containing DNA. Due to the importance of identifying skeletal remains in the forensic context, as they are frequently the only available sources of DNA, age estimation models based on DNA methylation level have recently been developed successfully using bones and teeth (Correia Dias et al. 2020; Kondo et al. 2021). In addition, Marquez-Ruiz et al. (Márquez-Ruiz et al. 2020) demonstrated that telomere length has limited usefulness as a supplementary marker for DNA methylation-based age estimation in tooth samples. Even at older ages, DNA methylation appeared to be more informative in predicting ages than telomere length.

Pigmentation: Retrieving information about externally visible characteristics such as eye, skin, and hair color from DNA can be suitable to provide valuable leads to locate the putative relatives for possible final identification of the human remains via STR profiling. In their paper, Draus-Barini et al. (2013) demonstrated the general suitability of validated HIrisPlex test system to successfully analyze DNA from bones and teeth of different ages stored under various conditions. Subsequently, the system was also successfully applied to DNA extracted from skeletal remains of World War II, bones attributed by multiple evidence to King Richard III of England, and a femoral bone attributed to Jörg Jenatsch, a leading freedom fighter during the 30 Year's War in Graubünden, Switzerland. Transition from classical method to next-generation sequencing allowed the analysis of the whole set of HIrisPlex-S markers in a single reaction on bone samples with different levels of DNA degradation (Kukla-Bartoszek et al. 2020). In addition, an 8-SNP multiplex assay was developed and validated by Mushailov et al. (2015), and the use of challenging samples, including bones, was tested. In 2019, Gokhman et al. (2019) presented a method for successfully reconstructing skeletal morphology via DNA methylation patterns. For more details on the evolution of variable

pigmentation in the human population, pigmentation-informative markers, and pigmentation prediction tools, refer to Chap. 8 (Podini and Gettings 2021).

Height: Another relevant externally visible characteristic used to describe human remains and create a biological profile is stature. Stature estimation from a skeleton is relatively easy, since it is based on the length of long bones. Taking into account the rarity of the circumstances in which all the bones of a skeleton are available for anthropological analysis, other methods were developed on other skeletal districts, such as bones of the feet and hands or even from the cranium, which, however, provide a wider range of errors.

Adult height is a polygenic trait characterized by normal distribution, high heritability, and non-Mendelian inheritance. Currently, height is evaluated using various single nucleotide polymorphisms SNPs (Ralf et al. 2015). Genetic Investigation of Anthropocentric Traits (GIANT) Consortium has identified 697 DNA variants statistically significantly associated with adult height in Europeans (Wood et al. 2014). The same markers were also tested in a non-European population, but the prediction accuracy obtained was lower than that obtained in the Europeans (Jing et al. 2019). Besides pigmentation traits, no molecular prediction tests are currently available for any other externally visible characteristics due to limited knowledge of genes and predictive DNA markers.

Biogeographical ancestry: A fundamental feature when building a biological profile considers the geographic origin of the individual in analysis, especially in light of the increase in migration and the modern multicultural society. The cranium is commonly the most characterizing skeletal region for ancestry identification, according to specific characteristics that it may display when considering a specific geographic area of origin (prognatism, palate shape, and dental shape). Some studies, however, attempted to identify ethnicity from postcranial bones, such as the femur.

The problem of ancestry estimation is further aggravated, as previously mentioned, by the multiculturalism of modern society, where well-defined boundaries between populations are day after day more complicated to observe, and specific ethnic groups are always more difficult to find. The constant awareness of the need for local methods for biological estimations has created an increased diligence toward the research for skeletal collections useful for enlarging and strengthening the existing ones.

Since ancestry can be considered as the genetic inheritance, each individual receives from their ancestors, and from population members that have occupied the same place of origin for long periods of time, biogeographical ancestry analysis can be carried out on population genetic markers found in an individual that can signal their origin from a particular geographic region. Therefore, the best way to assign an individual to a certain population via genetic testing is to use ancestry informative markers (AIMs) – markers characterized by important differences in allele frequencies among populations (Bulbul and Kidd 2021). The biogeographical

origin prediction based on ancestry informative marker analysis can also provide valuable information in aged bone samples, as proposed by Romanini et al. (2015).

Craniofacial features: It is a well-known fact that facial reconstruction is not an identification method; however, it can provide useful information to more accurately describe facial variation than an eyewitness account and to help trigger a suspicion of identity. Craniofacial reconstruction employed in the context of forensic investigation aims at estimating the facial appearance associated with an unknown skull through the reconstruction of soft tissues for victim identification. It is a technique based on facial anatomy and its relationship to the underlying skull. However, some have suggested that inter-individual variation in facial morphology is, in most cases, primarily determined by genetic variation (Baynam et al. 2013; Hopman et al. 2014). The potential of DNA-based facial construction is forensically of great interest in investigation activity or in promoting the reinvestigation of cold cases. Normal facial shape is known to be highly heritable (Ng et al. 2017). In recent years, separate studies have highlighted several genetic loci associated with facial shape phenotypes (e.g., Cha et al. 2018; Xiong et al. 2019). However, it is currently not possible to use genomic data to develop an accurate sketch of a person's face, and no data is available in the literature on bone samples.

Conclusion

While various cultures and religions view death differently, most are cognizant that relatives and friends must be able to grieve for their loved ones and have a remembrance place. Therefore, the identification process of the victims from different contexts, such as war, cases of political, ethnic, or religious violence and mass disaster, is a moral duty in all contexts and the main purpose of forensic activity for legal, social, and religious matters. The identification process emerged as a very complex activity. This activity is heavily conditioned by the varying conditions in which bodies are discovered, depending on the cause of death and any subsequent postmortem trauma the bodies experience. A skeleton is often all that remains of a body, and at times, the information the anthropological analysis can provide is limited or needs to be confirmed. In all these cases, DNA analysis is required and essential to identify unknown skeletons or provide information about the physical appearance of the deceased to compare with the physical description of acquaintances or photographs. Because in many forensic instances bones often represent the only available material that can be used for personal identification, the choice of the best available skeletal element and of suitable and reliable protocols for degraded bone samples is fundamental to achieving analytical success. In addition, it is important to note that, in order to avoid destructive sampling and preserve the original structure of the skull, teeth, and bones for future

morphological/anthropological studies, minimally invasive methods have been developed.

In conclusion, additional investigative tools are necessary when the forensic anthropological investigation does not have valid instruments to identify unknown remains, and from such a perspective, the support of other disciplines and novel technology is essential, particularly that of molecular anthropology focusing on the analysis of degraded DNA.

References

Baynam G, Walters M, Claes P et al (2013) The facial evolution: looking backward and moving forward. Hum Mutat 34:14–22. https://doi.org/10.1002/humu.22219

Boonyarit H, Mahasirimongkol S, Chavalvechakul N et al (2014) Development of a SNP set for human identification: a set with high powers of discrimination which yields high genetic information from naturally degraded DNA samples in the Thai population. Forensic Sci Int Genet 11:166–173. https://doi.org/10.1016/j.fsigen.2014.03.010

Bulbul O, Kidd KK (2021) Forensic ancestry inference: data requirements, analysis methods, and interpretation of results. In: Pilli E, Berti A (eds) Forensic DNA analysis technological development and innovative applications. Apple Academic Press, Florida, USA

Campos PF, Craig OE, Turner-Walker G et al (2012) DNA in ancient bone – where is it located and how should we extract it? Ann Anat 194:7–16. https://doi.org/10.1016/j.aanat.2011.07.003

Caplova Z, Obertova Z, Gibelli DM et al (2018) Personal identification of deceased persons: an overview of the current methods based on physical appearance. J Forensic Sci 63:662–671. https://doi.org/10.1111/1556-4029.13643

Cha S, Lim JE, Park AY et al (2018) Identification of five novel genetic loci related to facial morphology by genome-wide association studies. BMC Genomics 19:481. https://doi.org/10.1186/s12864-018-4865-9

Correia Dias H, Corte-Real F, Cunha E, Manco L (2020) DNA methylation age estimation from human bone and teeth. Aust J Forensic Sci. https://doi.org/10.1080/00450618.2020.1805011

Dabney J, Meyer M (2019) Extraction of highly degraded DNA from ancient bones and teeth. In: Methods in molecular biology. Humana Press, New York, NY, pp 25–29

Dabney J, Knapp M, Glocke I et al (2013) Complete mitochondrial genome sequence of a middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proc Natl Acad Sci U S A 110:15758–15763. https://doi.org/10.1073/pnas.1314445110

Dash HR, Rawat N, Das S (2020) Alternatives to amelogenin markers for sex determination in humans and their forensic relevance. Mol Biol Rep 47:2347–2360

Dayal MR, Spocter MA, Bidmos MA (2008) An assessment of sex using the skull of black South Africans by discriminant function analysis. Homo 59:209–221. https://doi.org/10.1016/j.jchb.2007.01.001

Draus-Barini J, Walsh S, Pośpiech E et al (2013) Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. Investig Genet 4. https://doi.org/10.1186/2041-2223-4-3

Emery MV, Bolhofner K, Winingear S et al (2020) Reconstructing full and partial STR profiles from severely burned human remains using comparative ancient and forensic DNA extraction techniques. Forensic Sci Int Genet 46:102272. https://doi.org/10.1016/j.fsigen.2020.102272

Fazi A, Gobeski B, Foran D (2014) Development of two highly sensitive forensic sex determination assays based on human DYZ1 and Alu repetitive DNA elements. Electrophoresis 35:3028–3035. https://doi.org/10.1002/elps.201400103

Garoufi N, Bertsatos A, Chovalopoulou ME, Villa C (2020) Forensic sex estimation using the vertebrae: an evaluation on two European populations. Int J Legal Med 134:2307–2318. https://doi.org/10.1007/s00414-020-02430-w

Gokhman D, Mishol N, de Manuel M et al (2019) Reconstructing Denisovan anatomy using DNA methylation maps. Cell 179:180–192.e10. https://doi.org/10.1016/j.cell.2019.08.035

Gonzalez A, Cannet C, Zvénigorosky V et al (2020) The petrous bone: ideal substrate in legal medicine? Forensic Sci Int Genet 47:102305. https://doi.org/10.1016/j.fsigen.2020.102305

Goodwin WH (2017) The use of forensic DNA analysis in humanitarian forensic action: the development of a set of international standards. Forensic Sci Int 278:221–227. https://doi.org/10.1016/j.forsciint.2017.07.002

Hansen HB, Damgaard PB, Margaryan A et al (2017) Comparing ancient DNA preservation in petrous bone and tooth cementum. PLoS One 12:e0170940. https://doi.org/10.1371/journal.pone.0170940

Hopman SMJ, Merks JHM, Suttie M et al (2014) Face shape differs in phylogenetically related populations. Eur J Hum Genet 22:1268–1271. https://doi.org/10.1038/ejhg.2013.289

Jing X, Sun Y, Zhao W et al (2019) Predicting adult height from DNA variants in a European-Asian admixed population. Int J Legal Med 133:1667–1679. https://doi.org/10.1007/s00414-019-02039-8

Kondo M, Aboshi H, Yoshikawa M et al (2021) A newly developed age estimation method based on CpG methylation of teeth-derived DNA using real-time methylation-specific PCR. J Oral Sci 63:54–58. https://doi.org/10.2334/josnusd.20-0138

Kukla-Bartoszek M, Szargut M, Pośpiech E et al (2020) The challenge of predicting human pigmentation traits in degraded bone samples with the MPS-based HIrisPlex-S system. Forensic Sci Int Genet 47:102301. https://doi.org/10.1016/j.fsigen.2020.102301

Kulstein G, Hadrys T, Wiegand P (2017) As solid as a rock – comparison of CE- and MPS-based analyses of the petrosal bone as a source of DNA for forensic identification of challenging cranial bones. Int J Legal Med. https://doi.org/10.1007/s00414-017-1653-z

Leclair B, Shaler R, Carmody GR et al (2007) Bioinformatics and human identification in mass fatality incidents: the World Trade Center disaster. J Forensic Sci 52(4):806–819

Lei G, Liu F, Liu P et al (2019) Worldwide tendency and focused research in forensic anthropology: a bibliometric analysis of decade (2008–2017). Legal Med 37:67–75. https://doi.org/10.1016/j.legalmed.2019.01.008

López-Lázaro S, Pérez-Fernández A, Alemán I, Viciano J (2020) Sex estimation of the humerus: a geometric morphometric analysis in an adult sample. Legal Med 47. https://doi.org/10.1016/j.legalmed.2020.101773

Madel M-B, Niederstätter H, Parson W (2016) TriXY – homogeneous genetic sexing of highly degraded forensic samples including hair shafts. Forensic Sci Int Genet 25:166–174. https://doi.org/10.1016/j.fsigen.2016.09.001

Márquez-Ruiz AB, González-Herrera L, Luna J d D, Valenzuela A (2020) DNA methylation levels and telomere length in human teeth: usefulness for age estimation. Int J Legal Med 134:451–459. https://doi.org/10.1007/s00414-019-02242-7

Mckinnon M, Higgins D (2021) Comparison of bone demineralisation procedures for DNA recovery from burned remains. Forensic Sci Int Genet 51:102448. https://doi.org/10.1016/j.fsigen.2020.102448

Mittnik A, Wang C-C, Svoboda J, Krause J (2016) A molecular approach to the sexing of the triple burial at the upper Paleolithic site of Dolní Věstonice. PLoS One 11:e0163019. https://doi.org/10.1371/journal.pone.0163019

Mushailov V, Rodriguez SA, Budimlija ZM et al (2015) Assay development and validation of an 8-SNP multiplex test to predict eye and skin coloration. J Forensic Sci 60:990–1000. https://doi.org/10.1111/1556-4029.12758

Ng MCY, Graff M, Lu Y et al (2017) Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry:

African ancestry anthropometry genetics consortium. PLoS Genet 13:e1006719. https://doi.org/10.1371/journal.pgen.1006719

Pedrosa M, Curate F, Batista de Carvalho LAE et al (2020) Beyond metrics and morphology: the potential of FTIR-ATR and chemometrics to estimate age-at-death in human bone. Int J Legal Med 134:1905–1914. https://doi.org/10.1007/s00414-020-02310-3

Pilli E, Vai S, Caruso MG et al (2018) Neither femur nor tooth: petrous bone for identifying archaeological bone samples via forensic approach. Forensic Sci Int 283. https://doi.org/10.1016/j.forsciint.2017.12.023

Pilli E, Castoldi E, Cattaneo C (2021) Forensic anthropology issues: a synergy between physical and molecular methods -the contribution of degraded DNA analysis to physical and forensic anthropology. In: Pilli E, Berti A (eds) Forensic DNA analysis technological development and innovative applications. Apple Academic Press, Florida, USA

Pluta D, Lebioda A, Jonkisz A, Dobosz T (2016) First successful DNA isolation and profiling from bone using an approach that is non-destructive toward bone surface. Arch Med Sadowej Kryminol 66:65–70. https://doi.org/10.5114/amsik.2016.62335

Podini D, Gettings KB (2021) Forensic analysis of externally visible characteristics: phenotyping. In: Pilli E, Berti A (eds) Forensic DNA analysis technological development and innovative applications. Apple Academic Press, Florida, USA

Prinz M, Carracedo A, Mayr WR et al (2007) DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). Forensic Sci Int Genet 1:3–12. https://doi.org/10.1016/j.fsigen.2006.10.003

Puerto MS, Egaña S, Doretti M, Vullo CM (2014) A multidisciplinary approach to commingled remains analysis: anthropology, genetics, and background information. In: Commingled human remains: methods in recovery, analysis, and identification. Elsevier Amsterdam, Netherlands, pp 307–335

Ralf A, van Oven M, Zhong K, Kayser M (2015) Simultaneous analysis of hundreds of Y-chromosomal SNPs for high-resolution paternal lineage classification using targeted semiconductor sequencing. Hum Mutat 36:151–159. https://doi.org/10.1002/humu.22713

Romanini C, Romero M, Salado Puerto M et al (2015) Ancestry informative markers: inference of ancestry in aged bone samples using an autosomal AIM-Indel multiplex. Forensic Sci Int Genet 16:58–63. https://doi.org/10.1016/j.fsigen.2014.11.025

Saragoni VG, Celis RC, Intriago M et al (2020) Overcoming the undetected inhibition of bone DNA extracts obtained by total demineralization. Forensic Sci Int Genet 48:102363

Sirak K, Fernandes D, Cheronet O et al (2020) Human auditory ossicles as an alternative optimal source of ancient DNA. Genome Res 30:427–436. https://doi.org/10.1101/gr.260141.119

Skoglund P, Storå J, Götherström A, Jakobsson M (2013) Accurate sex identification of ancient human remains using DNA shotgun sequencing. J Archaeol Sci 40:4477–4482. https://doi.org/10.1016/J.JAS.2013.07.004

Ubelaker DH, Shamlou A, Kunkle A (2019a) Contributions of forensic anthropology to positive scientific identification: a critical review. Forensic Sci Res 4:45–50

Ubelaker DH, Shamlou A, Kunkle AE (2019b) Forensic anthropology in the global investigation of humanitarian and human rights abuse: perspective from the published record. Sci Justice 59:203–209

Wood AR, Esko T, Yang J et al (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46:1173–1186. https://doi.org/10.1038/ng.3097

Xiong Z, Dankova G, Howe LJ et al (2019) Novel genetic loci affecting facial shape variation in humans. elife 8. https://doi.org/10.7554/eLife.49898

Zelic K, Pavlovic S, Mijucic J et al (2020) Applicability of pulp/tooth ratio method for age estimation. Forensic Sci Med Pathol 16:43–48. https://doi.org/10.1007/s12024-019-00200-8

Part II

Usefulness of Various Techniques of DNA Profiling in Solving Cases

Short Tandem Repeat Mutations in Paternity Analysis

17

Uthandaraman Mahalinga Raja, Usharani Munuswamy, Rajshree Raghunath, Thilaga Dhanapal, and Mahalakshmi Nithyanandam

Contents

U. Mahalinga Raja (✉) · U. Munuswamy · R. Raghunath · T. Dhanapal · M. Nithyanandam
Forensic Sciences Department, DNA Division, Chennai, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_15

Abstract

Polymorphic autosomal short tandem repeats (STRs) are a class of genetic markers that are routinely being used for parentage testing purposes in forensic science laboratories. They are composed of tandemly arranged repeat units (typically 2–7 bp in length) and are widespread throughout the human genome. Among the various types of repeats, tetranucleotide repeats have become the widely utilized systems due to their specific advantages. STRs have become markers of choice in Forensic DNA typing due to the ease with which they can be studied. STR DNA typing involves amplification using Polymerase Chain Reaction (PCR), capillary electrophoresis, and subsequent detection of the amplified fragments using laser-induced fluorescence of the size-separated DNA fragments. A panel of STR markers are being routinely used to solve paternity and kinship cases. Together with lineage markers such as Y-STRs and mitochondrial DNA (mtDNA), they form a formidable set of markers that can resolve disputed paternity cases with high confidence. Nevertheless, ambiguous results can be obtained in a proportion of cases due to incompatible genotypes between the disputed parent and the child, which requires careful interpretation of the DNA typing results. We finish this chapter by discussing a difficult paternity dispute case in which the concurrent paternal and maternal inconsistencies were observed in the questioned child.

Keywords

Short tandem repeats (STRs) · Mutations · Inconsistencies · Paternity testing

Introduction

The advent of DNA fingerprinting has revolutionized the field of forensic investigation. It provides a reliable approach not only in human identity establishment but also in parentage analysis. Short tandem repeat (STR) typing is the most preferred method of DNA fingerprinting that is widely applied in forensic casework, paternity disputes, immigration processes, disaster victim identification, and identification of missing persons. However, in a subset of cases, STR inconsistencies can be obtained which adds to the complexity of analysis, which needs careful interpretation of the results.

In this chapter, we provide a brief outline on the various DNA markers, with particular attention to STR markers and the statistical considerations in parentage testing. We also discuss the mutations/inconsistencies that can be observed between the parent and the offspring during routine parentage testing and their statistical considerations. We then provide a brief summary on the supplementary DNA markers and methodologies which can throw light on the molecular basis of these inconsistencies. We conclude the chapter by presenting a paternity testing case in which concurrent paternal and maternal inconsistencies were observed in the questioned child.

Types of Samples Received for Parentage Testing

A variety of samples can be received by the forensic DNA typing laboratories for parentage analysis. These samples are usually good quality reference samples, for example, venous blood collected on FTA cards, buccal swabs, or rinses. Consequently, DNA extraction is straightforward. Difficult samples can also be received at the DNA typing laboratory that include samples from medically terminated pregnancies (Bauer et al. 2002), formalin fixed tissues (Ullah et al. 2017), among others. Since formalin crosslinks DNA, it may result in DNA fragmentation and base modification; hence, suitable DNA extraction methodologies need to be selected to obtain typable quality of DNA from samples stored in formalin or formalin-fixed paraffin-embedded blocks (Hykin et al. 2015). In addition appropriate methodologies need to be followed to avoid maternal tissue contamination of the fetal tissues.

Markers Used for DNA Fingerprinting

The human nuclear DNA has 22 pairs of autosomal chromosomes and a pair of sex chromosomes (XX in Females and XY in Males). Apart from the nucleus, DNA is also present in the mitochondria. A number of molecular genetic markers have been developed for human identification and paternity testing, like Variable Number of Tandem Repeats (VNTR), Short Tandem Repeats (STRs), Single Nucleotide Polymorphisms (SNPs), and Insertion Deletion markers (InDels) (Teama 2018). Of which autosomal STRs are an important class of markers that are widely utilized in forensic studies due to their high variability, discriminatory power, and reliability (Butler 2006).

Short Tandem Repeat Markers

Short tandem repeats (STRs), also referred as microsatellites or simple sequence repeats (SSRs), are the primary genetic markers used in forensic genetics. They are made of tandemly arranged repeat units that are 2–7 bp in length (Butler 2006). STRs are present uniformly throughout the human genome including autosomal, Y, and X chromosomes. STRs are classified as di-, tri-, tetra-, penta-, hexa-, and heptanucleotide repeats on the basis of length of the core repeat unit (Fan and Chu 2007). Since tetranucleotide repeats have lower mutation rates (Chakraborty et al. 1997) and are amplified with high fidelity (Radtkey et al. 2000), they are preferred over other STRs for forensic DNA fingerprinting studies.

Autosomal and Y Chromosomal Short Tandem Repeats

Autosomal STRs

A set of 24 STRs loci are commonly used throughout the world for forensic DNA typing studies with differences based on regional legislations, like the Combined DNA Index System (CODIS) (Hares 2015) and European Standard Set (ESS)

(Goodwin and Peel 2012). STR allele nomenclature has been developed over the years by the DNA Commission of the International Society for Forensic Genetics (ISFG) (Gettings et al. 2015a). It follows a simple integer-based naming of the STR alleles representing the number of repeat motifs (e.g., 10–25 repeats). Variant alleles with additional nucleotides are represented by adding a decimal point followed by the number of nucleotides in the incomplete repeat (e.g., TH01 9.3 allele).

Y Chromosomal STRs

Y chromosome is nonrecombining and patrilineally inherited; hence, Y-Chromosomal STR analysis is used in forensic genetics and in inferring population histories. The STRs on the Y chromosome can be studied using the same methodology as autosomal STRs. It can be particularly helpful in duo paternity cases where the mother is not available for comparison. In the forensic genetics context, Y-STRs can be helpful in sexual assault cases to specifically amplify the male component from the male/female DNA mixtures.

Y-STRs mutation rates range from 3.78×10^{-4} to 7.44×10^{-2} (Gusmão et al. 2005; Ballantyne et al. 2010). Single Y-STR repeat mutations are strongly favored over multiple repeat mutations. Factors that influence Y-STR mutations are total repeat number, repeat complexity, length of the repeat motif, and father's age (Gusmão et al. 2005).

Single Nucleotide Polymorphism (SNP) Markers

Single nucleotide polymorphisms (SNPs) are the most abundant class of human polymorphisms in the human genome that can be valuable supplementary markers for forensic DNA studies. The advantages of SNPs over STRs are their lower mutation rates, ability to type severely degraded samples, and the availability of high throughput methodologies to study SNPs (Sobrino and Carracedo 2004).

Autosomal SNPs

Autosomal SNPs are diallelic markers and may require more number of loci, than STRs, to be profiled. When sufficient number of SNPs are profiled, they can provide match probabilities comparable to that of STRs (Seo et al. 2013). Since SNP analysis can be used to successfully analyze even highly degraded samples, several SNP panels have been evaluated for use with forensic genetics, including the widely studied SNP for ID panel with 52 SNPs (Musgrave-Brown et al. 2007). A specific advantage of SNP markers over STRs is that they display a much less overall mutation rate compared to autosomal STRs (Phillips et al. 2008).

Mitochondrial DNA Markers

Apart from nucleus, mitochondria is the only other cellular organelle that has a genome on its own. Mitochondrial DNA (mtDNA) analysis has several advantages when it comes to forensic DNA testing like lack of recombination, matrilineal

inheritance, resistance to degradation, and high copy number compared to nuclear DNA (Amorim et al. 2019). The mitochondrial control regions and the HV1 and HV2 regions are targeted in human identification and genomic diversity studies (Butler and Levin 1998). This increased copy number of mtDNA compared to the nuclear DNA is advantageous for DNA analysis in cases with low/degraded DNA content. Hence, it is easier to obtain a typable mtDNA profile from biological samples which are highly degraded like those obtained from mutilated remains, mass disasters, and other difficult forensic samples.

The lack of repair mechanisms and the low fidelity of the mitochondrial DNA polymerases make mtDNA prone to germline and somatic mutations (Amorim et al. 2019). Thus, the average mutation rates of the mitochondrial genome are significantly higher than the nuclear genome. Studies have also shown that natural radioactivity can increase the incidence of germline mtDNA mutations in the offsprings (Forster et al. 2002). Nevertheless, mtDNA typing can be especially informative when it comes to studies on forensic genetics or population histories.

Autosomal STRs in Human Identification and Parentage Testing

Autosomal STRs follow a Mendelian pattern of inheritance; thus, in an individual, under each of the STR locus, one allele is contributed by the biological father and the other allele is contributed by the biological mother. When a sufficient number of STRs are profiled, they can be used to discriminate unrelated individuals and to infer biological relationships (Butler 2006).

In forensic practice, two types of human identification challenges exist. First, direct comparison, is to compare the STR-based DNA fingerprinting profile of the reference or suspect with that of the crime scene evidence, where both alleles at all tested loci need to match. Second, indirect or familial comparisons in parentage testing cases require one allele to match between the parent and the offspring at all the tested loci. The former is used to link or exonerate a suspect to/from the crime (using Random Match probability (Coble and Bright 2019)), while the latter is used to resolve paternity disputes, identify missing person, disaster victim identification, or in the context of immigration (using Probability of Paternity – W (Gjertson and Morris 1995)).

Utility of Other Autosomal and Lineage Markers in Parentage Testing

Y-Chromosome and mitochondrial DNA constitute uniparentally transmitted genetic information. They are transmitted to offsprings without recombination. Y-Chromosome is patrilineally inherited; hence, Y-STRs are used to study the paternal lineage. Mitochondrial DNA is matrilineally inherited and is used to study the maternal lineage. All the members of the lineage will have identical profiles across generations for these markers, barring the differences generated by mutations; hence, lineage markers such as Y-STRs and mtDNA markers have a lower discriminative

power compared to the autosomal STRs. However, they can provide supplementary information required for reporting cases with ambiguous autosomal STR typing results (Lindner et al. 2014).

STR Typing Methodology

STR typing is fragment analysis method which does not account for differences at the sequence level for the alleles. The most common approach to identify STR alleles is by amplification of the STRs using multiplexed polymerase chain reaction (PCR) and electrophoresis. STR typing involves the measurement of the amplicon size to determine the number of repeats present in each of the alleles of the DNA profile. Early methodologies for STR typing used silver-stained polyacrylamide gels. Considering the potential of the STRs in human identification, STR typing methodologies underwent rapid strides during the 1990s in tandem with the Human Genome Project. Currently STR typing uses fluorescence detection of size-separated amplicons using Capillary Electrophoresis on Genetic analyzers. Further data analysis is carried out on commercially available software which provides precise allele designations. The current STR typing methodologies have several advantages such as amenability to automation, requirement of low initial quantity of DNA (~1 ng), less labor intensiveness, and faster results.

Statistical Analysis of the Paternity Test Results

DNA profiles of the Alleged Father (AF), Questioned Child (QC), and the Biological Mother (BM) can be compared to deduce the biological relationship among the trio. In trio tests, offsprings inherit half of all the autosomal alleles from each parent; hence, after accounting for the contribution of alleles from the BM, obligate paternal alleles can be determined. The presence of obligate paternal alleles at all tested loci in the AF shows that he cannot be excluded as the Biological Father of the QC. The obtained results are then subjected to statistical tests to ascertain their validity.

The statistical strength of the genetic evidence can be assessed by calculating Combined Paternity Index (CPI) (Kaiser et al. 1983), Probability of Paternity (W) (Gjertson and Morris 1995), and the Probability of Exclusion (PE) (Chakraborty et al. 1974). For each tested locus, Paternity Index (PI) is calculated as a likelihood ratio of the two competing hypothesis that the AF is the true biological father or an unrelated random man is the biological father. Combined Paternity Index (CPI) can then be calculated by multiplying the individual PI values for each tested locus. A high CPI value can support the hypothesis that the AF is the biological father and a CPI cutoff value (CPI $\geq$ 10,000) is generally considered to be satisfactory to draw conclusions on a true paternity (Gao et al. 2019).

The Probability of Paternity (W) can be calculated using Bayes' theorem, based on a defined prior probability and the results of DNA typing. The probability of paternity (W) enables the DNA typing results to be expressed numerically as a

percentage, thus providing a quantitative estimate for the obtained results (Henke et al. 1999). Assuming a prior probability of 0.5, W can be calculated using the formula:

$$W = CPI/CPI + 1$$

STR Inconsistencies/Mismatches in Parentage Testing

STR typing is a powerful tool that is usually adequate for resolving paternity disputes. But in a proportion of cases, the analyst may observe inconsistencies/mismatches in a limited number of loci that increases the complexity of the case. It will be incorrect to exclude the AF from the paternity of the QC based on isolated inconsistency/inconsistencies during paternity testing and the possibility of mutations should to be considered. STR inconsistencies/mismatches can occur due to germline mutations at STR loci during meiosis (Dauber et al. 2012) or due to changes at whole chromosomal, segmental, and smaller genomic regions.

The mutation rates of commonly used tetranucleotide STR markers in the forensic setting are relatively low compared to the di-nucleotide repeats (Chakraborty et al. 1997; Fan and Chu 2007). However, the possibility of mutations cannot be ruled out as mutations play an indispensable role in variation, diversity, and evolution of STRs (Willems et al. 2014). STR mutations are usually integer mutations (Slooten and Ricciardi 2013) which are caused by the insertion or deletion of entire repeat unit/s. Determination of STR mutations and mutation rates is based on a set of general assumptions: The shortest mutational step is assumed to be the actual one, if the origin of mutation cannot be deduced as paternal or maternal then they are considered as inconclusive, and if incongruent homozygous alleles are present in the homozygous parent and the homozygous child, null alleles are assumed (Veremeichyk et al. 2015).

Polymerase strand slippage during DNA replication is the widely accepted mechanism and Stepwise Mutation Model (SMM) is the widely accepted model that explains the STR mutation events (Fan and Chu 2007). Forensic literature on mutations in paternity testing has shown that single-step mutation (expansion or contraction of the core repeat by one unit) is the most common type of inconsistency in STR testing. Multistep mutation events (up to four steps), though rare, have also been reported (González-Herrera et al. 2020). At least three genetic mismatches are needed to exclude an AF from the paternity of the QC (Dauber et al. 2012). Table 1 provides data on population-specific STR mutation rates from different populations (Zhao et al. 2015; Hamester et al. 2019).

Three hypotheses need to be considered when inconsistencies are observed between the tested parent and the child, which includes existence of meiotic mutations, or that a first degree relative of the AF being the biological father, or that an unrelated random man is the biological father. Appropriate likelihood ratios can be calculated for providing an opinion on paternity. Though residual ("mock") W value, calculated by omitting the inconsistent loci from calculations, can

provide a rough estimate of strength of the results (Gjertson and Morris 1995), it is necessary to include the mutation rates of the inconsistent locus for obtaining reliable W values.

In the following sections, we discuss the statistical considerations of STR mutations, factors influencing STR mutations, the molecular basis behind the observation of these inconsistencies, and the limitations of STR analysis in such cases. We further provide information of the alternative and complementary marker resources and techniques that can help in such complex cases.

Calculation of Mutation Rate

The Paternity Testing Commission of the International Society of Forensic Genetics has laid down guidelines for the Paternity Testing Laboratories the ways to deal with inconsistencies, among others. Under Section R1.3.1, the commission recommends to modify (using the population specific mutation rates for the inconsistent STR loci) the PI for possible mutation patterns when isolated mismatches are observed to obtain the likelihood ratio for the hypothesis being tested (Gjertson et al. 2007). Thus, in cases where genetic inconsistencies between the AF and the QC are observed, it cannot be simply ignored, but the CPI needs to be corrected by using the locus-specific mutation rate for the inconsistent system (Ochiai et al. 2016). The formula for calculating the locus-specific PI in case of an inconsistency is

$$\mathrm{PI} = \mu/\bar{\mathrm{A}}$$

where μ is the locus specific mutation rate and $\bar{\mathrm{A}}$ is the average Probability of Exclusion (PE).

Avuncular Index

"Avuncular Index (AI) is defined as the likelihood ratio that tests the hypothesis that the tested man is the paternal uncle of the child versus the hypothesis that the tested man is unrelated to the child" (Morris et al. 1988). Thus, AI provides the theoretical framework to evaluate the probability of the AF for being the brother of the biological father. AI can be useful in cases where inconsistencies are observed in one or two loci between the AF and QC, while all other loci strongly suggest paternity. In immigration-related paternity analysis, Avuncular Index (AI) is usually calculated, in addition to Paternity/Maternity Index (Wenk et al. 2005), to rule out the possibility of a blood relative impersonating a parent or a child, for example, to rule out brother of the tested man for being the actual father of the questioned child.

AI is calculated using the formula

$$\mathrm{AI} = (\mathrm{PI} + 1)/2$$

The likelihood ratio is calculated using Combined Paternity Index (CPI) and the Combined Avuncular Index (CAI) to test the hypothesis that the AF is the biological father versus the hypothesis that he is the first degree relative of the AF: CPI/CAI.

Factors Influencing Short Tandem Repeat Mutations

Published studies have shown that several factors influence STR mutations. Among the factors that influence STR mutation are the allele size, locus polymorphism, parental origin, and age. Studies have shown that shorter alleles tend to undergo repeat expansion mutations, while longer alleles tend to undergo repeat contraction mutations (Aşicioglu et al. 2004). Paternal origin mutation rates are generally higher (Table 1) compared to that of Maternal origin mutations. This has been explained by the fact that spermatogenesis occurs at a higher rate compared to oogenesis, which increases the probability of mutations during the meiotic process. A positive correlation between mutations and parental age is also observed (Hohoff et al. 2006; Zhao et al. 2015). The base composition of the STR and the flanking regions is also shown to have an effect on the rate of mutations, for example, the mutation rate is higher in STRs with high AT content (Fan and Chu 2007). Finally, the STR mutation rate is found to be higher in the coastal areas compared to the inland areas (Zhao et al. 2015). Understanding mutational events and incorporating population-specific mutation rates in the calculation of CPI can help in the appropriate interpretation of the DNA typing results.

Molecular Basis of STR Inconsistencies

STR inconsistency between the tested parent and the questioned child can occur due to a variety of molecular reasons like primer binding site mutation, repeat expansion or contraction (due polymerase slippage), and whole or segmental chromosomal aberrations (Aşicioglu et al. 2004; Slooten and Ricciardi 2013; Hamester et al. 2019). Sequence variations like insertions, deletions in the flanking regions can also contribute to the inconsistencies. However, nucleotide substitutions within the amplicons and between the primer regions generally go unnoticed as STR typing is a fragment analysis technique that relies on the detection of the mass of amplicon rather than its sequence.

Primer Binding Site Mutations

Primers for amplification of STRs are designed based on several considerations and target highly conserved sequences that flank the actual STRs. Nevertheless, sequence variations in primer binding sites, especially those at the 5′ end, can destabilize the annealing to the target site or can either block or reduce the efficiency of amplification (Clayton et al. 2004b). These point mutations are generated by transitions and transversions during DNA replication. Such mutations can result in allele dropout and are termed as null or silent allele. Null alleles

Table 1 Summary of short tandem repeat mutation literature from various populations

Population	Number of cases	Mutations observed	Overall mutation rate	Repeat deletion versus repeat expansion	Paternal mutations versus maternal mutations	Cases with one-step mutation	Cases with two-step mutation	Cases with three-step mutation	Cases with four-step mutation
Brazil (Hamester et al. 2019)	5171	193	1.3×10^{-3}	77:77	6:1	189 (97.9%)	4	–	–
China (Zhao et al. 2015)	19,037	678	1.2×10^{-3}	Data not provided	1.43:1	487 (69.8%)	135	58	18
Poland (Droździok et al. 2018)	953	38	1.2×10^{-3}	21:17	4.6:1	38 (100%)	–	–	–
Nigeria (Hohoff et al. 2009)	337	10	1.6×10^{-3}	6:4	8:1	10 (100%)	–	–	–
Afghanistan (Hohoff et al. 2006)	333	19	1.6×10^{-3}	9:7	6:1	19 (100%)	–	–	–
Japan (Ochiai et al. 2016)	221	7	8.68 X 10^{-4}	4:1	6:1	5 (100%)[a]	–	–	–
Caucasoid (Dauber et al. 2012)	Not defined	65	1.3×10^{-3}	31:22	4.2:1	62 (95.4%)	2	–	1
Turkey (Aşicioglu et al. 2004)	59	12	1.8×10^{-3}	7:4[b]	11:1	12 (100%)	–	–	–

Belarusian (Veremeichyk et al. 2015)	11,744	283	9.2×10^{-4}	117:127	5.26:1	280 (98.23%)	1	1	1
Ecuadorian (Gaviria et al. 2017)	16,310	523	3.8×10^{-4}	Data not provided	4.97:1	Data not provided	Data not provided	Data not provided	Data not provided
Colombian (Burgos et al. 2019)	11,077	251	1.1×10^{-2}	Data not provided	Data not provided	247 (98.4%)	3	1	–

[a]Two inconclusive cases
[b]One inclusive case

can be identified as a discrepancy during concordance studies between different commercially available STR typing kits and as STR inconsistencies during parentage testing (Yao et al. 2018).

The tested loci show a homozygous allelic pattern in the parent and the child in case of silent alleles. Though the occurrence of such inconsistencies is relatively rare due to the comprehensive strategies behind selection of primer sets, it can still occur in a subset of cases. These mismatches between the parent and the offspring might result in inaccurate statistical interpretations (Yao et al. 2018). Primer binding site mutations can be confirmed by reanalysis with a different STR typing kit (from a different manufacturer) which utilizes primers that target a different flanking region for the STR locus. In addition, Massively Parallel Sequencing strategies can also provide information regarding these variations (Yang et al. 2014).

Strand Slippage Mutation

Strand slippage mutations also known as DNA slippage, polymerase slippage, and slipped strand mispairing is the widely accepted mechanism of STR mutations (Aşicioglu et al. 2004). Stepwise mutation model is the most accepted model for the STR mutation (Payseur and Cutter 2006). Due to their repetitive nature, replication of STR may sometimes result in mispairing between the template strand and the newly synthesized strand. This mispairing may result in looping out of one or more of the STR repeat units (Levinson and Gutman 1987). These changes are usually identified and eliminated by cellular mismatch repair mechanisms. However, if the DNA synthesis continues, this may result in alterations in the repeat numbers in the newly synthesized DNA molecule. If the strand slippage occurs during meiosis, it can result in the STR mutation in the offspring (Aşicioglu et al. 2004). Factors that affect the strand slippage mutation rates include length, location, and sequence of the repeat unit, in addition to sex and age (Fan and Chu 2007).

Uniparental Disomy

DNA is inherited from both the parents; however, under rare conditions, a child can inherit both the copies of a particular chromosome from a single parent. This molecular anomaly is known as Uniparental Disomy (UPD) (Guzmán-Alberto et al. 2019). UPD does not involve aneuploidy (a change in the number of chromosomes), and during parentage testing, UPD can result in isolated inconsistencies which will be characterized by the presence of both the alleles at the STR loci from the one parent and the absence of the STR allele from the other. This will be true for all the STRs that are present in the chromosome that is inherited based on UPD (Guzmán-Alberto et al. 2019).

UPD is further classified into two types based on whether two identical copies are inherited [uniparental isodisomy (iUPD)] or different homologs are inherited [uniparental heterodisomy (hUPD)] (Niida et al. 2018). iUPD may occur due to the nonseparation of sister chromatids during second meiotic division. hUPD may occur due to the nonseparation of bivalent chromatids during meiosis I (Ting et al. 2007).

Anomalous Band Patterns

Normally, in an individual, each STR locus being tested yields one (homozygous) or two (heterozygous) peaks. Anomalous band patterns such as triallelic patterns (and aberrant diallelic patterns) can be observed, albeit rarely, during routine parentage testing analysis (Clayton et al. 2004a; Raja et al. 2020). The triallelic band pattern is classified as type I or type II based on the peak heights (relative intensity). In type I, the three peaks are of uneven intensities and can result from a somatic mutation (primer binding site or repeat number mutations) at the heterozygous locus early during embryonic development or due to chimerism. While in type II, the three peaks are of even intensities and can result from intrachromosomal segmental duplications (Lukka et al. 2006; Raja et al. 2020) or aneuploidy (Huel et al. 2007). For an extensively reviewed of anomalous band patterns and their classification, the reader is referred to (Clayton et al. 2004a).

Complementary Markers to Study Inconsistencies in Parentage Testing

Though the regular panels of autosomal STR markers are sufficient for parentage testing, in some cases isolated inconsistencies can be encountered, which adds complexity to the interpretation of the DNA typing results. A combination of a small number of inconsistencies along with a high Likelihood Ratios (LR) may require a careful assessment of the biological relationships. In such situations, an expanded set of STRs are usually typed to provide a clear unambiguous exclusion of the tested man or to increase the LR.

When the autosomal STR markers are insufficient in providing an unambiguous result, it may be necessary to utilize supplementary genetic markers to provide an opinion on paternity. For example, alternative batteries of genetic markers like autosomal Insertion-Deletion (InDel) markers (González-Herrera et al. 2020), Single Nucleotide Polymorphism (SNP) markers (Børsting and Morling 2011), and Retrotransposon Insertion Polymorphisms (RIPs) markers (González-Herrera et al. 2020) have proven as useful supplementary markers in paternity cases with ambiguous STR typing results.

SNPs are diallelic markers that can perform well when STR markers are unable to resolve parentage testing cases due to mutations. A major advantage of SNP markers for parentage testing is their low mutation rates (10^{-8}) compared to STR markers (10^{-3} to 10^{-4}) (Lindner et al. 2014). In addition SNPs do not produce stutter which can be of help in DNA mixture analysis (LaRue et al. 2012). In the medical setting, SNP analysis can be used to systematically study chromosomal alterations like mutations, deletions, and duplications. Similarly comparison of the SNP data from trios (father, child, mother) can provide an additional layer of information that can help in determining the genetic basis of the inconsistency observed during parentage testing. For example, SNP data from trios can provide information on whether the

STR inconsistency observed during parentage testing is due to chromosomal alterations or due to a nonparental relationship (Ting et al. 2007).

InDels are a major class of polymorphisms in the human genome that have considerable potential in human identification and kinship testing (Pereira et al. 2009). In recent years, the utility of InDels as supplementary markers for human identification is being increasingly recognized due to their lower mutation rate and smaller amplicon size (LaRue et al. 2012). Similar to SNPs, InDels are diallelic markers. Thus, more number of loci needs to be profiled to match the efficiency of STR markers. However, InDel profiling can provide information needed to conclude the paternity testing when mismatches are presented.

Next Generation Sequencing

It is important to characterize the mutational event to resolve complex relationship cases. Since traditional STR typing by Capillary Electrophoresis (CE) is based on the fragment size, it is difficult to identify the parental source of mutations. As CE cannot sufficiently resolve complex paternity cases in which mutations are observed, it can result in false exclusions. Sequencing of the ambiguous STR loci can provide a precise description of the structure of the STR allele and the flanking regions, thereby helping in the identification of the source of mutations (Narkuti et al. 2007). Though Sanger sequencing can be used to further characterize the STR loci, the laborious nature of the procedure and additional steps required to discern heterozygous alleles has made forensic geneticists to look for other avenues.

NGS has been especially useful in these cases, as it can distinguish mutational events from true exclusions. Massively Parallel Sequencing (MPS) or Next Generation Sequencing (NGS) is used in forensic DNA typing (Yang et al. 2014) due to its multiple advantages like high throughput, higher discrimination power, decreased cost, and the continued improvement. The use of NGS technology provides distinct advantages over CE, as it helps in the identification of newer alleles with sequence variations and can also help in the identification of the source of STR mutations (Gettings et al. 2015a).

Several studies have demonstrated the usefulness of NGS in forensic DNA typing studies. For example, using NGS has led to the unequivocal identification of the parental source of D21S11 mutations in a majority of the cases, besides the identification of newer D21S11 alleles with sequence variations (Rockenbauer et al. 2014). The identification of newer alleles can directly translate into higher statistical power for STR-based analysis of parentage testing. The use of NGS in forensic DNA typing is an important research area that is undergoing rapid developments in the recent times. Since NGS can also provide information on the flanking region variations, SNPs and InDels markers, NGS offers a possibility of combining relevant autosomal STR-SNP-InDel markers (Gettings et al. 2015b).

Mutations in simple STR may still be difficult to ascertain because of their lower sequence-based information content. The addition of more compound/complex STR

repeats may increase the number of alleles and may provide a higher discrimination power. This also offers the possibility of accurately identifying the parental source of mutations. But NGS methodology may require developmental validations, quality assurance measures, and internal laboratory validations before it can be implemented in routine forensic casework (Gettings et al. 2015a, 2016).

Case Study

Concurrent Paternal and Maternal Inconsistencies at Two Autosomal STR Loci In A Case of Paternity Testing.

Background

A difficult paternity is reported here, in which the child acquired concurrent paternal and maternal mutations at two STR loci (D3S1358 and D13S317, respectively). Results obtained with *AmpFlSTR® Identifiler® Plus* were reconfirmed with *Promega PowerPlex® Fusion System* and *GlobalFiler™* kit. The probability of paternity and maternity (W) were 99.9999999999% and 99.99999999%, respectively (for the 15 STR markers included in the *Identifiler® Plus kit,* after incorporating the mutation rates for D3S1358 and D13S317). In addition, matching 17 Y-STR and 12 X-STR loci clearly established the parentage in this complex paternity case.

Methods

DNA was eluted from blood samples collected on FTA microelute cards from the Alleged Father, Questioned Child (male), and the Biological Mother. The eluted DNA was profiled for 15 autosomal STR loci included in the *AmpFlSTR® Identifiler® Plus* kit (CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX, vWA). An inconsistency was observed at D3S1358 which was attributed to the paternal side, and another inconsistency was observed at D13S317 which was attributed to the maternal side. Since two mismatches (one from each parent) were observed in the child, the results were then reconfirmed using *Promega PowerPlex® Fusion System* and the *GlobalFiler™* kit. In addition, the sex chromosomal STRs were profiled using *AmpFLSTR® Yfiler™ kit* and *Investigator® Argus X-12 kit*. All amplification reactions were performed as per manufacturer's instructions and appropriate positive and negative controls were included. The amplified PCR products were then subjected to capillary electrophoresis on Genetic Analyzer 3130XL/3500XL and genotyped using Genemapper ID-X v 1.5 and Genemapper ID-X v 1.6.

Results

The genotype of the alleged father, questioned child, and biological mother for D3S1358 locus were 17, 15/18, 15 and for D13S317 were 11/14, 14, 12, respectively (Fig. 1a, b). Maternity is assumed; however, probability of maternity was also calculated based on the obtained results. To confirm the genetic relationship among the trio, a battery of tests were employed to check for consistency of the results and to extract information on additional autosomal and sex chromosomal STR loci (Table 2). The results were consistent with that of AmpFℓSTR® Identifiler plus kit and no further inconsistencies were observed in the additional markers tested. In all, 23 autosomal STRs were genotyped using the three autosomal STR kits and further statistical analysis confirmed the genetic relationship among the trio.

Furthermore, AmpFℓSTR® Yfiler™ and Investigator® Argus X-12 kit were employed to study the Y and X chromosomal STR loci. A complete match was obtained for the alleged father and the questioned male child with 17 Y-STR loci (AmpFℓSTR® Yfiler™ – Applied Biosystems) (Table 3). Match was also obtained for the biological mother and the questioned male child with 12 X-STR loci (Qiagen Investigator® Argus X-12 kit) (Table 4).

Statistical Analysis of the DNA Typing Results

The probability of paternity and probability of maternity (W) were 99.9999999999% and 99.99999999%, respectively (after incorporating the mutation rates for the inconsistent loci) which further supported the genetic relationships among the trio (Table 5). To our knowledge, this is the first case, in which concurrent maternal and paternal inconsistencies were observed in the child.

Discussion

The D3S1358 mutation was of paternal origin and a gain of a single repeat unit was evident (17 to 18), while D13S317 mutation was of maternal origin and a gain of two repeat units was evident (12 to 14). Since it is a case of disputed paternity, the two-step mutation in maternity was assumed. Nevertheless, the maternity index and probability of maternity was also calculated. These mutations have occurred independently of each other (during paternal and maternal meioses). Data from the American Association of Blood Banks (AABB) (2003) (accessed from STRbase – 18th May, 2020) show that these loci display moderate mutation rates – D3S1358 (maternal – 0.015%, paternal – 0.13%, and combined – 0.12%) and D13S317 (maternal – 0.04%, paternal – 0.14% and combined – 0.14%). Though germline mutations in STR loci are frequent in older parents, in this case, the age of the biological mother and the alleged father were 19 and 21, respectively.

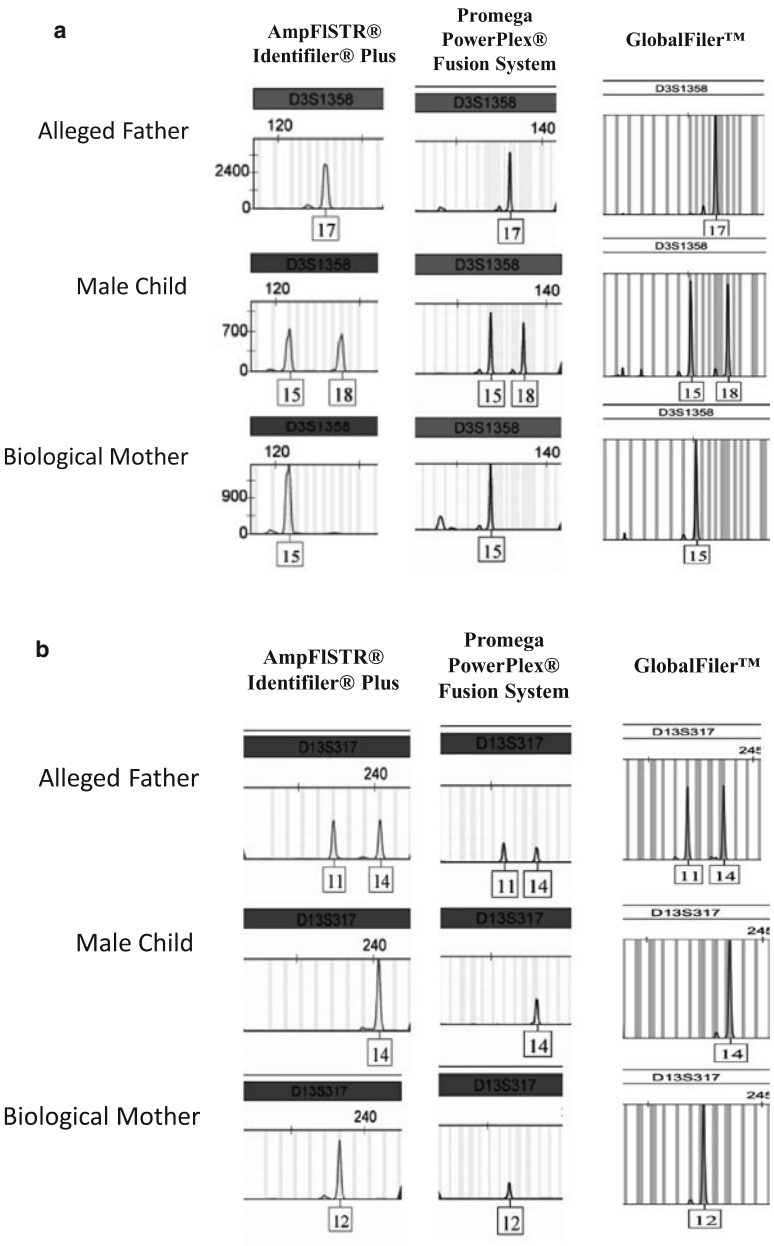


Fig. 1 (**a**) Electropherogram showing paternal allele mismatch at D3S1358 Locus. Analysis performed using AmpFlSTR® Identifiler® Plus, Promega PowerPlex® Fusion System, and GlobalFiler™ kit. (**b**) Electropherogram showing maternal allele mismatch at D13S317 Locus. Analysis performed using AmpFlSTR® Identifiler® Plus, Promega PowerPlex® Fusion System, and GlobalFiler™ kit

Table 2 Autosomal STR typing results for the Trio using AmpFLSTR® Identifiler®, PowerPlex® 21 System, and GlobalFiler™ kits. Inconsistencies are marked in italics and underlined (# Y chromosome STR and Y Indel hence did not amplify in Biological mother)

		AMPFlSTR® Identifiler® Plus Kit			Promega Powerplex Fusion Kit			GlobalFiler™ Kit		
S No	Locus	Alleged father	Male child	Biological mother	Alleged father	Male child	Biological mother	Alleged father	Male child	Biological mother
1	D8S1179	14,14	14,14	14,14	14,14	14,14	14,14	14,14	14,14	14,14
2	D21S11	30.2,32.2	30,30.2	30,32.2	30.2,32.2	30,30.2	30,32.2	30.2,32.2	30,30.2	30,32.2
3	D7S820	7,11	11,12	8,12	7,11	11,12	8,12	7,11	11,12	8,12
4	CSF1PO	10,12	10,12	12,12	10,12	10,12	12,12	10,12	10,12	12,12
5	D3S1358	*17,17*	*15,18*	15,15	*17,17*	*15,18*	15,15	*17,17*	*15,18*	15,15
6	TH01	9,9	9,9.3	8,9.3	9,9	9,9.3	8,9.3	9,9	9,9.3	8,9.3
7	D13S317	11,14	*14,14*	*12,12*	11,14	*14,14*	*12,12*	11,14	*14,14*	*12,12*
8	D16S539	10,10	10,13	13,14	10,10	10,13	13,14	10,10	10,13	13,14
9	D2S1338	18,22	18,23	23,25	18,22	18,23	23,25	18,22	18,23	23,25
10	D19S433	13,14.2	14.2,16	14,16	13,14.2	14.2,16	14,16	13,14.2	14.2,16	14,16
11	vWA	17,18	17,17	16,17	17,18	17,17	16,17	17,18	17,17	16,17
12	TPOX	10,10	8,10	8,12	10,10	8,10	8,12	10,10	8,10	8,12
13	D18S51	14,14	14,15	13,15	14,14	14,15	13,15	14,14	14,15	13,15
14	D5S818	11,11	11,12	11,12	11,11	11,12	11,12	11,11	11,12	11,12
15	FGA	23,24	22.2,24	22.2,23	23,24	22.2,24	22.2,23	23,24	22.2,24	22.2,23

16	D1S1656	–	–	–	14,14	12,14	12,17.3	14,14	12,14	12,17.3
17	D2S441	–	–	–	11,11	10,11	10,11	11,11	10,11	10,11
18	D10S1248	–	–	–	14,16	16,17	16,17	14,16	16,17	16,17
19	Penta E	–	–	–	14,16	5,16	5,17	–	–	–
20	Penta D	–	–	–	6,12	9,12	9,11	–	–	–
21	D12S391	–	–	–	18,18	18,22	22,22	18,18	18,22	22,22
22	D22S1045	–	–	–	11,15	11,14	11,14	11,15	11,14	11,14
23	SE33	–	–	–	–	–	–	28.2,30.2	16,28.2	16,19
24	AMELOGENIN	X,Y	X,Y	X,X	X,Y	X,Y	X,X	X,Y	X,Y	X,X
25	DYS391	–	–	–	11	11	#	11	11	#
26	Yindel	–	–	–	–	–	–	2	2	#

Table 3 Y-chromosome STR profiles from the alleged father and the male child using AmpFLSTR® Yfiler™ kit (Applied Biosystems)

Locus	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a	DYS385b	DYS393	DYS391	DYS439	DYS635	DYS392	Y GATA H4	DYS437	DYS438	DYS448
Alleged father	16	13	25	31	16	16	11	14	13	11	10	24	11	12	14	11	20
Male child	16	13	25	31	16	16	11	14	13	11	10	24	11	12	14	11	20

Table 4 X-chromosome STR profiles from the biological mother and the male child using Investigator® Argus X-12 kit (Qiagen)

Locus	Amelogenin	DXS7132	DXS7423	DXS8378	DXS10074	DXS10079	DXS10101	DXS10103	DXS10134	DXS10135	DXS10146	DXS10148	HPRTB
Biological mother	X,X	16,19	11,11	12,12	37,41	16,16	28,33.2	17,28	14,15	28,31	18,19	13,15	24.1,25.1
Male child	X,Y	19	11	12	41	16	33.2	28	14	28	18	13	25.1

Table 5 Summary of paternity statistics

Paternity statistics	
Residual (Mock) analysis for 14 STR without D3S1358[a]	
Combined Paternity Index (CPI)	6,613,731,100,000
Probability of Paternity (W)	99.9999999999
Combined Avuncular Index (CAI)	99.9999999
Single Inconsistency at D3S1358[a]	
Combined Paternity Index (CPI)	1439809258898.52
Probability of Paternity (W)	99.9999999999
Combined Avuncular Index (CAI)	3168823691.49
Likelihood Ratio CPI/CAI	454.36

[a]Assuming true maternity, contribution, and homozygosity at D13S317 locus

Conclusions

STR inconsistencies during DNA typing results are to be expected in a proportion of cases during routine paternity analysis. Including supplementary genetic markers and additional statistical considerations can provide information needed to resolve such difficult cases.

References

Amorim A, Fernandes T, Taveira N (2019) Mitochondrial DNA in human identification: a review. PeerJ 7:e7314. https://doi.org/10.7717/peerj.7314

Aşicioglu F, Oguz-Savran F, Ozbek U (2004) Mutation rate at commonly used forensic STR loci: paternity testing experience. Dis Markers 20:313–315. https://doi.org/10.1155/2004/643086

Ballantyne KN, Goedbloed M, Fang R et al (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 87:341–353. https://doi.org/10.1016/j.ajhg.2010.08.006

Bauer M, Thalheimer A, Patzelt D (2002) Paternity testing after pregnancy termination using laser microdissection of chorionic villi. Int J Legal Med 116:39–42. https://doi.org/10.1007/s004140000197

Børsting C, Morling N (2011) Mutations and/or close relatives? Six case work examples where 49 autosomal SNPs were used as supplementary markers. Forensic Sci Int Genet 5:236–241. https://doi.org/10.1016/j.fsigen.2010.02.007

Burgos G, Posada Y, Florez-Misas A et al (2019) An update of STR mutation rates from paternity tests analyzed in a 14 year period (2005–2018) at IdentiGEN lab, Universidad de Antioquia, Colombia. Forensic Sci Int Genet Suppl Series 7:530–531. https://doi.org/10.1016/j.fsigss.2019.10.078

Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. J Forensic Sci 51:253–265. https://doi.org/10.1111/j.1556-4029.2006.00046.x

Butler JM, Levin BC (1998) Forensic applications of mitochondrial DNA. Trends Biotechnol 16:158–162. https://doi.org/10.1016/S0167-7799(98)01173-1

Chakraborty R, Kimmel M, Stivers DN et al (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci 94:1041–1046. https://doi.org/10.1073/pnas.94.3.1041

Chakraborty R, Shaw M, Schull WJ (1974) Exclusion of paternity: the current state of the art. Am J Hum Genet 26:477–488

Clayton TM, Guest JL, Urquhart AJ, Gill PD (2004a) A genetic basis for anomalous band patterns encountered during DNA STR profiling. J Forensic Sci 49:1207–1214

Clayton TM, Hill SM, Denton LA et al (2004b) Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR® SGM Plus™ kit. Forensic Sci Int 139:255–259. https://doi.org/10.1016/j.forsciint.2003.10.004

Coble MD, Bright J-A (2019) Probabilistic genotyping software: an overview. Forensic Sci Int Genet 38:219–224. https://doi.org/10.1016/j.fsigen.2018.11.009

Dauber E-M, Kratzer A, Neuhuber F et al (2012) Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions. Forensic Sci Int Genet 6: 381–386. https://doi.org/10.1016/j.fsigen.2011.07.015

Droździok K, Kabiesz J, Tomsia M et al (2018) Mutation analysis of short tandem repeats in a population sample from Upper Silesia (southern Poland). Leg Med (Tokyo) 33:1–4. https://doi.org/10.1016/j.legalmed.2018.03.010

Fan H, Chu J-Y (2007) A brief review of short tandem repeat mutation. Genomics Proteomics Bioinformatics 5:7–14. https://doi.org/10.1016/S1672-0229(07)60009-6

Forster L, Forster P, Lutz-Bonengel S et al (2002) Natural radioactivity and human mitochondrial DNA mutations. Proc Natl Acad Sci U S A 99:13950–13954. https://doi.org/10.1073/pnas.202400499

Gao H, Wang C, Zhang R et al (2019) Application of CPI cutoff value based on parentage testing of duos and trios typed by four autosomal kits. PLoS One 14:e0225174. https://doi.org/10.1371/journal.pone.0225174

Gaviria A, Vela M, Fiallos G et al (2017) Mutation rates for 29 short tandem repeat loci from the Ecuadorian population. Forensic Sci Int Genet Suppl Series 6:e229–e230. https://doi.org/10.1016/j.fsigss.2017.09.103

Gettings KB, Aponte RA, Vallone PM, Butler JM (2015a) STR allele sequence variation: current knowledge and future issues. Forensic Sci Int Genet 18:118–130. https://doi.org/10.1016/j.fsigen.2015.06.005

Gettings KB, Kiesler KM, Faith SA et al (2016) Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Sci Int Genet 21:15–21. https://doi.org/10.1016/j.fsigen.2015.11.005

Gettings KB, Kiesler KM, Vallone PM (2015b) Performance of a next generation sequencing SNP assay on degraded DNA. Forensic Sci Int Genet 19:1–9. https://doi.org/10.1016/j.fsigen.2015.04.010

Gjertson DW, Brenner CH, Baur MP et al (2007) ISFG: recommendations on biostatistics in paternity testing. Forensic Sci Int Genet 1:223–231. https://doi.org/10.1016/j.fsigen.2007.06.006

Gjertson DW, Morris JW (1995) Assessing probability of paternity and the product rule in DNA systems. Genetica 96:89–98. https://doi.org/10.1007/BF01441154

González-Herrera LJ, García-Aceves ME, Domínguez-Cruz MD et al (2020) A four-step mutation at D22S1045 in one complex paternity case when the brother of the alleged father hypothesis is evaluated. Int J Legal Med 134:1647–1652. https://doi.org/10.1007/s00414-020-02312-1

Goodwin W, Peel C (2012) Theoretical value of the recommended expanded European Standard Set of STR loci for the identification of human remains. Med Sci Law 52:162–168. https://doi.org/10.1258/msl.2011.011068

Gusmão L, Sánchez-Diz P, Calafell F et al (2005) Mutation rates at Y chromosome specific microsatellites. Hum Mutat 26:520–528. https://doi.org/10.1002/humu.20254

Guzmán-Alberto JC, Martínez-Cortes G, Rangel-Villalobos H (2019) Inference of maternal uniparental disomy of the entire chromosome 2 from a paternity test. Int J Legal Med 133:71–75. https://doi.org/10.1007/s00414-018-1811-y

Hamester FIR, da Silva DS, Leboute APM et al (2019) Slippage mutation rates in 15 autosomal short tandem repeat loci for forensic purposes in a Southeastern Brazilian population. Electrophoresis 40:2873–2876. https://doi.org/10.1002/elps.201800491

Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. Forensic Sci Int Genet 17:33–34. https://doi.org/10.1016/j.fsigen.2015.03.006

Henke L, Fimmers R, Josephi E et al (1999) Usefulness of conventional blood groups, DNA-minisatellites, and short tandem repeat polymorphisms in paternity testing: a comparison. Forensic Sci Int 103:133–142. https://doi.org/10.1016/S0379-0738(99)00077-8

Hohoff C, Schürenkamp M, Börchers T et al (2006) Meiosis study in a population sample from Afghanistan: allele frequencies and mutation rates of 16 STR loci. Int J Legal Med 120:300–302. https://doi.org/10.1007/s00414-006-0091-0

Hohoff C, Schürenkamp M, Brinkmann B (2009) Meiosis study in a population sample from Nigeria: allele frequencies and mutation rates of 16 STR loci. Int J Legal Med 123:259–261. https://doi.org/10.1007/s00414-008-0307-6

Huel RLM, Bašić L, Madacki-Todorović K et al (2007) Variant alleles, triallelic patterns, and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia. Croat Med J 48:494

Hykin SM, Bi K, McGuire JA (2015) Fixing formalin: a method to recover genomic-scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing. PLoS One 10:e0141579. https://doi.org/10.1371/journal.pone.0141579

Kaiser L, Seber GAF, Opitz JM (1983) Paternity testing: I. Calculation of paternity indexes. Am J Med Genet 15:323–329. https://doi.org/10.1002/ajmg.1320150216

LaRue BL, Ge J, King JL, Budowle B (2012) A validation study of the Qiagen investigator DIPplex® kit; an INDEL-based assay for human identification. Int J Legal Med 126:533–540. https://doi.org/10.1007/s00414-012-0667-9

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221. https://doi.org/10.1093/oxfordjournals.molbev.a040442

Lindner I, von Wurmb-Schwark N, Meier P et al (2014) Usefulness of SNPs as supplementary markers in a paternity case with 3 genetic incompatibilities at autosomal and Y chromosomal loci. Transfus Med Hemother 41:117–121. https://doi.org/10.1159/000357989

Lukka M, Tasa G, Ellonen P et al (2006) Triallelic patterns in STR loci used for paternity analysis: evidence for a duplication in chromosome 2 containing the TPOX STR locus. Forensic Sci Int 164:3–9. https://doi.org/10.1016/j.forsciint.2005.11.006

Morris JW, Garber RA, D'Autremont J, Brenner CH (1988) The avuncular index and the incest index. In: Mayr WR (ed) Advances in forensic haemogenetics. Springer Berlin Heidelberg, Berlin/Heidelberg, pp 607–611

Musgrave-Brown E, Ballard D, Balogh K et al (2007) Forensic validation of the SNPforID 52-plex assay. Forensic Sci Int Genet 1:186–190. https://doi.org/10.1016/j.fsigen.2007.01.004

Narkuti V, Vellanki RN, Gandhi KPC et al (2007) Microsatellite mutation in the maternally/paternally transmitted D18S51 locus: two cases of allele mismatch in the child. Clin Chim Acta 381:171–175. https://doi.org/10.1016/j.cca.2007.01.032

Niida Y, Ozaki M, Shimizu M et al (2018) Classification of uniparental isodisomy patterns that cause autosomal recessive disorders: proposed mechanisms of different proportions and parental origin in each pattern. Cytogenet Genome Res 154:137–146. https://doi.org/10.1159/000488572

Ochiai E, Osawa M, Tamura T et al (2016) Effects of using the GlobalFiler™ multiplex system on parent–child analyses of cases with single locus inconsistency. Legal Med 18:72–74. https://doi.org/10.1016/j.legalmed.2015.12.009

Payseur BA, Cutter AD (2006) Integrating patterns of polymorphism at SNPs and STRs. Trends Genet 22:424–429. https://doi.org/10.1016/j.tig.2006.06.009

Pereira R, Phillips C, Alves C et al (2009) A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis 30:3682–3690. https://doi.org/10.1002/elps.200900274

Phillips C, Fondevila M, García-Magariños M et al (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Forensic Sci Int Genet 2:198–204. https://doi.org/10.1016/j.fsigen.2008.02.002

Radtkey R, Feng L, Muralhidar M et al (2000) Rapid, high fidelity analysis of simple sequence repeats on an electronically active DNA microchip. Nucleic Acids Res 28:E17. https://doi.org/10.1093/nar/28.7.e17

Raja UM, Thilaga D, Mahalakshmi N et al (2020) Identification of a rare tri-allelic inheritance at the D8S1179 locus in a case of paternity testing. Meta Gene 24:100658. https://doi.org/10.1016/j.mgene.2020.100658

Rockenbauer E, Hansen S, Mikkelsen M et al (2014) Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. Forensic Sci Int Genet 8:68–72. https://doi.org/10.1016/j.fsigen.2013.06.011

Seo SB, King JL, Warshauer DH et al (2013) Single nucleotide polymorphism typing with massively parallel sequencing for human identification. Int J Legal Med 127:1079–1086. https://doi.org/10.1007/s00414-013-0879-7

Slooten K, Ricciardi F (2013) Estimation of mutation probabilities for autosomal STR markers. Forensic Sci Int Genet 7:337–344. https://doi.org/10.1016/j.fsigen.2013.01.006

Sobrino B, Carracedo A (2004) SNP typing in forensic genetics: a review. In: Forensic DNA typing protocols. Humana Press, Totowa, pp 107–126

Teama S (2018) DNA polymorphisms: DNA-based molecular markers and their application in medicine. In: Liu Y (ed) Genetic diversity and disease susceptibility. InTech

Ting JC, Roberson EDO, Miller ND et al (2007) Visualization of uniparental inheritance, Mendelian inconsistencies, deletions, and parent of origin effects in single nucleotide polymorphism trio data with SNPtrio. Hum Mutat 28:1225–1235. https://doi.org/10.1002/humu.20583

Ullah S, Garg RK, Noor F (2017) DNA perspectives of fixed and paraffin embedded human tissues as resource materials for the identification. Egypt J Forensic Sci 7:23. https://doi.org/10.1186/s41935-017-0027-5

Veremeichyk VM, Borovko SR, Kuzub NN (2015) Mutation rates at 17 STR loci in the Belarusian population. Forensic Sci Int: Genetics Suppl Ser 5:e314–e316. https://doi.org/10.1016/j.fsigss.2015.09.125

Wenk RE, Gjertson DW, Chiafari FA, Houtz T (2005) The specific power of parentage exclusion in a child's blood relatives. Transfusion 45:440–444. https://doi.org/10.1111/j.1537-2995.2005.04271.x

Willems T, Gymrek M, Highnam G et al (2014) The landscape of human STR variation. Genome Res 24:1894–1904. https://doi.org/10.1101/gr.177774.114

Yang Y, Xie B, Yan J (2014) Application of next-generation sequencing technology in forensic science. Genomics Proteomics Bioinformatics 12:190–197. https://doi.org/10.1016/j.gpb.2014.09.001

Yao Y, Yang Q, Shao C et al (2018) Null alleles and sequence variations at primer binding sites of STR loci within multiplex typing systems. Legal Med 30:10–13. https://doi.org/10.1016/j.legalmed.2017.10.007

Zhao Z, Zhang J, Wang H et al (2015) Mutation rate estimation for 15 autosomal STR loci in a large population from Mainland China. Meta Gene 5:150–156. https://doi.org/10.1016/j.mgene.2015.07.006

Utility and Applications of Lineage Markers: Mitochondrial DNA and Y Chromosome

18

Sara Palomo-Díez and Ana María López-Parra

Contents

S. Palomo-Díez
Laboratory of Forensic and Population Genetics, Legal Medicine, Psychiatry and Pathology Department, Medicine School, Complutense University of Madrid (UCM), Madrid, Spain

Grupo de Ciencias Forenses: Genética y Toxicología forenses, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain

International University of La Rioja (UNIR), Logroño, Spain
e-mail: sarapalomodiez@med.ucm.es

A. M. López-Parra (✉)
Laboratory of Forensic and Population Genetics, Legal Medicine, Psychiatry and Pathology Department, Medicine School, Complutense University of Madrid (UCM), Madrid, Spain

Grupo de Ciencias Forenses: Genética y Toxicología forenses, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain
e-mail: amlopezp@ucm.es; amlopezparra@med.ucm.es

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_16

Abstract

Mitochondrial DNA and Y chromosome genetic markers are characterized by their lineage markers, so they are inherited without recombination phenomena and nearly unaltered from generation to generation, with the difference that the mitochondrial DNA is transferred from mothers to all her descendants and the Y chromosome is inherited from father to all his sons (males). On one hand, these characteristics provide them the impossibility of being used to individualize or identify a specific individual; but on the other hand, these same characteristics bring a rich pool of possibilities like the study of families, lineages, or biogeographical origin, or the possibility of obtaining information from degraded samples or minimal signs, which in specific cases could provide interesting supplementary information or even, in the worst cases, be the only information available.

This chapter will be focused on the main characteristics, utility, and applications of these kinds of markers, focusing the attention on general applications and forensic applications in particular.

Keywords

Y chromosome · Mitochondrial DNA · Lineage markers · Biogeographical origin · Nonrecombining markers

Introduction

Lineage markers include mitochondrial DNA (mtDNA) and Y chromosome markers that are transferred directly from generation to generation, either from mother to child, in the case of mtDNA, or from father to son. Y chromosome markers and mtDNA sequence information are treated as linked markers with the entire profile inherited as a block.

Advantages of lineage markers are their capacity for estimation of the biogeographical origin or exclusion in paternity testing or identification of missing persons. Their principal disadvantage is their null capacity for establishing direct relationships between individuals.

Mitochondrial DNA

What Is the Mitochondrial DNA?

In eukaryotic cells, it is possible to discern two different kinds of DNA, the nuclear DNA located in the nucleus and inherited fifty-fifty by the paternal and the maternal lineages and the DNA located in the mitochondrial organelles, mtDNA.

The polymorphisms located along the full mtDNA have been deeply studied since its full sequencing took place in 1981 (Anderson et al. 1981) and have supported interesting information about individuals and human populations. Furthermore, in the forensic field, the mtDNA can provide relevant information in some cases, especially on degraded samples characterized by low-template DNA.

Localization, Origin, and Structure

mtDNA is located in the organelles responsible for respiratory, named mitochondria (Fig. 1), and responsible for obtaining ATP (adenosine triphosphate) for the cell maintenance and functions.

These organelles can divide and reproduce by themselves. They have their DNA which can be also replicated, and during cell division, it is secreted independently and randomly. Also, they have their ribosomes, to produce amino acids. All these facts point that mitochondria could be archaebacteria which was able to do the

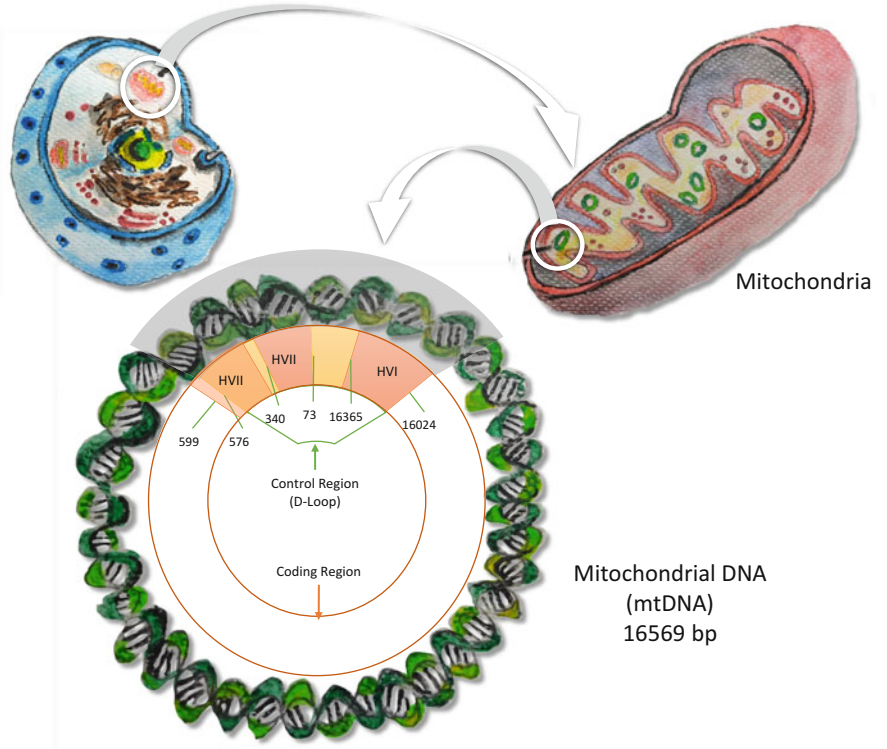


Fig. 1 mtDNA location and structure

oxidative phosphorylation to obtain energy (ATP), which was phagocytosed by a primitive eukaryotic anaerobic cell, providing to it this cellular respiration capacity.

The human mtDNA is concretely located in the mitochondrial matrix, and it is about 16.6 kb double chain circular molecule without histone proteins and with an asymmetric distribution in its strings because of its different density. One of its chains is denominated by heavy string (H), rich in guanines and adenosines; and the other is light string (L), rich in cytosines and thymines.

mtDNA is composed of 16.569 bp divided into two different regions (Fig. 1): a coding region and a noncoding region, also named control region, D-loop, or regulation region (Terra Pinheiro 2010; Crespillo Márquez and Barrio Caballero 2018).

From the forensic science point of view, the most interesting mtDNA region is the noncoding region, because it contains the most genetic polymorphisms, while from a phylogenetic perspective, it must be necessary to study the full mitochondrial genome.

The control region is divided into three regions: hypervariable regions I, II, and III. The hypervariable region I (HVI) ranges from the nucleotide number 16024 to the 16365 (being composed of 268 bp), the hypervariable region II (HVII) ranges from the nucleotide number 73 to the 340 (composed of 268 bp), and finally, the hypervariable region III (HVIII) ranges from 438 to 574 (composed of 137 bp) (Fig. 1). HVI and HVII show nucleotides more or less prone to mutation denominated hotspots. The D-loop region has a higher nucleotide substitution rate than nuclear DNA, mainly in the three hypervariable regions (Terra Pinheiro 2010).

mtDNA Characteristics

The own mtDNA characteristics, like kind of inheritance, recombination absence, and the upper number of molecules per cell, allow a greater probability of results obtained on problematic samples, like human skeletal remains or minimal signs, than autosomal or sex chromosome markers. This type of DNA shedding some light on those more complex cases to solve, or even in specific casuistic kinship, when the maternal lineage could provide interesting information to establish a familial relationship.

Moreover, this uniparental genetic marker shows specific geographical and ethnic distributions, involving the possibility of assigning a particular origin to an individual.

The main mtDNA characteristics are:

The High Number of Copies

In contrast to nuclear DNA, which only has two copies, the number of mtDNA molecules is much higher. On one hand, the number of DNA molecules oscillates from 2 to 10 per mitochondria. On the other hand, each cell contains hundreds to thousands of mitochondrial organelles, depending on the metabolic activity of the cell, being increased, for example, in oocytes or neurons. In any case, this kind of

DNA constitutes among 1–2% of the full cellular DNA. As a consequence, each cell contains hundreds to thousands of mtDNA copies. This high mtDNA molecules number makes the obtaining of results from critical samples, with low-template DNA or bad-quality DNA preservation, possible.

Matrilineal Inheritance

In mammals, during fecundation, only the spermatozoon's head penetrates the oocyte (which only contains nuclear genetic material, never mtDNA). In general, no spermatozoon's mitochondria are incorporated into the zygote. Rarely, some mitochondria could be incorporated into the oocyte (Schwarz and Vissing 2002), but in any case, the relative quantity of father's mtDNA comparing with mother's mtDNA is insignificant, and, overall, not enough to be detected in genetic analysis. Furthermore, the oocyte has recognition and elimination mechanisms of the father's original mitochondria (Crespillo Márquez and Barrio Caballero 2018). Behind, it is generally considered that mtDNA inheritance is exclusively maternal, being transmitted from the mother to all her descendants.

Heteroplasmy

The heteroplasmy is the coexistence of different mtDNA sequences in the mtDNA molecules' population in the cell, in a tissue, or in an individual (Ivanov et al. 1996).

The main cause of the heteroplasmy is that during the mtDNA replication, when mtDNA molecules replicate themselves independently, some chains could suffer mutation phenomena, and these mutations are inherited in some mitochondria, cells, or tissues. This fact leads to the heteroplasmic phenomena that can remain or disappear in subsequent generations of mitochondria. During the primary oocyte production, a limited group of mtDNA molecules is transferred to these new oocytes randomly, because when a heteroplasmic cell goes into division, the mitochondrial inheritance is random. So, each new oocyte could have different proportions of the different mtDNA variants. The mutated chains proportion in the cell depends on the bottleneck phenomena during the human oogenesis (Fig. 2).

The mutating molecules proportion can be different among the cell, or even among tissues, so that the heteroplasmic phenomenon supposes important repercussions to forensic genetic analyses and results interpretation. So, this phenomenon must be taken into account when comparing relatives, or even when comparing different samples from the same individual (Schwarz and Vissing 2002).

There is no consensus about the heteroplasmy frequency; on one hand, many authors support that it is infrequent; but on the other hand, other authors have determined that the frequency depends on the kind of sample analyzed, observing that meanwhile in blood cells it is not a frequent phenomenon, it is much more frequent in the hair (Terra Pinheiro 2010; Crespillo Márquez and Barrio Caballero 2018). Furthermore, the most common is to find chains of mtDNA that differ in a single position, with differences being rarely observed in two or more positions. Systematic studies have revealed that heteroplasmy occurs more often in tissues with high metabolic activity, and when mtDNA molecules pass through narrow bottlenecks during development.

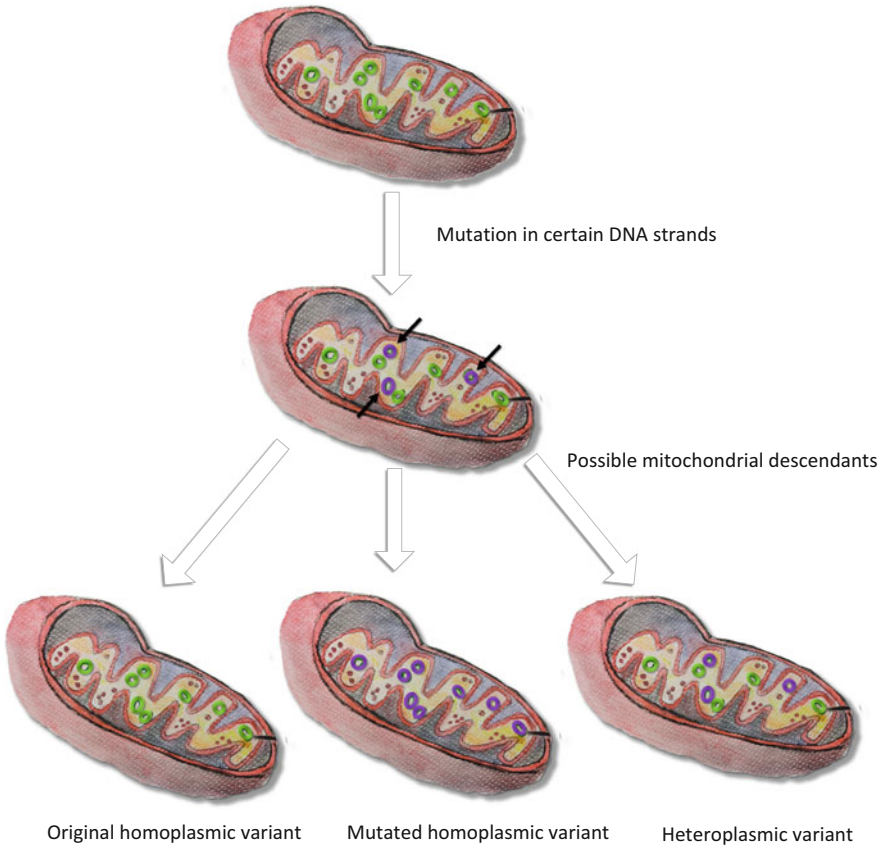


Fig. 2 mtDNA variant segregation

It is possible to differ two different kinds of heteroplasmy: point heteroplasmy and length heteroplasmy. The first one is referred to as a punctual position change, and the most commonly encountered has been the observation of two or even three heteroplasmic sites in a single sample. The second one, which is much more frequent than the positional heteroplasmy, is caused because of the variation of the number of bases in a homopolymeric tract.

To prove the presence of a point heteroplasmy, it is needed to observe the double peak in the same position in both chains, L and H. Moreover, it must be observed the phenomenon of both reading the forward and the reverse sequences, the sequence could be legible and above the baseline, and the lower peak must be at least 20% of the total area of the higher peak.

Concerning the length heteroplasmy, it is usually observed the HVI mono-polymeric transition on the 16,189 position, conducting to a homopolymeric tract between the 16,183 and 16,193 positions, because of the polymerase slippage. Therefore the interpretation of the sequence from this position is complicated and

must be analyzed by reading the complementary sequence or amplifying it with different primer pairs. Another typical length heteroplasmic tract is between positions 569 and 573 of HVIII, or between 309 and 315 positions of HVII. Although this phenomenon is quite common in the above-cited regions, there is no routine practice for its interpretation and reporting in forensic casework.

At forensic DNA analysis, the heteroplasmy phenomena can suppose a problem at the interpretation of the results, especially when it is necessary to differentiate it from a mixed profile, so that mtDNA usually is not taken into account in mixed profiles interpretation. Nevertheless, when the heteroplasmy diagnosis is sure, it also could be an excellent individualizing indicator (Ivanov et al. 1996).

One of the handicaps of heteroplasmy at the time of results interpretation is that, usually, the affected positions coincided with evolutionary hotspot mutations. Nevertheless, the list of heteroplasmic occurrences in Irwin et al. (2009) is a useful resource to compare observations in newly generated data. Moreover, recent studies have suggested that the incidence of heteroplasmy may be higher than previously reported. However, evaluation and validation of these observations for forensic purposes are subject to future research.

Looking at the interpretation of the results, from a forensic genetics point of view, it is important to note that the differences in both point and length heteroplasmic phenomena do not constitute evidence for excluding two otherwise identical haplotypes as deriving from the same source or same maternal lineage. The same point heteroplasmy observed in identical haplotypes may increase the strength of the evidence (Ivanov et al. 1996).

As is collected in recommendation number 10 of Parson et al. (2014) in forensic casework, laboratories must establish their interpretation and reporting guidelines for observed length and point heteroplasmy. The evaluation of heteroplasmy depends on the limitations of the technology and the quality of the sequencing reactions as well as the experience of the laboratory.

High Mutation Rate and Neutrality

mtDNA is not associated with histone proteins, and it achieves lower levels of compaction than nuclear DNA. So, mtDNA is much more susceptible to suffering the action of the free radicals. Furthermore, the mtDNA polymerase has a poor correction activity, so errors made during replication remain. These are the three reasons why mtDNA shows a higher mutation rate than nuclear DNA.

Despite this high mutation rate, mtDNA is much more polymorphic because its noncoding character allows that some mutations were fixed without any deleterious effect. Because of that, it is possible to speak about mtDNA neutrality.

Degradation Resistance

The small size and the circular structure of the mtDNA drive to less degradation of the molecule, because the endonucleases cannot damage it. Furthermore, the higher copy number compared with nuclear DNA makes there were much more mtDNA molecules available for replication, despite the endonucleases action. This

characteristic is fundamental in forensic genetics, particularly when facing with critical samples with low-template DNA.

NUMTs (Nuclear Mitochondrial DNA Fragments)

The term NUMTs refers to the nuclear DNA sequences which are homologous to the mtDNA. So, some mtDNA sequences are frequently transferred to the nucleus and incorporated into nuclear DNA. On one hand, its phenomenon supposes a problem when interpreting mtDNA analysis results because it is possible to amplify non-specific regions, but on the other hand, these NUMTs could be very useful to individualize specific persons (Bouhlal et al. 2013).

Analysis Methodologies and Results Interpretation

Nowadays, forensic genetic laboratories have standardized mtDNA analysis protocols. It is a relevant issue to interlaboratory comparing results or to carry out counter-expertises.

Into the mtDNA analysis process, it is possible to discern four main steps: extraction, quantifying, amplification, and sequencing. Also, last year's massive sequencing techniques have experienced great development in forensic genetics, and all the advantages of these new analysis strategies point out that they will replace the classic techniques.

Nevertheless, it will not be focused the attention on the technical strategies of analysis in this chapter, spotlighting on the analysis and interpretation of results.

Single-Nucleotide Polymorphisms (SNPs) and Other mtDNA Variations

The biallelic marker, SNPs, or single-nucleotide polymorphisms, comes from mutations that usually produce the change of a nucleotide base by another nucleotidic base in a punctual position of the DNA chain. This is the most frequent genetic variation in the human genome, and it is the main genetic variation in mtDNA.

SNPs show relevant characteristics to forensic genetics, especially when facing critical or degraded samples that usually contain fragmented DNA chains. The length of the DNA fragment to analyze to look for a specific SNP position is much shorter than to analyze an STRs. Nevertheless, its power of discrimination is lower than STRs, but this handicap can be overcome by the analysis of a higher number of analyzed loci.

Looking at the mtDNA characteristics, the SNP variation in the coding region is fewer than the variations observed in the control region, so the second one is much more interesting to individual discrimination.

Conversely, studying the coding region mutations can provide interesting medical information about specific pathologies; concretely there are about 250 diseases associated with mtDNA coding region mutations (Terra Pinheiro 2010). In forensic genetics, it is possible to analyze the noncoding region and also the coding region fragments located between genes, or silent positions of coding genes, which are considered noncoding regions.

Moreover, when analyzing an mtDNA sequence, to detect punctual position variations in comparison with an mtDNA reference sequence, the main changes that could be observed are:

- Transitions: It is a kind of substitution, specifically when a purine (adenine or guanine) is substituted by another purine or when a pyrimidine (cytosine or thymine) is substituted by another pyrimidine. These are the most often observed substitutions.
- Transversions: It is a kind of substitution, specifically when a purine is substituted by pyrimidine. This second kind of substitutions is rarely observed, showing lower frequencies than transitions.
- Insertions: A new nucleotidic base incorporation.
- Deletions: The loss of a nucleotidic base.

Nomenclature: IUPAC-IUB Code and How to Use It

The International Society for Forensic Genetics (ISFG) commission developed the first mtDNA typing recommendations in 2000 (Carracedo et al. 2000), where included was a specific point about the nomenclature. Nowadays there is a pool of actualized recommendations to assure the quality of the mitochondrial sequence (Parson et al. 2014). In any case, a universal nomenclature has been established to define all the nitrogenated bases or the different nucleotidic bases combinations in the cases of detecting heteroplasmic phenomena or mixtures. This universal nomenclature is based on the codes established by the International Union of Pure and Applied Chemistry and the International Union of Biochemistry (IUPAC-IUB) (Table 1).

To analyze an mtDNA sequence, it is essential to compare this sequence with a reference sequence and to take note of all the differences among both sequences. In this sense, it is possible to detect different kinds of discrepancies, and taking into account the above IUPAC-IUB codes, it is possible to nominate each one of these

Table 1 IUPAC-IUB codes

Nitrogenated base or nitrogenated bases combination	IUPAC-IUB code
Adenine	A
Thymine	T
Guanine	G
Cytosine	C
Adenine + guanine	R
Cytosine + thymine	Y
Adenine + cytosine	M
Guanine + thymine	K
Cytosine + guanine + thymine	B
Adenine + guanine + thymine	D
Adenine + cytosine + thymine	H
Adenine + thymine + guanine + cytosine	Y

discrepancies considering also the specific position of the difference and its characteristics.

Substitutions are named by the position of the string point changed preceded by the original nitrogenated base and followed by the new nitrogenated base; for example, the notation T16362C indicates a substitution (transition) from thymine to cytosine at the position 16,362; or the notation T239G indicates a substitution (transversion) from thymine to guanine at the position 239.

In the case of insertions, the number position is preceded by a script and followed by the new nitrogenated base incorporated. And finally, the deleted positions are represented by the position deleted followed by the letters "DELL" of "del."

As it is possible to observe, up to now all the codes are represented by capital letters. Nevertheless, the exposed nomenclature has some variations to differentiate the modifications that occurred in a mixture of two sequences. For example, code 152c would indicate the mixture of a transition and a deletion at nucleotide position T152. An example of a mixture composed of an undeleted variant at position C299, and a deleted variant in this same position (C299DEL), is named c299.

In summary, as it is collected at recommendation number 8 of Parson et al. (2014), IUPAC conventions using capital letters shall be used to describe differences to the rCRS and (point heteroplasmic) mixtures. Lowercase letters should be used to indicate mixtures between deleted and non-deleted (inserted and non-inserted) bases.

When there is more than one insertion in a poly C tract, the first insertion is denominated by the number position followed by a point and a 1 plus the C incorporated, for example, 309.1C; the second insertion will be named 309.2C, the third 309.3C, etc. The main difficulty of the insertions in these poly C tracts is that in a tract of, for example, seven Cs (between the positions 302 and 309), it is not possible to know where the insertion has taken place.

Results and Interpretation

According to the ninth recommendation of the DNA Commission of the ISFG for mtDNA typing (Parson et al. 2014), the alignment and notation of mtDNA sequences should be performed in agreement with the mitochondrial phylogeny. There are different tools to assist with the notation of mtDNA sequences, some of them available at http://empop.org.

The main objective of the mtDNA analysis is to compare the obtained result in a doubtful sample with the reference samples (undoubted). Each one of them is previously compared with the reference sequence (rCRS). The analysis results are expressed as discrepancies concerning this reference sequence. This pool of discrepancies defines the haplotype.

Once both (doubtful and reference samples) are defined, it is possible to compare them, obtaining three possible results:

- **Exclusion**: Both sequences are undoubtedly different, showing at least two different positions among them.

- **Non-exclusion**:
  - When both sequences show the same discrepancies concerning the reference sequence. In this case, the coincidence must be statistically evaluated.
  - When both sequences show the same discrepancies concerning the reference sequence and one of these discrepancies is a heteroplasmic position. In this case, the fact of observing the same heteroplasmy in both sequences reinforces the coincidence.
  - When both sequences show the same discrepancies concerning the reference sequence, except in one position, where only one of them shows a heteroplasmy.
- **Inconclusive result**: When there is only one discrepancy among the two analyzed samples and there is no heteroplasmy evidence or when this same situation occurs and also one of the samples shows another heteroplasmy in another position.

Haplotype and Haplogroup Concepts

Haplotype Determination

The haplotype of a specific sample is defined by the set of variations concerning the reference sequence that it shows and which are inherited as a block. From the forensic point of view, the haplotype usually is the main mitochondrial result objective, because it provides information about the maternal lineage being shared by all the individuals maternally linked.

Mitochondrial Reference Sequences (CRS, rCRS, RSRS)

As it has been hinted, to carry out the analysis and interpretation of mtDNA results, it is necessary to compare sequence results with a reference sequence. The first mtDNA reference sequence was named Cambridge reference sequence (CRS) or also Anderson sequence, named in this way in honor of the first author of the Sanger researching group, published in 1981 (Anderson et al. 1981). In 1999, the group of Andrews and collaborators reanalyzed the same material and determined some discrepancies from the first results, determining the new revised Cambridge reference sequence (rCRS), which nowadays is generally accepted as standard (Andrews et al. 1999) (disposable in the GenBank (NC_012920.1) and MITOMAP (http://www.mitomap.org/mitoseq.htm)).

It has been proposed a new alternative reference sequence, the RSRS (Behar et al. 2012), based on the fact that the rCRS is not appropriate from the phylogenetic point of view. So, Behar and collaborators, through this new sequence, offer a new sequence, which could be accepted as a phylogenetically more correct model. Nevertheless, the change of the reference rCRS sequence by this new proposal would carry interpretation difficulties; and so the ISFG recommends the use of the rCRS sequence as a reference.

The haplotype is established noting all the differences between the sequence and the reference sequence, noting each as was explained in point 2.

Haplogroup Diagnosis

An interesting characteristic of the mtDNA is how the mtDNA is nearly inalterable inherited through multiple generations along with human history, which has allowed establishing maternal lineages preserved along the time. This characteristic makes phylogenetic studies possible among different human populations. So, it has been possible to establish genetic mtDNA differences among human populations. Furthermore, through the mtDNA analysis, it is possible to infer an approximation of the biogeographical origin of an individual.

This fact conduces attention to the haplogroup term, defined as the set of haplotypes or sequences, which share certain specific polymorphisms, and which have a common origin. Haplogroups constituting an interesting tool in populational evolution studies.

Most of these haplogroups are defined by specific polymorphisms in specific positions, both in the control region and the codifying region, which are diagnostic positions. Moreover, because of the mtDNA nature, specific combinations of this polymorphism in specific positions are characteristics from determinate geographical areas, Europe, Asia, and Africa, and the different clades are subdivided into other subclades (Van Oven and Kayser 2009). The result is a specific mtDNA geographical distribution.

PhyloTree

All the mtDNA haplogroups are collected and organized in an enormous phylogenetical tree called PhyloTree (https://www.phylotree.org/) (Van Oven and Kayser 2009).

All the mtDNA haplogroups are denominated by the different alphabet letters, followed by numbers to indicate the subdivisions, or sub-haplogroups, into each one of them.

The PhyloTree is based on the hypothesis that all the actual mtDNA variants come from the same ancestor (mitochondrial Eve), which lived in Africa approximately 200,000 years ago, from which the rest of the current haplogroups come from.

Databases and Haplogroup Assignation Tools

The DNA Commission of the International Society for Forensic Genetics recommends the use of the PhyloTree to haplogrouping, but the manual haplogroup assignment can be laborious, and it could be facilitated by software applications to automate this task, like the software, EMMA, implemented in EMPOP VER.3. This software has the advantage that provides the affiliation of a sample together with the haplogroup distribution of the dataset providing useful information about the composition of the population, taking into account both the PhyloTree and an internal database. There are different additional tools for the haplogroup assignment, like HaploGrep (https://haplogrep.i-med.ac.at/) (Weissensteiner et al. 2016), or mtDNA Manager (http://mtmanager.yonsei.ac.kr/) (Lee et al. 2008).

Furthermore, the databases are not only useful to assign the haplogroup, but also they are essential to determine the meaning of finding the same haplotype in two independent individuals, depending on the frequency of appearance of this

haplotype in a given population; so the choice of an adequate database is an extremely important factor, because the database would accurately represent the pool of potential contributors of the mtDNA type without reference to the geographic origin of the suspect in question. It can be possible with homogenous populations, where a single local or regional database can be expected to represent the potential contributors.

To this aim it is necessary to use large databases, usually used by the different laboratories; the most used are the following:

- SWGDAM (Scientific Working Group on DNA Analysis Methods).
- GenBank (www.ncbi.nlm.nih.gov/Genbank).
- mtDB: regulated by Uppsala University (Sweden) (http://www.mtdb.igp.uu.se/),
- MITOMAP: from California University (http://www.mitomap.org).
- EMPOP (www.empop.org): which has also a tool to estimate mitochondrial DNA haplogroups using the maximum likelihood approach EMMA.

Statistical Evaluation

Once the haplotype is established by comparing it with a reference sequence, it is possible to compare the haplotype obtained of this doubtful sample with the reference sample. Once this comparison is performed, it is essential to assess statistically the coincidence among these sequences, to know the probability of specific biological remains belongs to a specific individual.

One of the most extended methods to the statistical evaluation is the *counting method*, based on the mounting of a determinate haplotype in the database, to estimate the frequency of this haplotype in the population. Nevertheless, when a determinate haplotype is not represented in a database, it could suppose a problem, because it could be interpreted as an extremely infrequent haplotype, but it could be that it is not represented in the database. On the other hand, the statistical parameter recommended by the ISFG to articulate the result is the LR (likelihood ratio), which is based on the contraposition of two alternative hypotheses. In both methods, the counting method and LR, it is necessary to make use of the databases commented on before. When the haplotype is not observed in the database, it is possible to apply some statistical parameter to correct. There are different mathematical tools to correct this deficiency derived from the number of sequences included in the databases. In forensic genetics, usually two different tools are applied: the Balding and Nichols method (1994) and the Holland and Parsons method (1999).

Both systems show similar results when the haplotype is not observed in the database. Nevertheless, when there are observed two or more similar sequences, the confidence intervals method must be used.

Forensic Applications of mtDNA

In contrast to nuclear markers, it must be said that mtDNA does not allow to individualize, and only is useful to establish maternal lineages.

Another disadvantage is that there are no larger databases to estimate the haplotype frequencies, conducting low discrimination power.

Furthermore, the high mutation rate of the mtDNA sometimes supposes a problem to compare the studied sequence with the reference sequence. For this reason, when finding one difference among them, it is not enough evidence for exclusion.

Nevertheless, this greater mutation rate of mtDNA than nuclear DNA makes the creation of species extremely useful in the matrilineal line for several generations, originating mtDNA variants.

Because of this higher mutation rate, matrilineal inheritance and no recombination, mtDNA can provide interesting supplementary data in many cases. The main forensic applications of mtDNA are:

Low-Template DNA Samples

Samples could show low-template DNA because of two main reasons: because they proceed from degraded samples and because the DNA extraction has been not sufficiently efficient.

In both cases, the higher number of mtDNA copies comparing with nuclear DNA is an advantage, so it has a higher probability of amplifying success.

Usually, the most common samples that show these low-template DNA problems are teeth, bones, or hairs; but it could be applied to any kind of sample.

One of the first examples where there was possible a human identification with mtDNA was the analysis of the skeletal remains of a disappeared child in 1984. It was possible to identify the child with the reference samples of his mother (Stoneking et al. 1991).

Also in 1990 a body of a female was discovered, in an advanced state of decomposition, and it was possible to identify her by comparing mtDNA sequences with a blood sample from a sister of the deceased female (Sullivan et al. 1992).

One of the most famous cases was the identification of Tsar Nicholas II and his family, also by skeletal remains analysis and by the study of mtDNA of living maternal relative of Tsar Nicholas II and Tsarina Alexandra (Gill et al. 1994; Ivanov et al. 1996).

Also, it has been an essential tool in many massive victims' events, like the case of the tsunami in Thailand in Southeast Asia, in 2004, when it might identify 200 victims by 200 reference relative samples, comparing mtDNA sequences from 258 tooth samples (Deng et al. 2005). Also in war events, such as the skeletal remains exhumed from a mass grave from the Spanish Civil War (1936–1939), where it was possible to identify a person by comparing mtDNA sequence with her putative sister (Ríos et al. 2010). In the same way, the remains of other missing soldiers were analyzed in Italy, both mtDNA and Y-STR markers, and compared with offspring of the Italian soldier Libero Zugni Tauro, showing, in this case, a clear exclusion (Piccinini et al. 2010). Another example is the case of the remains of eight people buried in a mass grave found at the cemetery Powązki Military in Warsaw, Poland, where the reference material was collected from the closest living relatives of Communist crimes' victims (1944–1956), and it was obtained a positive mtDNA match between six putative victims and six living relatives (Ossowski et al. 2016).

The Massive Parallel Sequencing have been applied also on mtDNA analysis, and thanks to these new technologies, it was possible to obtain the whole mitochondrial genome from a skeleton excavated at the presumed site of the Greyfriars friary in Leicester (2012) and compare it with saliva samples of the modern relatives of Richard III, obtaining a positive match allowing the identification (King et al. 2014).

But the skeletal remains (bones or teeth) are not the only kind of critical sample that can provide interesting mtDNA results; also the hairs, nails, or even paraffin-preserved tissues could be critical DNA sources (Cuenca et al. 2020).

When the Objective Is to Detect a Kinship Through the Maternal Lineage

mtDNA is especially interesting when there are no direct relatives to establish kinship, for example, the case of the Romanov family (Gill et al. 1994; Ivanov et al. 1996), or even in different war victims' identification, like the case of the Spanish Civil War (Ríos et al. 2010; Palomo-Díez et al. 2019) or other multiple-victim conflicts (Ossowski et al. 2016) that have been solved many years after occurring. In all these cases, the absence of direct relatives (daughters, or sons) results in difficulties in performing the identification by autosomal STRs, but it is possible to analyze mtDNA and compare it with a maternal relative.

However, when the results point to confirm identity by mtDNA, contrary to nuclear DNA, in the case of mtDNA, they do not refer to an individual but a group of individuals of the same maternal lineage.

To Approximate the Biogeographical Origin

The mitochondrial haplogroups are defined by specific combinations of polymorphisms which are characteristics from determinate geographical áreas: Europe, Asia, and Africa; and the different clades are subdivided into other subclades (Van Oven and Kayser 2009). The result is a specific mtDNA geographical distribution. So, it is possible to approximate the biogeographical origin of the individual, by the study of mitochondrial haplogroups.

A summary of the global mtDNA haplogroups distribution could be the next: The macro-haplogroup L, composed of the clades L0, L1, L2, and L3, is the most ancient haplogroup, and it is deeply spread in Africa. Later, the haplogroup L3 went out of Africa originating two Euroasiatic macro-haplogroups M and N. Next, N diverged into the large subclade R (European distribution), which includes a large set of haplogroups (B, F, HV, H, J, K, P, T, U, V, A, I, S, W, X, and Y); meanwhile, M diverged into the haplogroups C, D, E, G, Q, and Z. Later, the haplogroups A, B, C, and D colonized the Asian and American continents. Nowadays, the most commonly observed haplogroups in Afro-American populations are L2a, L1c, L1b, and L3b (Allard et al. 2005). Finally, the main haplogroups found in Asian populations are haplogroups M and N.

The haplogroup information can provide interesting information about the human population's origins and movements and the opportunity to analyze specific historical events with genetic information. But from the forensic point of view, it could be interesting, also, to study the biogeographical origin or the ancestrality of an individual. Nevertheless, it must be taken into account that this origin is only by

the maternal lineage, and this data could be completed with the Y chromosome analysis or even with autosomal DNA, to acquire a more complete view. Together this can help to resolve criminal acts.

New Perspectives: Twins Differentiation

In the last years, there have been observed different characteristics of mtDNA variations along with the individual life that could be employed to differentiate among twins, or also to inquire about the age of the individual.

Some studies about the mtDNA differences among monozygotic twins have been performed through a deep mtDNA genomic analysis, paying special attention to the NUMTs and the different heteroplasmic patterns among twins, both in coding and noncoding regions. It has been observed that the patterns heteroplasmy changes with age and in different ways in twins (Bouhlal et al. 2013).

It is essential to differentiate between heteroplasmy and replication mistakes (Bouhlal et al. 2013; Chocron et al. 2019). Several studies have presented compelling evidence for the key roles of mtDNA in age-related pathology (Chocron et al. 2019).

The future perspectives in mtDNA analysis point to the development of new massive parallel sequencing technologies (MPS), because it has been proven that MPS lead to the advantage of obtaining a large number of results with considerably less labor-intensive than Sanger sequencing, even analyzing different kinds of critical samples (Cuenca et al. 2020).

Y Chromosome

Origin of the Y Chromosome

The most widely accepted hypothesis since the 1960s is that the sex chromosomes of mammals evolved from a pair of homologous autosomal chromosomes (Ohno 1967). Probably, the Y chromosome appeared because one of the proto-sex chromosomes acquired a sex-determining locus (Graves 1998). In 1967, Ohno et al. proposed that, during the evolution of sex chromosomes, the X chromosome would have retained most of the autosomal genes it contained, while the Y chromosome would have lost them except for those genes involved in sex determination. Hence, the Y chromosome has been considered an extremely degenerate X chromosome (Skaletsky et al. 2003).

The continued loss of genes on the Y chromosome has been partly due to its inability to eliminate mutant alleles by recombination with a homologous chromosome, whereby mutations accumulate and many genes degenerate (Muller's ratchet) (Jobling et al. 2004).

Structure of the Y Chromosome

The Y chromosome is a small, acrocentric chromosome with only ~60 Mb, that is, approximately 1% of diploid DNA (Skaletsky et al. 2003). Two types of regions can

be distinguished fundamentally: nonrecombinant region and pseudoautosomal region (Fig. 3). The nonrecombinant region (NRPY, NRY) occupies 95% of the Y chromosome. Initially, and since it has no homolog, in this region there could be no recombination during meiosis. Therefore, some scientists considered that its true name is a male-specific region (MSY) because it is specific to males and undergoes recombination.

A mosaic of heterochromatic sequences and three classes of euchromatic sequences can be distinguished within this region: X chromosome transposon (X-transposed), degenerate segments of the X chromosome (X-degenerate), and ampliconic sequences (Skaletsky et al. 2003).

The X-transposed genes share 99% identity with the X chromosome (Xq21). The X-degenerate elements are composed of pseudogenes or single-copy genes with 60–96% homology to X-linked genes. The ampliconic segments are composed largely of sequences that exhibit marked similarity (99.9%) to other sequences in the MSY. The amplicons are located in seven segments that are scattered across the euchromatic long arm and proximal to the short arm and whose combined length is 10.2 Mb.

The ampliconic sequences exhibit by far the highest density of genes. Skaletsky et al. (2003) identified nine distinct MSY-specific protein-coding gene families, with copy numbers ranging from two, three, four, six, to approximately 35 units (TSPY). These coding families encompass roughly 60 transcription units. The ampliconic sequences include at least other 75 transcription units. All nine protein-coding families in the ampliconic regions are expressed predominantly or exclusively in tests. The ampliconic sequences exhibit by far the lowest densities of LINE1 and total interspersed repeat elements among the euchromatic sequence classes.

The euchromatic zone with 23 Mb codes for 27 different proteins with an average of 1.2 genes per Mb (Skaletsky et al. 2003), which is poor compared to the 717 genes in 160 Mb of the X chromosome (4.5 genes/Mb) (Jobling et al. 2004). The

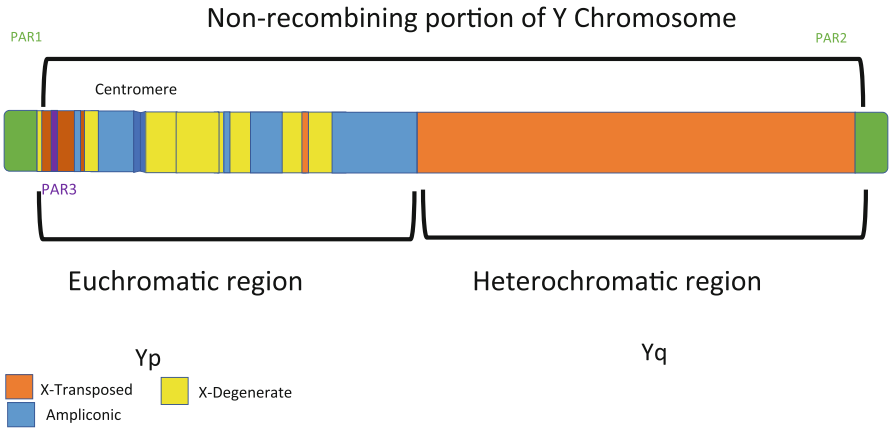


Fig. 3 Y chromosome structure

heterochromatic zone is usually variable in length. In some individuals, it represents more than half the length of the chromosome (Jobling et al. 2004).

At the ends of the arm of the chromosome are located the pseudoautosomal regions PAR1 (Yp) and PAR2 (Yq). These regions are present on both the X and the Y chromosome and recombine during meiosis (Mangs and Morris 2007). PAR1 with 2.6 Mb reflects the ancient origin of the sex chromosomes of mammals, being a region of forced recombination during male meiosis (Jobling et al. 2004). PAR2 with 0.32 Mb would be an evolutionarily more recent acquisition, specific to humans, and with little importance in chromosomal segregation (Jobling et al. 2004). PAR3 was identified in 2013 (Veerappa et al. 2013) and is approximately 2.3 Mb in length. It shares more than 98% sequence homology with Xq21.3 but was only identified in 2% of the general population.

Palindromes, which are inverted-repeat sequences, stand out on the Y chromosome. The ampliconic region contains eight palindromes (P1–P8) that span 5.7 Mb of the Y chromosome long arm. The most striking feature of the MSY palindromes is the high arm-to-arm sequence similarity. Excluding differences due to insertions and deletions, the arm-to-arm nucleotide identity present in the MSY reference sequence is always higher than 99.93% (Trombetta and Cruciani 2017). Of the protein families that encode Y chromosome, six families are located exclusively in palindromes, and two have some members in palindromes. The only family of the TSPY gene is not found in a palindrome.

In contrast to its scarce gene pool, the Y chromosome is rich in different types of repeats such as SINEs (short interspersed nuclear elements), endogenous retroviruses, and duplications. This causes increased susceptibility to recombinant rearrangements between non allelic homologous resulting in unusually high levels of structural polymorphisms between humans (Jobling et al. 2004). Among recombinant rearrangements, ectopic gene conversion stands out. The Y chromosome presents the longest uninterrupted stretches of gene conversion identified within the human genome (Hallast et al. 2013). It has been proposed that ectopic gene conversion may help in protecting them against deleterious mutations.

Given the high number of structural polymorphisms present on the Y chromosome, it is possible that natural selection eliminated chromosomes with an excessive or too low copy number in families of genes with multiple copies, as it is suspected to act in the gr/gr deletion in the AZFc region (Repping et al. 2003). Different studies have been carried out seeking to associate a haplotype or haplogroup with a protein phenotype without having obtained definitive results.

The choice of Y chromosome polymorphisms present in NRY in the analysis of the genetic structure of human populations is based fundamentally on three properties: present exclusively in males, located on a chromosome that has no homolog and does not present recombination. These properties have the consequence that the Y chromosome is inherited from parents to male children without changes, except for mutations that gradually accumulate, and that allows the genetic history of populations or individuals to be studied.

Polymorphisms of the Y Chromosome

Since the discovery of the first Y chromosome DNA polymorphism in 1985, the number and type of markers have increased exponentially. The ones most analyzed today are microsatellites and single polymorphisms nucleotides.

Y-STRs

Microsatellites or STRs (short tandem repeats) are highly polymorphic regions composed of a sequence of two to six base pairs that repeat in tandem, with the variation in the number of times the unit repeats the sequence. Microsatellites can be simple, when they contain chains of uninterrupted repeating units (monomers) with the same sequence, or complex, if they consist of more than one type of repeating unit and/or contain interruptions (Kayser et al. 2004).

Microsatellites were discovered almost simultaneously on autosomal chromosomes and on the Y chromosome (Y-STRs). In a first analysis of the human genome sequence, it was concluded that microsatellites constituted 3% of the genome.

Since the Y-STRs are located in the recombinant region, the set of Y-STRs that are analyzed in an individual is inherited as a block. The combination of alleles that are present on a Y chromosome is called a haplotype.

In 1997, the European forensic community established that the minimum haplotype (MH) used in identification should be constituted by nine systems, to standardize the use of Y-STRs in its field. An extended haplotype was defined later that includes three markers more. Currently, the main commercial kits used in the forensic field comprise a minimum of 23 Y-STRs.

In 2010, Ballantyne et al. described a set of 13 Y-STRs with mutational rates above 1.6x10–2 (rapidly mutating or RM Y-STRs), whereas most Y-STRs, including those currently used in forensic genetics, have mutational rates in the order of 1x10–3 or lower. The principal advantages of the RM Y-STRs are that they improve differentiation of unrelated individuals (Ballantyne et al. 2014) and they allow distinguishing closely related from more distantly related males. Recently, Ralf et al. (2020) have described new 12 RM YSTRs (Table 2). The set of newly identified 12 RM Y-STRs has a mean mutation rate of 2.6x10–2, which is higher compared with the set of previously identified 13 RM Y-STRs. Moreover, when combining all 25 RM Y-STRs, male relative differentiation capacity was 44% for the father-son pairs (1 meiosis), 69% for the brothers and grandfather-grandson pairs (2 meioses), 83% for the uncle-nephews (3 meioses), and 90% for the cousins (4 meioses) being differentiated by at least 1 mutation, respectively. For paternal relatives separated by eight meioses and above, over 99% were differentiated with this set of 25 RM Y-STR markers (Ralf et al. 2020).

Up to 41 Y-STRs are currently in the commercial kit the SureID® PathFinder Plus kit (HEALTH Gene Technologies Co., Ltd.), which includes the 17 loci from the Yfiler® kit, 14 rapidly mutating Y-STR loci, and 10 low-medium mutation loci, and 3 Y-indels.

Table 2 Rate mutation of Y-STRs. Classification according to how it appears in the corresponding publications

Marker	Combined mutation rate	N° mutations/meiotic transfers	Classification	Reference
DYF1001	0,0520	84/1616	RAPIDLY	Ralf et al. (2020)
DYS5724	0,0464	75/1616	RAPIDLY	Ralf et al. (2020)
DYF1000	0,0359	58/1616	RAPIDLY	Ralf et al. (2020)
DYR88	0,0291	47/1616	RAPIDLY	Ralf et al. (2020)
DYS712	0,0272	44/1616	RAPIDLY	Ralf et al. (2020)
DYS711	0,0266	43/1616	RAPIDLY	Ralf et al. (2020)
DYS1012	0,0192	31/1616	RAPIDLY	Ralf et al. (2020)
DYF1012	0,0179	29/1616	RAPIDLY	Ralf et al. (2020)
DYF404S1	0,0166	12/722	RAPIDLY	Fan et al. (2021)
DYS1007	0,0155	25/1616	RAPIDLY	Ralf et al. (2020)
DYS518	0,0148	40/2697	RAPIDLY	Fan et al. (2021)
DYS627	0,0148	43/2907	RAPIDLY	Fan et al. (2021)
DYS1010	0,0142	23/1616	RAPIDLY	Ralf et al. (2020)
DYS685	0,0142	23/1616	RAPIDLY	Ralf et al. (2020)
DYS576	0,0134	56/4165	RAPIDLY	Fan et al. (2021)
DYS1003	0,0130	21/1616	RAPIDLY	Ralf et al. (2020)
DYF387S1	0,0122	36/2945	RAPIDLY	Fan et al. (2021)
DYS527A/B	0,0122	9/739	RAPIDLY	Fan et al. (2021)
DYS449	0,0102	28/2758	RAPIDLY	Fan et al. (2021)
DYS570	0,0098	38/3865	RAPIDLY	Fan et al. (2021)
DYS533	0,0066	5/754	RAPIDLY	Fan et al. (2021)
DYS458	0,0063	58/9161		YHRD

(continued)

Table 2 (continued)

Marker	Combined mutation rate	N° mutations/meiotic transfers	Classification	Reference
DYS460	0,0055	14/2532		YHRD
DYS481	0,0054	4/747	RAPIDLY	Fan et al. (2021)
DYS549	0,0053	4/749	RAPIDLY	Fan et al. (2021)
DYS439	0,0048	69/12583		YHRD
DYS456	0,0045	41/9162		YHRD
DYS389II	0,0044	72/16334		YHRD
DYS635	0,0043	43/10009		YHRD
YGATAH4	0,0029	30/10193		YHRD
DYS385	0,0028	82/29715		YHRD
DYS444	0,0027	2/733	SLOWLY	Fan et al. (2021)
DYS389I	0,0026	43/16272		YHRD
DYS391	0,0024	41/17419		YHRD
DYS390	0,0022	38/16905		YHRD
DYS19	0,0021	38/18027		YHRD
DYS448	0,0015	14/9163		YHRD
DYS522	0,0014	1/702	SLOWLY	Fan et al. (2021)
DYS447	0,0014	1/712	SLOWLY	Fan et al. (2021)
DYS437	0,0013	16/12585		YHRD
DYS393	0,0012	20/16197		YHRD
DYS643	0,0010	3/3071	SLOWLY	Fan et al. (2021)
DYS461	0,0006	1/1695	SLOWLY	Baeta et al. (2018)
DYS525	0,0006	1/1712	SLOWLY	Baeta et al. (2018)
DYS561	0,0006	1/1783	SLOWLY	Baeta et al. (2018)
DYS392	0,0005	9/17350		YHRD
DYS438	0,0004	5/12606		YHRD
DYS388	0,0000	0/752	SLOWLY	Baeta et al. (2018)
DYS426	0,0000	0/1735	SLOWLY	Baeta et al. (2018)
DYS485	0,0000	0/1730	SLOWLY	Baeta et al. (2018)
DYS557	0,0000	0/748	SLOWLY	Fan et al. (2021)
DYS593	0,0000	0/722	SLOWLY	

(continued)

Table 2 (continued)

Marker	Combined mutation rate	N° mutations/meiotic transfers	Classification	Reference
				Fan et al. (2021)
DYS596	0,0000	0/749	SLOWLY	Fan et al. (2021)
DYS645	0,0000	0/712	SLOWLY	Fan et al. (2021)

Y-STRs are already being included in the new forensic kits for massively parallel sequencing (MPS) platforms. For example, ForenSeq™ DNA Signature Prep (Verogen) contains 29 auSTRs, 9 X-STRs, 24 Y-STRs, 86 auSNPs, 56 AIM-SNPs, and 22 SNPs of phenotypic information, or PowerSeq 46GY system that analyzes 22 auSTRs and 23 Y-STRs. Although MPS platforms are already widely implemented in many laboratories of forensic genetics, there are still some aspects that should be considered in this type of DNA analysis. It is necessary to have a definitive standard of MPS-based STR nomenclature, which is in progress led by the work of the STRAND ISFG Working Group.

Y Chromosome Microsatellite Mutation Rates

According to the hypothesis of male-driven evolution, there are a larger number of replications in male gametogenesis compared to the female one. Furthermore, the number of replications in spermatogenesis increases with time favoring the accumulation of mutations. Estimates of the alpha factor (the ratio of the mutation rate of males to females) based on Y and X chromosome sequences are 2.8 (95% confidence interval; 2.3–3.4), confirming that mutation rates in the male germline are higher than in the female one (Ebersberger et al. 2002).

The microsatellite polymorphisms of the Y chromosome used in the study of human populations are found in the nonrecombinant region and noncoding areas, so that their variability will be mainly due to mutations and is, therefore, selectively neutral. The most accepted mutation model for microsatellites is the stepwise mutation model (SMM). The original model assumes that a mutation alters the length of a set of repeating blocks through the addition or removal of a repeating unit at a fixed rate (a process that is independent of repeat length).

There is an alternative model that proposes that microsatellite length is determined by the balance between two types of mutational forces: length mutations and point mutations. Thus, the length mutations favor the increase in the size of the microsatellites, while the point mutations break the sets of large blocks of repeats into small units, thus explaining why the microsatellites do not expand continuously.

Mutation rates have been estimated by examining descendants of pedigrees that are separated by many generations, using cell lines, direct analysis of mutations in parents and children with confirmed paternity, and analysis in sperm. Mutation rates have also been carried out from the number of mutations in the branches of a

network of haplotypes of a population or within haplogroups defined by SNPs. From all these works a series of properties of the mutation process can be deduced (e.g., Jobling et al. 2004):

- Most mutations involve the increase or loss of a repeat unit.
- In general, the mutation rate increases concerning the length of the allele. Longer repeating unit blocks tend to contract, while shorter repeating unit blocks tend to expand. If the length of the repetition blocks is plotted against the mutation rates, a bell-shaped curve is observed that presents the allele frequency distribution in which the highest point represents the critical length of the modal allele.
- Dinucleotide loci mutate more rapidly than trinucleotides and tetranucleotides.
- Blocks of repeating units without breaks mutate faster than blocks of repeating units with breaks.

The mutation process appears to be variable according to the loci and type of repeat, with specific allelic mutation rates. It has also been previously suggested that some Y-STR markers may exhibit mutation rate differences between populations explained by different underlying Y-SNP haplogroups (Claerhout et al. 2018). The characteristics of the Y chromosome make it more sensitive to the forces acting on populations such as genetic drift, bottlenecks, or founding effects followed by isolations.

As regards the father's age, which may affect Y-STR mutability, contradictory results exist. Numerous papers describe a positive correlation. Subsequently, older fathers present more mutations (e.g., Gusmão et al. 2005). However, other studies do not report such a correlation, or only a small effect, which may be caused by sample size or differences in the complexity or sequence motifs between the studied Y-STRs.

In the same manner, it remains unclear if the DNA sequence of the repeat motif affects the Y-STR mutability (Ralf et al. 2020).

Interpretation of Y-STR Results in Forensic Analysis

Due to the lack of recombination, Y-STR loci are inherited along the paternal lineage. Therefore, the product rule of multiplying single-locus allele frequencies cannot be applied, and haplotype frequencies are needed for estimating Y-STR match probabilities. It is known that the haplotype distribution of Y-STRs is correlated with geographical and ethnolinguistic populations. Thus, the estimation of haplotype frequencies requires a reference database and a statistical model that considered their relationship with populations. Recently, DNA Commissions of the International Society of Forensic Genetics have published recommendations on the interpretation of Y-STR results in the forensic analysis (Roewer et al. 2020).

Some mutational effects rarely seen in autosomal STRs are more pronounced in Y-STRs. Like occurs with large-scale deletions, insertions and conversions are responsible for higher numbers of null and multiple alleles at certain loci. Large-scale deletions that can be greater than one megabase in size can affect some Y-STRs located in proximity. Some markers, as DYS385, show more than one allele because

the sequence has some copies on the Y chromosome. Also, duplications and triplications of some Y-STRs have been reported. Allelic dropout may occur due to molecular reasons, such as chromosomal rearrangements, deletions, primer binding site mutations, or stochastic effects caused by low amounts of DNA template or DNA degradation (Butler 2014). Moreover, the frequency of some mutational events depends on the population. All this is of forensic relevance because a pattern can be erroneously interpreted as allelic dropout, DNA contamination, and mixture, which may affect the evidential value of a DNA profile.

The probability of a match is evaluated using count or model-based estimators of the profile frequency. The proposed approaches with variable estimators (surveying, coalescence, discrete Laplace) mainly concern the frequency estimation of rare haplotypes, whereas frequencies of common haplotypes can be estimated using counts and Clopper-Pearson confidence intervals. Concerning the problem of a great number of unique haplotypes in the database, Brenner (2010) proposed the use of probabilistic estimates, as the kappa model, instead of frequency estimates. This model allows stronger evidence when a rare Y-STR haplotype that has never been seen in the database before matches with the suspect.

The most important database of Y-STR haplotypes is Y Chromosome Haplotype Reference Database (YHRD) (https://yhrd.org/). Currently (release 63, 8/2020), this database includes 321,471 haplotypes for minimal haplotype and information about 1363 populations. Haplotype frequencies are reported in metapopulations (geographically dispersed human population samples with shared genetic ancestry). The hierarchy of metapopulations incorporates linguistic and geographical terminology. But YHRD does not renounce the use of a national database, defined by political characteristics, or specific populations.

About the use of databases in forensic cases, as a comparison between a trace donor and a suspect, the location of the crime scene, and not the population from where the suspect is original, must determine the choice of the metapopulation. Only if additional information is provided by the court, the examiner should also report on a specific metapopulation (Roewer et al. 2020).

When the examiner compares the haplotype of the trace donor and the suspect, if both profiles differ by at least one allele, the conclusion must be “exclusion,” in which the verbal statement “someone other than the person (suspect) is the source of the DNA.” If the trace profile and suspect possess identical alleles at each locus, the conclusion is “non-exclusion” and needs evaluation. The recommendation is a quantitative assessment of the value of the match metapopulation of the suspect, the formulation of an alternative hypothesis, and the use of the likelihood ratio to evaluate the findings and a verbalization (Roewer et al. 2020).

When the examiner analyzes the haplotypes, he does not forget that some kits included RM Y-STRs. DNA Commission of the ISFG suggests the use of RM Y-STRs to further analyze trace and reference samples in case of a match for possible exclusion.

Y-SNPs

Biallelic markers, also known as binary markers, biallelic markers, or unique event polymorphisms (UEP), include both single-base substitutions in DNA (single-nucleotide polymorphisms or SNPs) and small insertions or deletions (indels or DIPs).

In 1997, Underhill et al. published the identification of 19 new biallelic polymorphisms based on PCR that had been identified by applying the DHPLC technique (Underhill et al. 1997). In this way, the doors were opened to the identification and typing of hundreds of new SNPs, insertions, and deletions. Y-SNP alleles can be designated as "ancestral" or "derived." The ancestral state of a Y-SNP is usually determined by comparison to a chimpanzee DNA sequence.

Estimates based on sequencing studies show that there is an average of 1 change per 10,000 bp between 2 randomly chosen NRY sequences, which is very low compared to the average rate of 1 SNP every 1000–2000 bp in the rest of the genome. Furthermore, in this region there is an abundance of unique SNPs (singleton); that is, in the derived state, they have only been found on a Y chromosome compared to the common ones.

Y Chromosome Consortium, in its nomenclature proposal, considers a haplogroup as to lineages in NRY defined by binary markers (Y-SNPs). The term haplotype refers to all sublineages of haplogroups that are defined by the combination of alleles of Y-STRs.

Origin and Rates of Mutation in Y Chromosome Biallelic Markers

Two processes generate biallelic markers: erroneous incorporation of nucleotides during replication and mutagenesis phenomena. However, for a sequence change to pass to the next generation, it must bypass DNA repair mechanisms (Jobling et al. 2004).

Mutation rates for Y chromosome biallelic markers are much lower than STRs (10–8 versus 10–3 mutation per generation), although there is evidence of recurrence in some Y-SNPs.

Regarding the identification of SNPs, it has been observed that the estimated SNP density is especially high in duplicate regions (from 0.69 per Kb to 1.33 per Kb; Jobling et al. 2004), indicating the possibility that some of those SNPs were PSVs (paralogous sequence variants or paralogs), that is, substitutions in very similar, nonallelic sequences, generated by duplication. It has been estimated that approximately 10% of the SNPs in the databases are PSVs. In the specific case of the Y chromosome, the frequency of paralogs exceeds the genomic average, being between 30 and 45% (Skaletsky et al. 2003). Also, an important property of paralogous sequences is their propensity for gene conversion – nonreciprocal transfer from one parallel copy to another – which may result in greater diversity in the sequences of the duplicated regions than the non-duplicated ones (Jobling et al. 2004).

Underhill et al. (1997) constructed the first phylogenetic tree from 22 SNPs. The low mutation rate of SNPs allows considering them as unique events in the evolution of humans. The absence of recombination and little recurrence allows the sequential succession of mutations to incorporate haplogroups in a phylogeny using the principle of maximum parsimony. This tree has been expanding as new SNPs have been identified. It is regularly updated by the International Society of Genetic Genealogy (ISOGG).

Applications of Y Chromosome

The Y chromosome testing is useful for several different applications of human genetics, including forensic evidence examination, paternity testing, historical investigations, biogeographical origin, human migrations, and genealogical research.

Forensic Analysis

The first application of the Y chromosome in forensic DNA is to support the presence of male DNA in a sample. The Y chromosome is present exclusively in males. For example, in the sexual assault with samples from azoospermic or vasectomized males, or samples without sperm, the analysis of markers of Y chromosome supports the presence of male DNA. Analysis of markers of Y chromosome can be used to verify males with mutations in the gen of amelogenin of the Y chromosome.

Moreover, interpretable results can be obtained in cases where autosomal results are limited by the presence of high levels of female DNA and minor amounts of male DNA. In cases of sexual assault with a mixture of samples of different males, as gang rape, it may be easier to analyze with markers of the Y chromosome than with autosomal STRs. But it is important not to forget that a match with markers of Y chromosome only means that this DNA is shared by all men of this lineage. Thus, a match of markers of Y chromosome between a suspect and the evidence has not the same meaning as in the case of autosomal STRs. Markers of Y chromosome cannot distinguish among paternal relatives.

But haplotyping of markers of Y chromosome is very useful for excluding suspects from involvement in a crime by demonstrating nonmatching haplotypes. This shared inheritance between men of the same lineage permits an increase in the number of relatives in comparison with missing person investigations, mass disaster victim identification, or familial searching. Analysis of Y-STRs in familial searching may require the use of RM Y-STRs to eliminate the more distant male relatives and to focus on the closely related ones.

Paternity Testing and Historical Investigations

In paternity testing, it must be used autosomal markers as auSTRs. But in cases where the father is dead or unavailable for testing, markers of Y chromosome could be used. Thus, for example, the analysis of markers of the Y chromosome was applied in the paternity dispute of US President Thomas Jefferson (Foster et al.

1998). Historically, a debate has existed over whether Thomas Jefferson was the father of some of the six children of his slave, Sally Hemings. Since Thomas Jefferson himself had no surviving sons, a direct comparison between him and Sally Hemings' offspring could not be made. Foster et al. (1998) analyzed DNA from other male members of the Jefferson family, five male-line descendants of two sons of Field (Thomas Jefferson's paternal uncle) and three male-line descendants of three sons of John Carr (Jefferson's nephews), and compared them with samples from two sons of Sally Hemings (male descendants of Thomas Woodson and Eston Hemings Jefferson).

Results of 11 Y-STRs and 7 Y-SNPs found that there was a match between descendants of Eston Hemings Jefferson and relatives of Thomas Jefferson and not with descendants of Thomas Woodson. Authors concluded that according to this result, "he was the biological father of Eston Hemings Jefferson." In a letter, authors clarified that Thomas Jefferson's brother could have been also the father of Sally Hemings' later children. Moreover, the exclusion with descendants of Thomas Woodson cannot exclude the paternity because there is the possibility of non-parenting events between Thomas Woodson (born in 1790) and their progenies across generations.

Genealogical Studies

In many modern societies, the transmission of the father's surname is linked with the biological inheritance of the male Y chromosome. The possible correlation between a specific surname and a specific haplotype or haplogroup required a series of conditions: (a) the surname must have a unique origin, (b) there should be no illegitimacy, which would introduce chromosomes from other surname groups, and (c) chromosomes associated with different surnames must have been unrelated at the time of surname establishment (Jobling, 2004). This link may be disrupted in cases of adoption, a surname change, a maternal surname transmission, a (hidden) baby exchange, and also independently co-founded surname, or an extra-pair paternity event. For example, the rate of non-biological paternity is very small (1–2%), estimated in different populations.

The first study about the heredity of surnames and Y-STR was done by Bryan Sykes with his surname (Sykes and Irven 2000).

The correlation of Y chromosome with surnames could be influenced by surname frequency and/or specific geographical origin of the surnames. King and Jobling (2009) investigated males with 150 different British surnames. They observed a better correlation in rare surnames.

Recently, Claerhout et al. (2020) have analyzed 2401 men from Belgium and the Netherlands. They observed the highest surname match frequency for the Belgium samples. The lower surname match frequency for the Dutch subset could be because the Dutch heredity surname traditions came into use later than in Belgium. So, the age of hereditary surname practices is an important factor for the correlation between surnames and Y chromosome. Besides, they observed a strong negative exponential correlation between the geographical distance and the surname match frequency. There are numerous studies about the correlation between surnames and Y-STRs

and/or Y-SNPs. For example, Martinez-Gonzalez et al. (2012) carried out a study on the surname Colon. One of the objectives of the work was to determine the origin of the famous discoverer who carried this surname. They analyzed samples of men with the surname Colon and surnames similar to Colon from Catalonia, Valencia, the Balearic Islands, southern France, and northern Italy.

They obtained a haplotype diversity in Italian men who shared the surname Colombo, similar to the general population. These researchers raised that the cause was that this surname was given until 1825 to the orphans hosted by the orphanage at the Ospedale Maggiore in Milan because a dove figured prominently in the crest of the Ospedale Maggiore. So, these men would share the surname but not the paternal lineage. About the origin of Christopher Columbus concerning Y chromosome haplotype, they obtained four lineages with a cumulative frequency of such lineages of 71, 87, and 82% in the Catalan, Valencian, and the Balearic Coloms, respectively, while they are only 18 and 0% in the Ligurian and Lombard Colombos. Therefore, identification would be much more likely for a Catalan than for a Ligurian Columbus (Martinez-Gonzalez et al. 2012).

Theoretically, the correlation between surnames and markers of Y chromosome could eventually provide a label of regional and familial relationships. This could be useful for familial searching to identify the surname of a perpetrator from his DNA left at the crime scene. The model of Claerhout et al. (2020) estimated a surname prediction sensitivity equal to 72%. About this, the case of the killer of Marianne Vaatstra is an example of the usefulness of genetic genealogy. In this case, police thought that the perpetrator likely comes from the region of the killer scene. They collected 6600 samples from local men. Before police analyzed Y-STR in the first set of 81 volunteers, they found 2 Y-STR haplotype matches with the semen trace found in the body's victim. Autosomal STR profiling excluded both men as likely suspects who had different surnames, but police found that they shared the same paternal ancestor at a time before the Dutch were forced to have their surnames registered. Thus, they decided to select samples from volunteers with these two surnames, and also, they decided to use RM Y-STRS to exclude distant male relatives, decreasing the suspect pool. Surprisingly, one of the volunteers with a Y-STR profile match showed an autosomal STR profile match with the semen trace. It is a clear example of how genetic genealogists could help to resolve a case.

On the other hand, the analysis of the surnames and markers of the Y chromosome has had a great interest in the general population. Thus, companies have emerged that, in addition to analyzing the Y chromosome markers, have developed projects for each surname, allowing communities of men who share the surname and with a great interest in knowing their origins to be created. On many occasions these citizen sciences have genealogical information, managing to identify their shared ancestors.

Biogeography Ancestry and Study of Migrations of Human

Y chromosome polymorphisms are applied for inferring the paternal biogeography ancestry. This is due to the absence of recombination. Once a mutation has occurred, it is inherited for all males of this paternal lineage. Moreover, the effective

population size of the Y chromosome is four times smaller than that of autosomes. For every four autosomal chromosomes, there is a Y chromosome if it is a male. This makes the Y chromosome more sensitive to genetic drift phenomena, and therefore more susceptible to presenting a population substructure and greater divergence between populations. Migrations driven by males, patrilocality, and polygyny increase Y chromosome differences over geographic distance.

One of its main applications has been in the analysis of the genetic structure of populations since its mutations could theoretically be related to population events in a specific geographic and temporal context. Because of the lower mutation rates of Y-SNPs relative to Y-STRs, geographic ancestry signatures are much longer at Y-SNPs, before being diluted via mutations to Y-STRs (Kayser et al. 2017). So, there are Y-SNP haplogroups with predominant occurrence in specific geographic regions. On some occasions, the geographic correlation with the Y-SNP haplogroups can be inferred from associated Y-STR haplotypes. Polymorphism of Y chromosomes has been used to trace biogeographic ancestry of populations to resolve important questions about population origins and migration history worldwide.

Final Conclusion

Along with this chapter, it has been possible to know the main characteristics, utility, and applications of mtDNA and Y chromosome.

Over the last years, mtDNA and Y chromosome typing have been widely used around the world to solve human identifications, to obtain ancestors' information, and to analyze biogeographical origin. The progress in DNA typing has been astonishing, going from the examination of small fragments in a matter of days to sequencing multiple entire DNA genomes in a couple of hours. It has been demonstrated by many referenced examples the great utility of these lineage markers, and there have been given some hints about the future of these markers in human genetic typing.

References

Allard MW, Polanskey D, Miller K, Wilson MR, Monson KL, Budowle B (2005) Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. Forensic Sci Int 148(2–3):169–179. https://doi.org/10.1016/j.jorsciint.2004.06.001

Anderson S, Bankier AT, Barrell BG, de Bruign MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290(5806):457–465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23(2):147

Baeta M, Núñez C, Vilaescusa P, Jiménez-Moreno S, Martínez-Jarreta B, de Pancorbo MM et al (2018) Assessment of a subset of Slowly Mutating Y-STRs for forensic and evolutionary studies. Forensic Sci Int Genet 34:E7–E12. https://doi.org/10.1016/j.fsigen.2018.03.008

Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64 (2–3):125–140

Ballantyne KN, Ralf A, Aboukhalid R, Achakzai NM, Anjos MJ, Ayub Q et al (2014) Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. Hum Mutat 35(8): 1021–1032

Behar DM et al (2012) A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90(4):675–684

Bouhlal Y, Martinez S, Gong H, Dumas K, Shieh JTC (2013) Twin mitochondrial sequence analysis. Mol Genet Genomic Med 2013 1(3):174–186

Brenner CH (2010) Fundamental problem of forensic mathematics – the evidential value of a rare haplotype. Forensic Sci Int Genet 4:281–291

Butler JM (2014) Advanced topics in -forensic DNA typing: interpretation. Elsevier

Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA Commission of the International Society for forensic genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110(2):79–85

Chocron ES, Munkàcsy E, Pickering AM (2019) Cause or Casualty: the role of mitochondrial DNA in aging and age-associated disease. Biochim Biophys Acta Mol Basis Dis 1865(2):285–297. https://doi.org/10.1016/j.bbadis.2018.09.035

Claerhout S, Vandenbosch M, Nivelle K, Gruyters L, Peeters A, Larmuseau MHD, Decorte R (2018) Determining Y-STR mutation rates in deep-routing genealogies: identification of haplogroup differences. Forensic Sci Int Genet 34:1–10. https://doi.org/10.1016/j.fsigen.2018.01.005. Epub 2018 Jan 17

Claerhout S, Roelens J, Van der Haegen M, Verstraete P, Larmuseau MHD, Decorte R (2020) Ysurnames? The patrilineal Y-chromosome and surname correlation for DNA kinship research. Forensic Sci Int Genet 44:102204

Crespillo Márquez MC, Barrio Caballero PA (2018) Genética Forense: Del laboratorio a los tribunales. Editorial: Díaz de Santos, Madrid. ISBN: 978-84-9052-213-4

Cuenca D, Battaglia J, Halsing M, Sheehan S (2020) Mitochondrial sequencing of missing persons DNA casework by implementing Thermo Fisher's precision ID mtDNA whole genome assay. Genes 2020(11):1303. https://doi.org/10.3390/genes11111303

Deng YJ, Li YZ, Yu XG, Li L, Wu DY, Zhou J, Yu J (2005) Preliminary DNA identification for the tsunami victims in Thailand. Genomics Proteomics Bioinformatics 3:143

Ebersberger I, Metzler D, Schwarz C, Pääbo S (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet 70(6):1490–1497

Fan G, Pan L, Tang P, Zhou Y, Liu M, Luo X (2021) Technical note: developmental validation of a novel 41-plex Y-STR system for the direct amplification of reference samples. Int J Legal Med 135(2):409–419. https://doi.org/10.1007/s00414-020-02326-9. Epub 2020 Jun 10

Foster EA, Jobling MA, Taylor PG, de Knijff P, Mieremet R, Zerjal T, Tyler-Smith C (1998) Jefferson fathered slave's last child. Nature 396(6706):27–28

Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6:130

Graves JAM (1998) Interactions between SRY and SOX genes in mammalian sex determination. BioEssays 20:264–269

Gusmão L, Sánchez-Diz P, Calafell F, Martín P, Alonso CA, Alvarez-Fernández F et al (2005) Mutation rates at Y chromosome specific microsatellites. Hum Mutat 26(6):520–528

Hallast P, Balaresque P, Bowden GR, Ballereau S, Jobling MA (2013) Recombination dynamics of a human Y-chromosomal palindrome: rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions. PLoS Genet 9:e1003666

Holland MM, Parsons T (1999) Mitochondrial DNA sequence analysis-validation and use for forensic casework. Forensic Sci Rev 11(81):21–50

Irwin J, Saunier JL, Niederstätter H, Parsons T et al (2009) Investigation of Heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. J Mol Evol 68(5):516–527. https://doi.org/10.1007/s00239-009-9227-4

Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ, Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of tsar Nicholas II. Nat Genet 12(12):417–420. https://doi.org/10.1038/ng0496-417

Jobling MA, Hurles M, Tyler-Smith C (2004) Human evolutionary genetics: origins, peoples and disease. Garland Science, New York

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A et al (2004) A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74(6):1183–1197

King TE, Jobling MA (2009) Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. Mol Biol Evol 26(5):1093–1102

King TE, Fortes GG, Balaresque P, Thomas MG, Balding D, Delser PM, Schürer K (2014) Identification of the remains of King Richard III. Nature Communications

Lee HY, Song I, Ha E et al (2008) mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. BMC Bioinformatics 9:483. https://doi.org/10.1186/1471-2105-9-483

Mangs AH, Morris BJ (2007) The human pseudoautosomal region (PAR): origin, function, and future. Curr Genomics 8:129–136

Martinez-Gonzalez LJ, Martinez-Espín E, Alvarez JC, Albardaner F, Rickards O, Martinez-Labarga C et al (2012) Surname and Y chromosome in southern Europe: a case study with Colom/Colombo. Eur J Hum Genet 20(2):211–216

Ohno S (1967) Sex chromosomes and sex-linked genes. Springer, Berlin/New York

Ossowski A, Diepenbroek M, Kupiec T, Bykowska-Witowska M, Zielińska G, Dembińska T, Ciechanowicz A (2016) Genetic identification of communist crimes' victims (1944–1956) based on the analysis of one of many mass graves discovered on the Powazki military cemetery in Warsaw, Poland. J Forensic Sci 61:1450

Palomo-Díez S, Gomes C, López-Parra AM, Baeza-Richer C, Cuscó I, Raffone E, García-Arumi E, Arroyo-Pardo E et al (2019) Genetic identification of Spanish civil war victims. The state of the art in Catalonia (Northeastern Spain). Forensic Sci Int Genetics Supp Series 2019 7:419–421

Parson W, Gusmao L, Hares DR, Irwin JA, Mayr WR, Morling N, Pokorak E, Prinz M, Salas A, Schneider PM, Parsons TJ (2014) DNA Commission of the International Society for forensic genetics: revised and extended guidelines for mitochondrial DNA typing. Forensic Sci Int Genet 13(2014):134–142

Piccinini A, Coco S, Parson W, Cattaneo C, Gaudio D, Barbazza R, Galassi A (2010) World war one Italian and Austrian soldier identification project: DNA results of the first case. Forensic Sci Int Genet 4(5):329–333. https://doi.org/10.1016/j.fsigen.2009.10.005

Ralf A, Lubach D, Kousouri N, Winkler C, Schulz I, Roewer L et al (2020) Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers. Hum Mutat 41(9):1680–1696

Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T et al (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35(3):247–251

Ríos L, Ovejero JI, Prieto JP (2010) Identification process in mass graves from the Spanish Civil War I. Forensic Sci Int 199(1–3). https://doi.org/10.1016/j.forsciint.2010.02.023

Roewer L, Andersen MM, Ballantyne J, Butler JM, Caliebe A, Corach D et al (2020) DNA commission of the International Society of Forensic Genetics (ISFG): recommendations on the interpretation of Y-STR results in forensic analysis. Forensic Sci Int Genet 48:102308

Schwarz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. N Engl J Med 347(8)

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423(6942):825–837

Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48(2):370–382

Sullivan KM, Hopgood R, Gill P (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Legal Med 105(2):83–86. https://doi.org/10.1007/BF02340829

Sykes B, Irven C (2000) Surnames and the Y chromosome. Am J Hum Genet 66(4):1417–1419

Terra Pinheiro MF (2010) Genètica forense: Perspectivas da identificaçao genética. Ediçoes Universidad Fernando Pessoa. ISBN: 978-989-643-050-4

Trombetta B, Cruciani F (2017) Y chromosome palindromes and gene conversion. Hum Genet 136 (5):605–619

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7(10):996–1005

Van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2):E386–E394

Veerappa AM, Padakannaya P, Ramachandra NB (2013) Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome transposed region (XTR) in the Y chromosome. Funct Integr Genomics 13:285–293

Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, Kronenberg F, Salas A, Schönherr S (2016) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucl Acids Res. https://doi.org/10.1093/nar/gkw233

Usefulness of the X-Chromosome on Forensic Science

19

Cláudia Gomes and Eduardo Arroyo-Pardo

Contents

Abstract

In kinship analyses, as well as in other forensic studies, autosomal markers are, generally, the tool of choice. However, there are some more complex situations where the study of these markers is not sufficiently informative and the analysis of the available set of markers proves to be inconclusive. Due to its characteristic transmission properties, the X-chromosome markers have emerged as an asset in certain genealogical situations, not only complementing and completing autosomal marker information but also by solving alone certain kinship studies irresolvable with autosomal markers. Examples of this application can occur in complex genealogical analyses, or when the alleged father is not available

C. Gomes · E. Arroyo-Pardo (✉)
Laboratory of Forensic and Population Genetics, Legal Medicine Psychiatry and Pathology Department, Medicine School of the Complutense University of Madrid, Madrid, Spain

Forensic Sciences: Forensic Genetics and Toxicology Group, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos IdISSC, Madrid, Spain
e-mail: clopes01@ucm.es; eduardoa@ucm.es

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_17

(in such cases the statistical evidence given by autosomal markers is, in most cases, ambiguous), or when the kinship alternatives are genetically indistinguishable using autosomal markers (e.g., as in the cases of half-siblings, avuncular, or grandparent-grandchild kinships).

Keywords

X-chromosome · X-STRs · X-InDels · X-SNP · Kinship analysis

Introduction

The specific mode of transmission of the X-chromosome is what motivates the interest of its study, since its transmission depends on the sex of the individuals, providing a different approach than that obtained using autosomal markers. In normal female cells, there is a pair of X-chromosomes that recombine with each other, just like autosomes. As for male cells, the X-chromosome has no homologue, only recombining with the Y chromosome at the level of the homologous PAR1 and PAR2 regions. In this way, a father transmits to all his daughters a full copy of his X-chromosome, while a mother randomly transmits a copy of one of his two X-chromosomes to daughters and sons, after recombination, as for autosomes.

In this sense, male individuals have only one X-chromosome, which does not undergo recombination in its greatest extent, transmitting itself as a haplotype, in the same way, to all daughters. On the other hand, female individuals have two X-chromosomes, so the knowledge of their haplotypic phase is only possible based on paternal information or, for markers, whose recombination rate is practically null, by analyzing her son(s) (Szibor et al. 2005).

Thus, the specific mode of transmission of the X-chromosome, as well as the pattern of recombination it presents, makes the study of genetic markers located on this chromosome crucial both in the field of population and forensic genetics. As will be explained below, the X-chromosome markers have different characteristics, allowing to infer events related to the history of populations that are not easily detectable through the study of other types of markers. On the other hand, its mode of transmission allows, in kinship analysis, to increase the information power available based only on autosomal and/or uniparental markers.

The X-Chromosome Evolution

There are two models of genetic sex determination in the animal kingdom: the XY sex-determination system is found in most mammals, including humans. In this model, males are the heterogametic sex, XY, and females are homogametic, XX. Birds, butterflies, and many reptiles have a ZW sex-determination system,

where females are the heterogametic sex, ZW, and males are homogametic, ZZ. In general, human sex chromosomes are believed to have arisen by evolutionary differentiation from a pair of autosomal chromosomes within the last 300 million years (Myr) (Ohno 1967; Bull 1983; Rice 1996; Lahn and Page 1999). Therefore, it has been generally accepted that evolutionary differentiation has been larger in the heterogametic sex (Y/W chromosome), while the X/Z chromosome remained more similar to autosomes. In the different taxa, the differential selective pressure between the two sexes causes the genetic composition to vary. For example, in *Drosophila* the X-chromosome has a shortage of genes that are selectively expressed in males (Parisi et al. 2003), while it occurs in reverse in mammals (Wang et al. 2001).

For the X/Y model, these two chromosomes started to differentiate approximately 160 million years ago, shortly before the separation of the metatherians (marsupials) and eutherians (placental mammals).

We can enumerate four evolutionary forces behind this differentiation process. First, hemizygosity: in males, the only existing copy of the X-chromosome does not allow recombination except in small regions called "pseudoautosomal regions" (Pseudo Autosomal Region or PARs) that maintain homology during male meiosis and are present both in the X-chromosome and in the Y chromosome. According to some authors, PAR sequences have followed an evolutionarily special differentiation concerning the rest of the chromosomes, as there is a different selective pressure between both sexes (Lahn and Page 1999; Bachtrog 2006). According to Gurbich and Bachtrog (2008), recessive or partially recessive mutations are more efficiently fixed on the X-chromosome due to hemizygosity. Thus, given that the X-chromosome is hemizygous in human males, the fixation is particularly accentuated in the case of mutations that are beneficial to males (Rice 1984; Hurst 2001). This effect can lead to an increase of male-specific genes on the Y chromosomes, while the opposite occurs in the case of X-chromosome: since this chromosome is transmitted 2 out of 3 times through females and only 1 out of 3 times through males, selection favors mutations beneficial for females. This can lead to an accumulation of female-beneficial mutations on the X-chromosome. A second evolutionary force behind sex chromosome differentiation is sexual antagonism due to the selective effect of mutations in both sexes. For instance, some alleles can benefit one sex and not the other or can be deleterious for one sex. This effect is strongly increased in the X-chromosome since this chromosome shows the female-biased transmission and is hemizygous in males (Vicoso and Charlesworth 2006; Rice 1984). Third, the temporary inactivation of the Y chromosome and an X-chromosome during primary spermatogenesis in many animal species makes the X-chromosome a poor place to find specific male genes. Inactivated chromosomes lead subsequently to a temporary inactivation of gene transcription in silenced genes. As a consequence, male-specific genes can be selectively favored in autosomes where temporal inactivation simply does not occur. These two factors mean that the X-chromosome has undergone a special differentiation throughout evolution (Gurbich and Bachtrog 2008). Finally, the main cause of phenotypic differentiation between sexes is the differential

expression of genes or the presence of sex-biased genes (Connallon and Knowles 2005; Rinn and Snyder 2005). Some genes are more expressed or exclusively expressed in one sex, and this may be particularly relevant in the case of sex-biased genes located on sex chromosomes, due to the owing to the sexual antagonism abovementioned.

These four evolutionary forces can easily show how a mutation of one allele of the Sox3 gene yielded the male sex-determining gene SRY (sex-determining region Y) on the proto-Y of this chromosome pair (Sutton et al. 2011). According to Gribnau and Grootegoed (2012), the region around SRY increased male beneficial genes and modifications limiting progressively the possibility for meiotic recombination between the evolving heterologous sex chromosomes. Subsequently, the male-specific region of Y (MSY), inherited from father to son, could never recombine, while the X-chromosome could recombine in the meiotic prophase of oogenesis. The result was a progressive differentiation of two new chromosomes.

Structural Characteristics

The X-chromosome is one of the most stable nuclear chromosomes, retaining its relatively unalterable genetic makeup at least in Eutherian mammals (Lahn and Page 1999; Kohn et al. 2004; Schaffner 2004). It has a large number of repeated sequences (Gunter 2005). In 2005, Ross et al. published the first draft of the euchromatic sequence of the X-chromosome. It is approximately 156 million base pairs (156 Mb) long, representing 5% of the total genome (Lander et al. 2001). According to Ross and collaborators, the X-chromosome has a low content of GC pairs, slightly below the genome average (41%), and one of the lowest concentrations of functional genes. According to the VEGA database (human VEGA68), this chromosome has a total of 1992 genes (12.7 genes per Mb) that code for 839 proteins. The X-chromosome contains 870 pseudogenes (5.6 pseudogenes per Mb) and also the largest known gene in the human genome, the dystrophin (DMD) locus in Xp21.1, which spans 2,220,223 bp. The genes present are relatively short: only 1.7% of the total chromosome sequence is composed of exons that transcribe 33% of the X-chromosome (Ross et al. 2005). Another remarkable characteristic of the X-chromosome is that interspersed repeats account for 56% of the euchromatic X-chromosome sequence, while the genome average is 45% (Ross et al. 2005).

According to Giorgetti et al. (2016), X-chromosome inactivation (XCI) involves major reorganization of the X-chromosome, given that inactivation implies silencing and heterochromatic conversion. Female mammalian development triggers XCI through the non-coding XIST RNA in one of the X-chromosomes. XIST acts by coating the chromosome in cis and induces silencing of almost all genes via its A-repeat region. There is no, however, a total inactivation since some genes avoid silencing in most cell types and others escape silencing only in some cases. These genes are called constitutive and facultative escapees, respectively.

However, although the low density of genes and chromosome inactivation is probably the most remarkable characteristics of X-chromosome structure, the main relevant trait for forensic use is the model of heritage in sexual chromosomes: two recombining copies in females and a single non-recombining copy in males (except for the PAR regions) creating haplotypes (Gomes 2011; Gomes et al. 2020b).

Inside Theria, the human X-chromosome long arm and a proximal portion of the short arm correspond to genes on the marsupial X-chromosome. This domain, the X-conserved region (XCR), is shared by sex chromosomes of all live-bearing mammals. Nevertheless, the X-added region (XAR) is autosomal in marsupials and translocated to eutherian sex chromosomes between the divergence of marsupials and placental mammals ~148 million years ago and the eutherian radiation ~100 million years ago (Graves 1995; Bininda-Emonds et al. 2007).

Types of Markers and Their Usefulness on Forensic Science: X-STRs, X-InDels X-SNPs

The special characteristics of the X-chromosome that are showed in the previous section are different from those of the haploid markers – Y chromosome and mitochondrial DNA – and also from those of the autosomes. For this reason, the X-chromosome can solve a forensic casuistry that other markers cannot solve, especially in certain cases of kinship, as explained later. Genetic markers considered on the X-chromosome for forensic applications are short tandem repeat (STR) events, single-nucleotide polymorphisms (SNPs), and insertion-deletion (InDels).

Work with X-STRs began in the early 1990s of the last century (Edwards et al. 1991, 1992; Hearne and Todd 1991; Sleddens et al. 1992) and received a major boost in 2007 with Szibor's seminal work (Szibor 2007). More recently, Gomes, I and collaborators (2020b) have carried out an extensive review of the literature related to X-STRs for forensic use (2020). While these authors have concluded that a total of 85 STRs have been used for forensic and population-based purposes, a much smaller number of STRs have been considered for commercial standard and multiplex protocols (Diegoli 2015). These markers have been used mainly for relationship analysis due to the high polymorphism of X-STRs, the technical ease of genotyping by PCR-FLP and the possibility of multiplexing suitable short amplicons in forensic cases where degraded DNA is often studied. On the contrary, X-InDels theoretically have the same properties as X-STR even though the lower polymorphism implies a much higher number of InDel loci to study and, therefore, the multiplexing of many more loci than in the case of the X-STR. Gomes, I and collaborators (2020b) report, between 2012 and 2020, 4 out of 5 InDel multiplexes through CE and 1 through massive parallel sequencing. Although these multiplexes are not as common as X-STRs in routine forensic analysis, some of them can be used as a complementary tool in complex kinship analysis, as demonstrated by Gomes, C et al. (2020a), when

computing X-InDels power of discrimination for males and females, as well as the mean exclusion chance for duos and trios in the global Spanish population study.

Finally, SNPs lag far behind both X-STRs and InDels, possibly due to the lack of standardization of the various methodologies used to type them (MALDI-TOF mass spectrometry, Massive Parallel Sequencing, Taqman probes, Snapshot). The compilation of X-chromosome-specific SNP genotyping systems by Gomes, I and collaborators (2020b) show a shortage of X-SNP. Besides, the information content of the X-SNPs does not amount to that of the X-STRs, and, subsequently, many more SNPs should be included in the multiplexes (Amorim and Pereira 2005). This adds complexity to the analysis and interpretation of the results.

Worldwide X-Chromosome Population Data (Population Databases)

Compiled data of DNA polymorphisms is a common practice in forensic and population genetics. Databases of the Y chromosome or autosomal STRs are periodically updated in several institutions in order to provide a powerful tool to the forensic research. This tool has been also of great interest for populational genetics. Gomes, I et al. (2020b) have extensively reviewed the literature searching for articles with X polymorphisms databases. As they report, articles show a great variability concerning the number of polymorphisms studied (mainly X-STRs) and the number and classification of populations. The most common polymorphisms considered in these studies are X-STRs, although InDels multiplexes have been also analyzed in several populations (Pereira et al. 2012; Gomes et al. 2020a). Also, in forensic and even in populational studies, DNA polymorphisms become popular when they are included in commercial kits. The first commercial kit for forensic applications was the Argus X-UL (Biotype, Dresden, Germany), containing four X-STRs (DXS8378, DXS7132, HPRTB, and DXS7423) situated in distant positions of the chromosome to avoid any linkage. Later on, the same commercial firm released Argus X-8 (DXS7132, DXS7423, DXS8378, DXS10074, DXS10101, DXS10134, DXS10135, and HPRTB). This kit considered four pairs of linked X-STRs. Then, the Argus X-12 (Qiagen, Hilden, Germany) expanded the former 8 X-STRs until 12 (DXS8378, DS10135, DXS10148, DXS7132, DXS10074, DXS10079, HPRTB, DXS10101, DXS10103, DXS7423, DXS10134, DXS10146) organized in 4 linkage groups. This last kit was optimized to release the final version Argus X-12 QS, which contained the same group of X-STRs, including now a "Quality Sensor."

Other commercial kits, such as Goldeneye DNA ID System 17X (Goldeneye Technology Co., Ltd., Beijing, China) and the AGCU X19 STR kit (Wuxi Sino-German Meilian Biotechnology Co., Jiangsu, China), were mainly used for testing Chinese populations. However, a decaplex design by the Portuguese and Spanish Speaking Group of the International Society for Forensic Genetics (GHEP-ISFG) is the most widely used system for X-STR genotyping (Gusmão et al. 2008). All these

X-STR genotyping systems consider the four LD groups described above. Therefore, allele and haplotype frequencies are often, but not always unfortunately, included together both for populational and forensic purposes.

Concerning the global population representation of the X-chromosome in the databases, an extensive study by Gomes, I and collaborators (2020b) has shown, apart from a lack of X-STR data information for many countries, a high heterogeneity among and inside continents. There is a lack of data for sub-Saharan African and American populations (except for Argentina, Brazil, and United States), while China is by far the best represented both by the number of publications and by ethnic groups.

The Statistical Approach: IBD, IBS, and Likelihood Ratio

When an alleged kinship between two individuals is tested in the absence of other family members, it is not easy to determine their kinship degree, since there is a possibility that related people do not share any allele, as well as two unrelated people may coincidentally have similar alleles (Wenk and Chiafari 2000). For example, two full brothers may not share any allele, both for autosomal, as for X-chromosome marker. On the other hand, one of the clues for kinship analysis is the frequencies of each allele involved in the genetic analysis, hence the enormous importance of correctly determining the reference population.

As explained by Gomes, I and collaborators (2020b), considering several generations beyond which individuals are assumed to be unrelated, genetic relatedness studies are based on the probability that allele sets are inherited from common ancestral alleles, that is, on the probability that the alleles are identical-by-descendants (IBD) (Jacquard 1974). In this way, two individuals are said to be related, if (at least) one of one individuals' allele can be IBD relative to (at least) one allele of the other (see Example 1). Finally, an individual is said to be inbred if he/she can have two IBD alleles in the same locus. This is the case with the son of two brothers (or any two individuals who are related to each other). In this specific case, both alleles have the same origin.

The alleles of unrelated individuals can be the "same," being designated by identical-by-state (IBS) alleles (Jacquard 1974; Pinto et al. 2012), without having the same origin or ancestry. For example, for an autosomal STR, two unrelated individuals are 13–18. IBS alleles may not be IBD if their coalescence is much earlier than that relationship or appear independently, by mutation (Blouin 2003); however, assuming no mutation, all IBD alleles are also IBS.

In Example 1, a case is shown where two alleles are IBD (A3 from the putative mother) and a case where two alleles being IBS; it cannot be guaranteed whether they are IBD or not (A3 alleles of the progeny). In Fig. 1 it is explained the schematic figures used on each Jacquard case (Jacquard 1974), representing always how two individuals can be related by sharing 2, 1, or 0 IBD alleles.

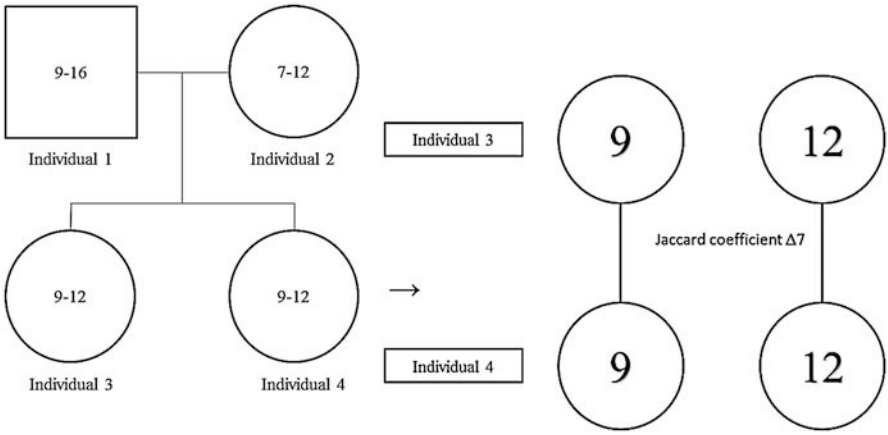
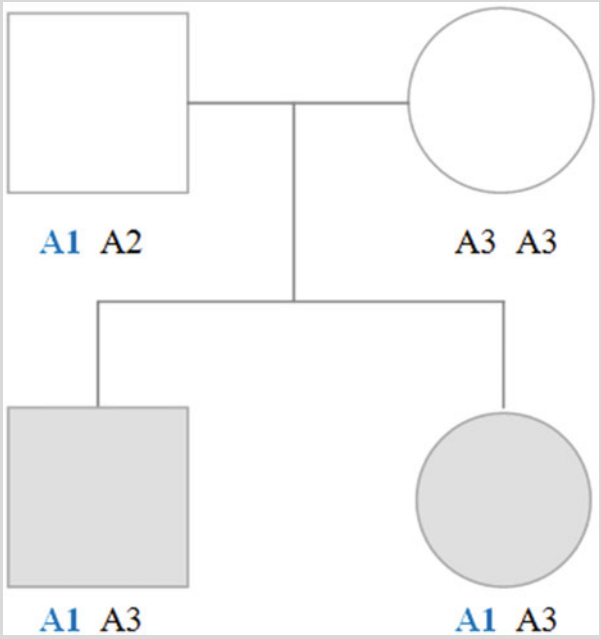


Fig. 1 Schematic explanation of how two persons can share two IBD alleles, represented by a Jacquard coefficient (Δ7)

Example 1



(continued)

If a pair of alleles of the parents is A1A2 (father) and A3A3 (mother), then the A1 of which descendant are IBD, since they are copies of the same parental allele. The same is not true for the A3 alleles of the offspring, which may or may not be copies of the same maternal allele; however, although they cannot be guaranteed to be IBD, these two alleles are IBS (Weir et al. 2006).

On Fig. 2 are the three cases where each individual is not the result of a consanguineous relationship, also known as non-inbred individual, represented by the Jacquard coefficient Δ7, Δ8, and Δ9.

As explained before, the characterization of the kinship between two individuals is based on the probability their alleles are IBD (Weir et al. 2006; Gomes et al. 2020b). Indeed, all statistical evaluation depends mainly on the IBD probabilities and the frequency of the alleles in the population. It is possible to evaluate the individual's genotypic information to compare the possibility that they are related by one or another relationship, through the calculation of likelihood ratios (LR).

When calculating LR, the probabilities of genotypes are compared under different alternative hypotheses (Weir et al. 2006). For example, in paternity tests, the probability of the individuals' genotypes compatible with paternity (P | H1) is usually compared with the probability of the observed genotypes belonging to two unrelated individuals in the population (P|H2), calculating LR = (P|H1) / (P|H2).

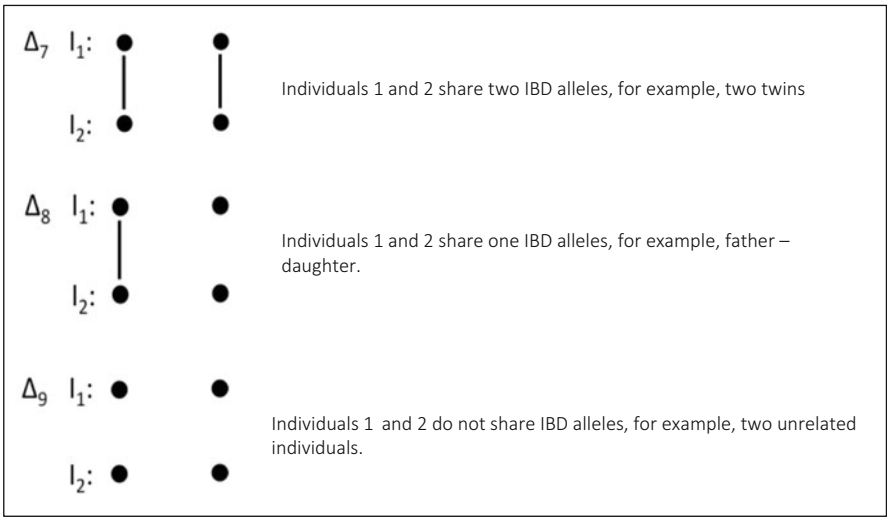


Fig. 2 Jacquard coefficient Δ7, Δ8, and Δ9, for two autosomal markers or two X-chromosome markers, but in that specific case, only considering two women

X-chromosome and Kinship Analysis

The difference between X-STRs and autosomal STRs is based on the fact that, in the first case, the sex of individuals determines the mode of transmission of genetic information (Szibor et al. 2003, Table 1). Thus, for pairs of individuals, the male-male (MM), female-female (FF), and female-male (FM) pairs must be considered separately. Depending on the case, the number of chances of IBD allele sharing varies (Pinto et al. 2010b), as each man has only one X-chromosome and each woman two. The number of hypotheses (represented, for X-chromosome, by each Jacquard coefficient with a capital lambda, Λ) also varies according to whether the woman is considered to be inbred, as explained by Pinto and collaborators (2012). Depending on the gender of the individuals, the patterns of sharing (partitions) of IBD alleles will be different.

Female-Female Duo (X^{FF})

For two women, as with autosomal markers, nine hypotheses for IBD partitions are considered. However, if only non-inbred women are considered, there are only three partitions (see Fig. 3), x_2^{FF}, x_1^{FF}, and x_0^{FF}, which represent, respectively, the probability of sharing 2, 1, or 0 IBD alleles. Note that, except for identical twins and similar to what happens for autosomal, there is no relationship between two women whose likelihood of sharing two IBD alleles is 1. However, there are genealogies that ($x_1^{FF} = 1$; $x_0^{FF} = x_2^{FF} = 0$), which means that two women so related share exactly one IBD allele at each locus. Finally, it should be noted that in the case of two unrelated women (and by definition), there is no IBD allele sharing, that is, the probability of sharing zero alleles is equal to 1 ($x_0^{FF} = 1$; $x_1^{FF} = x_2^{FF} = 0$). However, that there are genealogies that relate two women, for whom $x_0^{FF} = 1$, as is the case of paternal great-grandmother-great-granddaughter, for example.

Table 1 Specific relationships and the probability of transmitting genetic information (Excluding Mutation events). (Source: Butler 2011, Chapter 15)

Inheritance	Autosomal markers (%)	ChrY markers	mtDNA (%)	ChrX markers (%)
Mother → son	50	Not applicable	100	100
Mother → daughter	50	Not applicable	100	50
Father → son	50	100%	0	0
Father → daughter	50	0%	0	100
Paternal grandmother → granddaughter	25	Not applicable	0	100
Maternal grandmother → granddaughter	25	Not applicable	100	25
Paternal grandfather → grandson	25	100%	0	0

Male-Male Duo (X^{MM})

Considering two men and considering that each one has only one X-chromosome, only two possibilities of sharing are observed: they either share IBD alleles or they do not share (x_1^{MM} and x_0^{MM}, respectively; see Fig. 3). Excluding incestuous situations, there is no genealogy for which $x_1^{MM} = 1$, and there are several where the IBD allele is not taken for granted (as for the unrelated). In fact, even for father-son comes $x_0^{MM} = 1$.

Female-Male Duo (X^{FM})

Considering a female-male duo, there will be, for a non-inbred woman, x_1^{FM} and x_0^{FM} representing, respectively, the likelihood that the duo will share 1 or 0 IBD allele on an X-chromosome marker (Fig. 3). It should be noted that in kinships such as father-daughter and mother-son, there is a mandatory sharing of an X-chromosome; therefore, the probability of sharing 1 IBD allele is equal to 1 ($x_1^{FM} = 1$; $x_0^{FM} = 0$); however, in relationships such as paternal grandmother-grandchild or paternal grandfather – there is no X-chromosome transmission, so the probability of sharing 0 IBD alleles is equal to 1 ($x_0^{FM} = 1$; $x_1^{FM} = 0$), as well as unrelated.

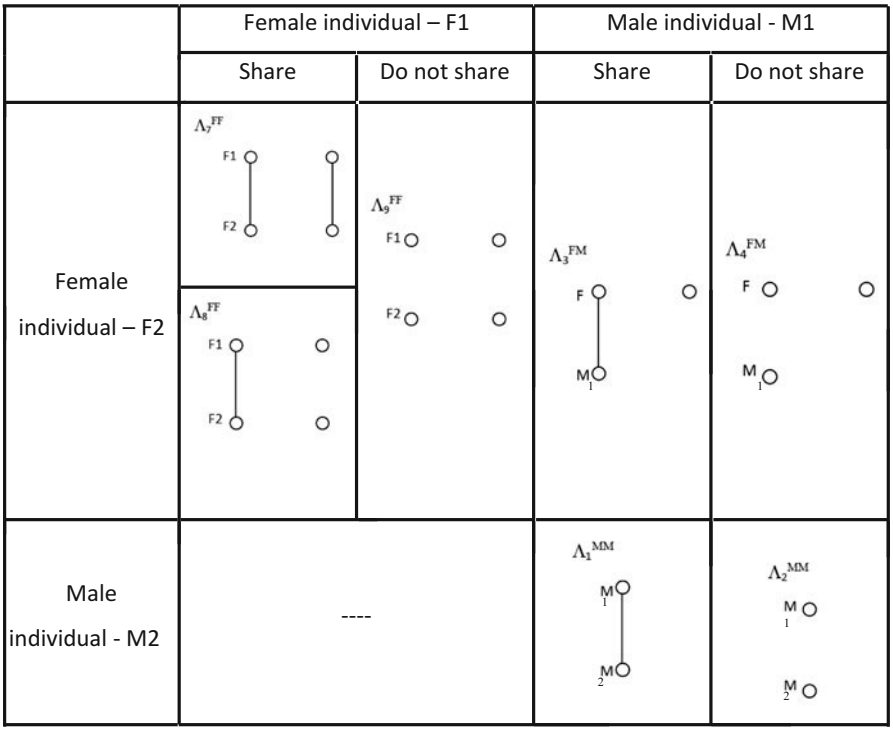


Fig. 3 Schematic summary of the IBD partitions for the X-chromosome, considering two women (F1 and F2), two men (M1 and M2), and a man/woman pair

As explained before, Jacquard's coefficients partitions Δ7, Δ8, and Δ9 represent the possibilities of sharing two, one, or none IBD alleles, respectively. When considering X-chromosome analysis, such hypothesis could also be represented by Λ_7 (the same as Δ7), Λ_8 (the same as Δ8), and Λ_9 (the same as Δ9), and their respective probabilities are represented by x_2, x_1, and x_0. The sum of the three probabilities must always be 1 ($x_2 + x_1 + x_0 = 1$).

Considering two twin sisters, x_2, x_1, and x_0 are very easy to determine, considering each X-chromosome locus analyzed. Since they must share the same information for each locus ($x_2 = 1$, obligation to share the same maternal and paternal information in each locus), so the probability of sharing any IBD allele (x_0) must always be 0 ($x_0 = 0$). In this sense, $x_2 = 1$, x_1, and x_0 are 0.

In the cases of parents-child, it is also easy to calculate these three probabilities, since between mother-son/daughter and father-daughter, it is mandatory they share exactly one IBD allele for each locus. So, considering these relations, $x_2 = 0$; $x_1 = 1$, and $x_0 = 0$.

This calculation can be complicated when it comes to calculating these probabilities in cases of kinship where there is no obligation to share IBD information at each locus. The most common kinship x_2, x_1, and x_0 probabilities can be consulted on Table S1 from Pinto and collaborators (2011).

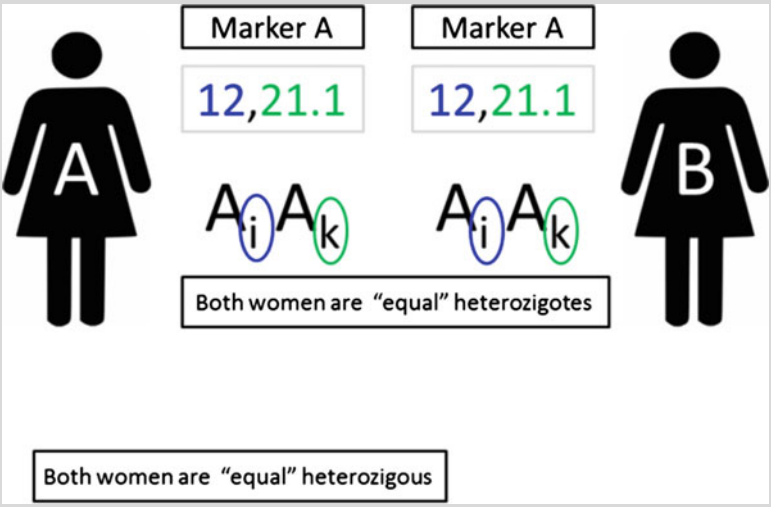
On the other hand, it is possible to calculate manually the probability for two non-twin women to be sisters. It is necessary to have in consideration three types of information:

(a) *The genotype of each woman for that specific X-STR.* This information is crucial to choose which joint genotypic probability must be used, since each formula takes into account both women genotypes. These formulas could be consulted on Pinto and collaborators (2012).
(b) *x_2, x_1, and x_0 values for that specific kinship.*
(c) *The populational frequency of each allele found on both women*. This value should be searched in a genetic database, specific for each considered population.

In Example 2, a probability of sisterhood is calculated, considering a random X-STR allele.

Example 2
Sisterhood probability considering a random X-STR ("Marker A"), two non-twin women (A and B), hypothetical values for the population frequency of each allele involved in this case, and the probabilities for each partition, x_2, x_1, and x_0.

(continued)



Considering their genotype for this specific marker, consulting the joint genotypic probabilities published by Pinto and collaborators (2011), it is possible to determine which formula should be applied to compute the sisterhood probability for this specific marker.

- $2x_2^{FF}f_if_k + x_1^{FF}f_if_k(f_i + f_k) + 4x_0^{FF}f_i^2f_k^2$, according to both women genotype for this specific marker.

- $x_2^{FF}=1/2$
- $x_1^{FF}=1/2$
- $x_0^{FF}=0$

These values are specific for this kinship (two non-twin sisters) and for the all kinships that belong to the same X-class.

- Hypothetical allele frequencies, in a specific population:
  - $f_{12}=0.01$
  - $f_{21.1}=0.05$

- $2x_2^{FF}f_if_k + x_1^{FF}f_if_k(f_i + f_k) + 4x_0^{FF}f_i^2f_k^2$

  $= 2 \times ½ \times 0.01 \times 0.05 + ½ \times 0.01 \times 0.05\ (0.01+0.05) + 4 \times 0 \times 0.01^2 \times 0.05^2$

  $=0{,}000515$. This value corresponds to the probability of sisterhood, for marker "A", considering two non inbred women, considering the frequency of these alleles (*i* and *k*) in a specific population.

(a) When the alleged father is a close relative of the biological father
(a1) The paternal grandfather-granddaughter relationship

In cases where it is intended to determine the probability of a presumed paternity, and where there may be a possibility that the presumptive father is a close relative of the real father, the informative power of the X-chromosome is substantially superior to that of autosomal markers (Szibor et al. 2003). Consider, for example, the situation in which two individuals, whose relationship is father-son, are analyzed as putative parents concerning a female child. In this case, only the real father will share IBD alleles with his daughter. Therefore, whenever there is no total sharing of the haplotype present in the father, which cannot be attributed to mutational events, paternity can be excluded.

Thus, when the alleged father is the paternal grandfather, the probability of compatibility with the alleged daughter (in fact, granddaughter) is equal to two random individuals from the population. Therefore, in most cases, the analysis of X-chromosome markers allows the exclusion of the paternal grandfather as a father.

(a2) The paternal uncle relationship – niece

In the event that two alleged parents (relative to a daughter) are siblings, the probability both brothers share X-chromosome IBD alleles by maternal side is 1/2, similar to what happens for autosomal markers (Szibor et al. 2003).

Thus, when analyzing a case of paternity where there is a suspicion that the presumed father will be the brother of the real father, the use of studying chromosome X markers is only advantageous since it allows the analysis of an additional number of markers, relative to markers autosomal tests usually analyzed.

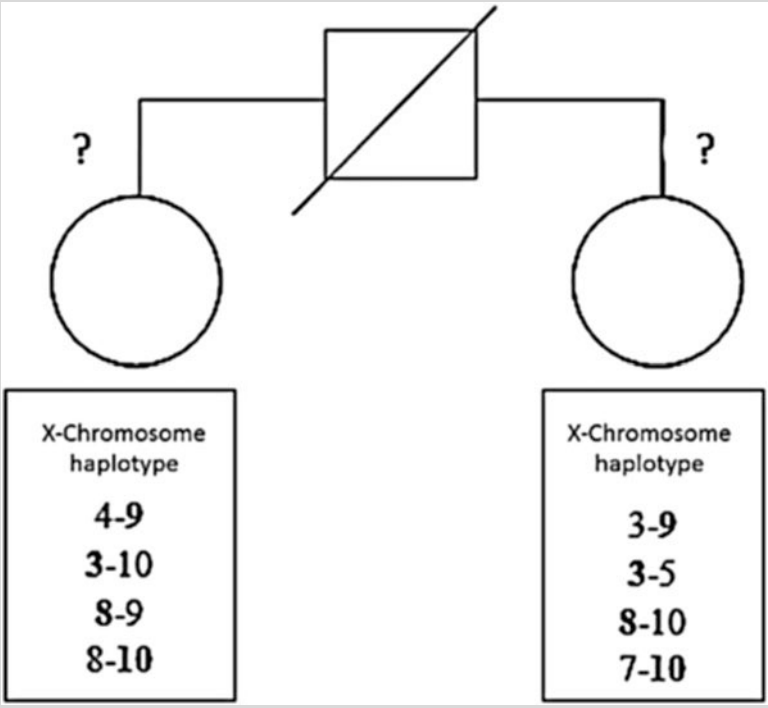
(b) When the biological father is absent – kinship relationships whose power of discrimination with autosomal markers is null
(b1) Half-sisters

The study of X-chromosome markers assuming a paternal half-sister relationship can unequivocally exclude any hypothesis of paternity, even when the alleged father is not accessible for analysis. In fact, due to their mode of transmission, two half-sisters must share the entire paternal haplotype (Szibor et al. 2003). However, when it is not possible to exclude the hypothesis of paternity, as with other markers, whenever possible it is important trying to access relevant information from other relatives, for example, mother or brother (see Example 3).

Example 3

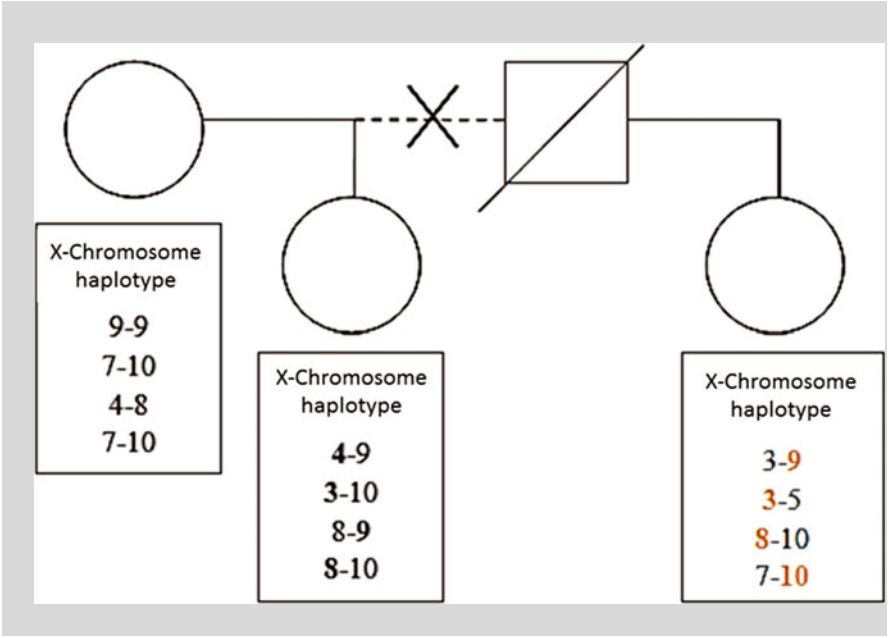
The usefulness of studying other biological relatives from both presumptive half-sisters.

In the absence of a presumed father, the analysis of two alleged paternal half-sisters may be sufficient to immediately refute the relationship under analysis (when they do not share any allele); however, there are cases, such as the one presented here, in which it is not possible to reject the kinship between both, being possible to define an alleged paternal haplotype (highlighted in the figure: 9-3-8-10, for the four represented markers.)



When accessing additional information, in this case, information about the haplotypes of the mother of one of the presumed half-sisters, it appears that the hypothesis of paternity may be excluded. In the figure, the paternal haplotypes of each of the alleged sisters are highlighted, verifying the impossibility of this relationship being true, since they do not share the same paternal haplotype.

(continued)



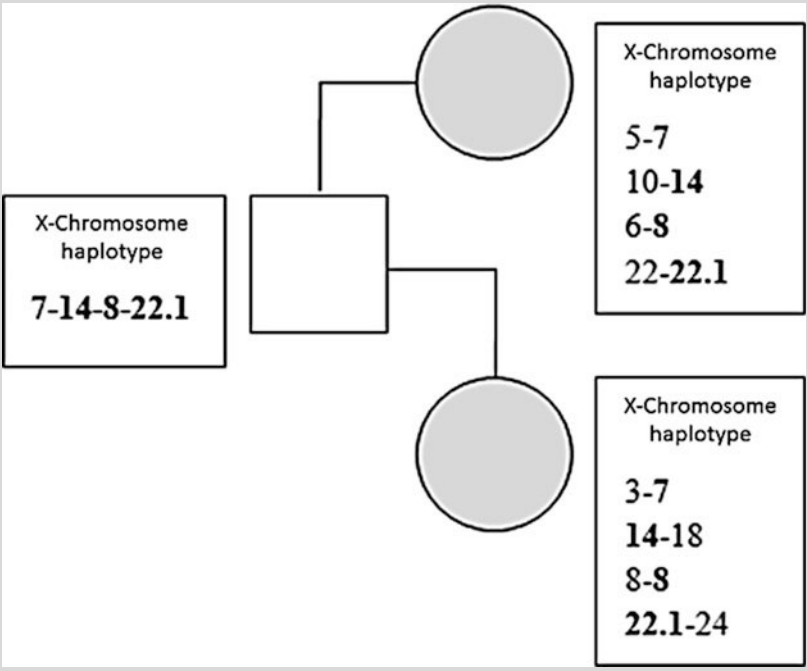
(b2) The paternal grandmother-granddaughter relationship

In cases where it is not possible to access samples from the alleged father due to his absence, it may be necessary to use samples from biological relatives close to him, such as the mother or a possible sister. It such cases, the analysis of autosomal markers may prove to be useless, since the exclusion power is null. There is no way to exclude the hypothesis that one woman is the grandmother of another, based solely on genetic information related to autosomal markers. However, concerning X-STRs, the alleged father's mother assumes a high importance, since she has all the alleles of the true father has (see Example 4). In this case, the exclusion power will not be null since the grandmother's genetic information must always be present in the granddaughter.

However, although this analysis is extremely useful, it should not be confused with a true paternity test, as for any other type of genetic marker. One would have to verify, with complete certainty, two premises: the presumable father is the son of the alleged grandmother, and it should be taken into account the possibility of the alleged grandmother having another son likely to be the real father.

Example 4
Paternal grandmother-granddaughter.

(continued)



Given its transmission, one of the paternal grandmother's haplotypes (highlighted in the figure) is passed on to her son and then fully transmitted to all her biological daughters.

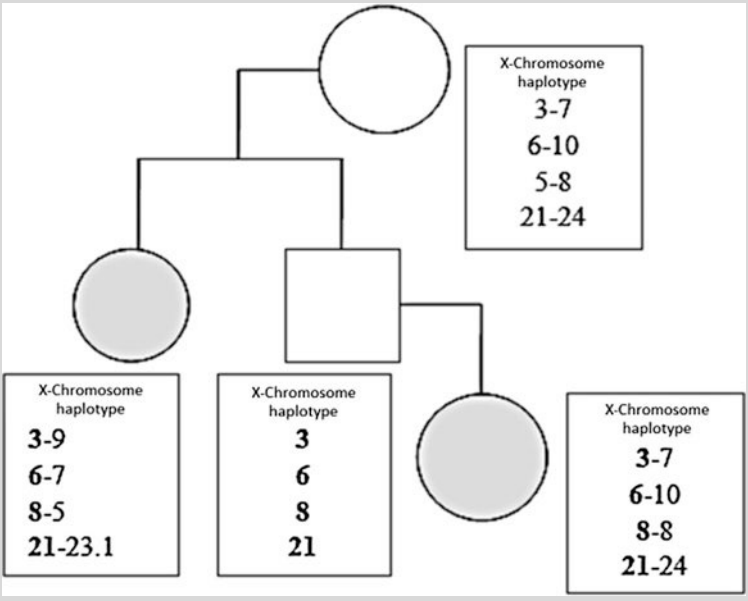
(b3) The paternal aunt-niece

In cases where it is not possible to have access to a biological sample of the alleged father, it may be useful to analyze the paternal aunt-niece relationship, since there is always the possibility of brother-sister sharing the same X-chromosome, transmitted by the mother. However, contrary to the hypothesis "paternal grandmother-granddaughter," the analysis of the X-chromosome does not allow to exclude any female individual from the paternal aunt-niece relationship, even when none of the alleles is shared. That is, the power of discrimination of X-chromosome markers is equal to that of autosomal markers, null.

Thus, in a case of a paternity dispute, it may not be as effective to use the analysis of a presumed paternal aunt, as the mother of the presumed father. However, if only information regarding the paternal aunt is available, the X-chromosome markers are more effective because the probability of sharing alleles from the same ancestor is twice as high as that obtained when using autosomal markers (see Example 5).

Again, this analysis does not allow extrapolating the result to a true paternity test. For this analysis to draw true conclusions, the alleged paternal aunt must be, in fact, the biological sister of the alleged father, as well as, it is also necessary to ensure that the alleged aunt does not have any other male brother possible to be the real father.

Example 5
Paternal aunt-niece.



In the present case, brothers share information regarding the X-chromosome, so the analysis of the paternal aunt-niece may be informative.

Indistinguishable Genealogies

Although they are not very frequent, there are some situations where it is necessary to evaluate alternative kinships for which the analysis of autosomal independent markers is irrelevant, since the probability associated with the different hypotheses of kinship has the same value (LR = 1). Such is the case of half-brothers, avuncular, and grandfather-grandchild, whose probabilities of sharing IBD alleles are the same, belonging to the same autosomal class. Thus, genealogies with the same IBD probabilities are not distinguishable through the analysis of autosomal independent markers (Pinto et al. 2009).

Classes Distinguishable by Analyzing X-Chromosome Markers

Theoretical studies carried out by Pinto et al. (2010a, b) demonstrate that some genealogies belonging to the same autosomal kinship class can, in theory, be distinguished by analysis of the X-chromosome. This is the case of the maternal grandmother-granddaughter and maternal aunt-niece relationships, for example. Figure 4 shows some examples of genealogies indistinguishable by autosomes, for which a distinction is expected when using the X-chromosome markers.

Kinship relationships are theoretically distinguishable because, concerning the X-chromosome, they belong to different kinship classes, having different allelic sharing probabilities. Thus, for an FF duo, three distinct kinship classes were determined involving two women related in the second degree (grandmother-granddaughter, aunt-niece, and half-sisters), each with probabilities of allelic sharing by different offspring. Regarding a MM duo, there are three different kinship classes (where individuals are related in the 2nd-degree). In the class where there is no allelic sharing, all kinships have men related by paternal side. Finally, considering an FM duo, four kinship classes were described where 2nd-degree kinships are grouped differently. All X-class can be accessed on Table S1 from Pinto and collaborators (2011) supplementary material.

Some of the cases that raise doubts in terms of alternative relationships can be resolved when there is information beyond genetics, with age being a factor that allows excluding (or considering differently) some of the relationships (Pinto et al. 2009), namely, grandparents-grandchild. However, in cases where only biological traces are accessible, such as in the case of an accident, war, or a massive disaster, sometimes it is not possible to exclude any kinship.

Incest Cases

In the case of a daughter, X-chromosome markers may provide crucial information about a possible incest case even without analyzing the putative father (Butler 2011; Gomes et al. 2020b), as explained below.

8.1 *The father of the daughter is also the father of the mother* (Butler 2011): in this case, mother and daughter will share the same X paternal haplotype, since they both have the same father. So, the child can be homozygous, if the mother has transmitted her paternal haplotype; so, the child receives the same information twice. If the mother transmits her maternal X-chromosome haplotype, the child will be heterozygous. In this case, X-chromosome profiles from both child and mother will be equal.

8.2 *The father of the daughter is also the brother of the mother* (Butler 2011): in this case, the heart of the matter is whether mother and father (who are also brothers) have the maternal haplotype in common. If they do not share the same maternal information for the analyzed X-markers, the genetic profile of the child will not be

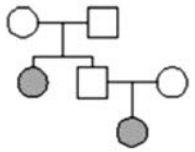
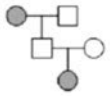
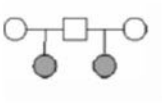
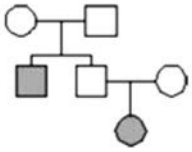
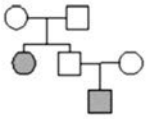
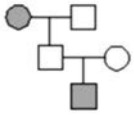
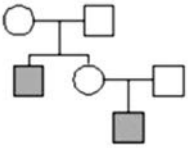
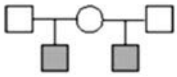
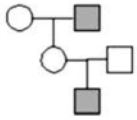
Two non – inbred women			
Kinship	X-class	Kinship	X-class
	$x_2^{FF} = 0$ $x_1^{FF} = ½$ $x_0^{FF} = ½$	 Paternal grandmother-granddaughter 	$x_2^{FF} = 0$ $x_1^{FF} = 1$ $x_0^{FF} = 0$
Paternal aunt-niece		Paternal half-sisters	
Non-inbred man and woman			
	$x_1^{FM} = ½$ $x_0^{FM} = ½$	 Paternal aunt - nephew 	$x_1^{FM} = 0$ $x_0^{FM} = 1$
Paternal uncle-niece		Paternal grandmother-grandson	
Two non-inbred men			
	$x_1^{MM} = ¼$ $x_0^{MM} = ¾$	 Maternal half-brothers 	$x_1^{MM} = ½$ $x_0^{MM} = ½$
Maternal uncle-nephew		Maternal grandfather-grandson	

Fig. 4 (continued)

distinguished from a non-inbred child. In such cases, analyzing X-chromosome information will be so informative as to increase the number of autosomal markers.

X-Chromosome and Forensic Analysis of Low Template DNA

The study performed by Gomes, C and collaborators (2020a) with X-InDels in the Spanish population has demonstrated the potential of this tool in identification, as in kinship cases. After computing allele and haplotype frequencies, forensic efficiency parameters were calculated – Power of Discrimination (PD) $_{\text{males}} = 99.999976\%$; $PD_{\text{females}} = 99.99999999998\%$; Mean Exclusion Chance (MEC) $_{\text{duos}} = 0.999$ and $MEC_{\text{trios}} = 0.99999$. Given the X-InDels characteristics, mainly its feasibility to be analyzed as short amplicons, it allows its use in degraded samples, as demonstrated by Gomes and collaborators (2019) in their study with very degraded samples from the Spanish Civil War (1936–1939). In this case, when analyzing autosomal STRs between a presumptive father-daughter duo, a dubious LR result was obtained due to an only one Mendelian incompatibility. When using X-InDels it was possible to immediately discard the paternity hypothesis, due to the five incompatibilities detected between both X-InDels profiles.

Conclusion

Considering its unique structure and mode of transmission, several studies have shown the X-chromosome as one of the most effective tools, as both a complement to the use of autosomal markers and an alternative in several cases of complex kinships, especially X- STRs markers. The effectiveness of other X-chromosome markers, namely, X-InDels, has also been demonstrated in cases where biological samples are highly degraded.

References

Amorim A, Pereira L (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. Forensic Sci Int 150:17–21. https://doi.org/10.1016/j.forsciint.2004.06.018

Bachtrog D (2006) A dynamic view of sex chromosome evolution. Curr Opin Genet Dev 16: 578–585

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. Nature 446: 507–512

Fig. 4 Example of distinguishable kinship classes by analyzing X-STRs. For each type of duo considered (FF, MM, and FM), the two classes presented belong to the same autosomal class, so the kinships are indistinguishable. Using X-STRs analysis, these genealogies are found in different kinship classes, so they are theoretically distinguishable

Blouin M (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends Ecol Evol 18(10):503

Bull JJ (1983) Evolution of sex determining mechanisms. Benjamin Cummings, Menlo Park

Butler J (2011) Advanced topics in forensic DNA typing: methodology, 1st edn. Academic. ISBN: 9780123745132

Connallon T, Knowles LL (2005) Intergenomic conflict revealed by patterns of sex-biased gene expression. Trends Genet 21:495–499

Diegoli TM (2015) Forensic typing of short tandem repeat markers on the X and Y chromosomes. Forensic Sci Int Genet 18:140–151. https://doi.org/10.1016/j.fsigen.2015.03.013

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746–756

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253. https://doi.org/10.1016/0888-7543(92)90371-x

Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Chen CJ, Kaplan N, Chang HY, Heard E, Dekker J (2016) Structural organization of the inactive X chromosome in the mouse. Nature 535 (7613):575–579. https://doi.org/10.1038/nature18589. Epub 2016 July 18. PMID: 27437574; PMCID: PMC5443622

Gomes C (2011) Aplicações forenses do estudo de 12 X-STRs: utilidade em diferentes casos de investigação de parentesco biológico. MSc thesis, Faculty of Science, University of Porto

Gomes C, Palomo-Díez S, Baeza-Richer C, López-Parra AM, Cuscó I, Garcia-Arumí E et al (2019) X-InDels efficacy evaluation in a critical samples paternity case: a Spanish Civil War case from the memorial of the Camposines (Tarragona, Spain). Forensic Sci Int Genet Suppl Ser 7(1): 494–495

Gomes C, Quintero-Brito JD, Martínez-Gómez J, Pereira R, Baeza-Richer C, Gay MA et al (2020a) Spanish allele and haplotype database for 32 X-chromosome Insertion-Deletion polymorphisms. Forensic Sci Int Genet 46:102262

Gomes I, Pinto N, Antão-Sousa S, Gomes V, Gusmão L, Amorim A (2020b) Twenty years later: a comprehensive review of the X chromosome use in forensic genetics. Front Genet 11:926. https://doi.org/10.3389/fgene.2020.00926

Graves JA (1995) The origin and function of the mammalian Y chromosome and Y-borne genes–an evolving understanding. BioEssays 17:311–320. https://doi.org/10.1002/bies.950170407

Gribnau J, Grootegoed A (2012) Origin and evolution of X chromosome inactivation. Curr Opin Cell Biol 24(3):397–404

Gunter C (2005) Genome biology: she moves in mysterious ways. Nature 434:279–280. https://doi.org/10.1038/434279a

Gurbich TA, Bachtrog D (2008) Gene content evolution on the X chromosome. Curr Opin Genet Dev 18:493–498. https://doi.org/10.1016/j.gde.2008.09.006

Gusmão L, Sánchez-Diz P, Alves C, Gomes I, Zarrabeitia MT, Abovich M et al (2008) A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. Int J Legal Med 123:227–234

Hearne CM, Todd JA (1991) Tetranucleotide repeat polymorphism at the HPRT locus. Nucleic Acids Res 19:5450. https://doi.org/10.1093/nar/19.19.5450

Hurst LD (2001) Evolutionary genomics. Sex and the X. Nature 411:149–150

Jacquard A (1974) The genetic structure of populations, biomathematics, vol 5. Springer, Berlin, pp 102–107

Kohn M, Kehrer-Sawatzki H, Vogel W, Graves JA, Hameister H (2004) Wide genome comparisons reveal the origins of the human X chromosome. Trends Genet 20:598–603. https://doi.org/10.1016/j.tig.2004.09.008

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964–967. https://doi.org/10.1126/science.286.5441.964

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. https://doi.org/10.1038/35057062

Ohno S (1967) Sex chromosomes and sex-linked genes. Springer, Berlin, Germany
Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B (2003) Paucity of genes on the Drosophila X chromosome showing male-biased expression. Science 299:697
Pereira R, Pereira V, Gomes I, Tomas C, Morling N, Amorim A et al (2012) A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. Int J Legal Med 126:97–105. https://doi.org/10.1007/s00414-011-0593-2
Pinto N, Gusmão L, Amorim A (2009) Distinguishing kinship from genealogical likelihoods. Forensic Sci Int 2:453–454
Pinto N, Gusmão L, Amorim A (2010a) Likelihood ratios in kinship analysis: contrasting kinship classes, not genealogies. Forensic Sci Int 4:218–219
Pinto N, Gusmão L, Amorim A (2010b) X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial. Forensic Sci Int 5: 27–32
Pinto N, Gusmão L, Amorim A (2011) X-chromosome markers in kinship testing: a generalisation of the IBD approach identifying situations where their contribution is crucial. Forensic Sci Int Genet 5(1):27–32
Pinto N, Silva P, Amorim A (2012) A general method to assess the utility of the X-chromosomal markers in kinship testing. Forensic Sci Int Genet 6(2):198–207
Rice WR (1984) Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735–742
Rice WR (1996) Evolution of the Y sex chromosome in animals. Bioscience 46:331–343
Rinn JL, Snyder M (2005) Sexual dimorphism in mammalian gene expression. Trends Genet 21: 298–305
Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D et al (2005) The DNA sequence of the human X chromosome. Nature 434:325–337. https://doi.org/10.1038/nature03440
Schaffner SF (2004) The X chromosome in population genetics. Nat Rev Genet 5:43–51. https://doi.org/10.1038/nrg1247
Sleddens HF, Oostra BA, Brinkmann AO, Trapman J (1992) Trinucleotide repeat polymorphism in the androgen receptor gene (AR). Nucleic Acids Res 20:1427. https://doi.org/10.1093/nar/20.6.1427-a
Sutton E, Hughes J, White S, Sekido R, Tan J, Arboleda V, Rogers N, Knower K, Rowley L, Eyre H et al (2011) Identification of SOX3 as an XX male sex reversal gene in mice and humans. J Clin Investig 121:328–341
Szibor R (2007) X-chromosomal markers: past, present and future. Forensic Sci Int Genet 1:93–99. https://doi.org/10.1016/j.fsigen.2007.03.003
Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74
Szibor R, Hering S, Kuhlisch E, Plate I, Demberger S, Krawczak M, Edelmann J (2005) Haplotyping of STR cluster DXS6801–DXS6809–DXS6789 on Xq21 provides a powerful tool for kinship testing. Int J Legal Med 119:363–369
Vicoso B, Charlesworth B (2006) Evolution on the X chromosome: unusual patterns and processes. Nat Rev Genet 7:645–653
Wang PJ, McCarrey JR, Yang F, Page DC (2001) An abundance of X-linked genes expressed in spermatogonia. Nat Genet 27:422–426
Weir B, Anderson A, Hepler A (2006) Genetic relatedness analysis: modern data and new challenges. Nat Rev 7:771–781
Wenk R, Chiafari F (2000) Distinguishing full siblings from half-siblings in limited pedigrees. Transfusion 40:44–47

Using Mitochondrial DNA in Human Identification

20

Pankaj Shrivastava, Manisha Rana, Pushpesh Kushwaha, and D. S. Negi

Contents

P. Shrivastava (✉)
DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

M. Rana
Department of Forensic Science, Guru Ghasidas University, Bilaspur, India

DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

P. Kushwaha
Dr. APJ Abdul Kalam Institute of Forensic Science and Criminology, Bundelkhand University, Jhansi, India

DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India

D. S. Negi
Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_18

Abstract

Mitochondrial DNA (mtDNA) has been employed for human identification in forensic caseworks since 1996. High copy number and uniparental inheritance are some of the advantageous characteristics of mitochondrial DNA (mtDNA) that yields useful results in degraded samples that fail to generate successful nuclear DNA (nuDNA) profiles. Presently, analysis of autosomal short tandem repeats (STR) is utilized for genetic fingerprinting in forensic caseworks. However, majority of the samples from skeletal remains of deceased persons encountered in caseworks such as mass disaster exposed to high temperature and harsh environment conditions are highly degraded. The mitochondrial DNA (mtDNA) profiling in addition to nuclear DNA in disaster victim and archaeological ancient samples establish the human identification. The mitochondrial DNA being a maternally inherited marker plays a crucial role in establishing phylogenic ancestry. Also, the mitochondrion being powerhouse of the cells plays a crucial role in human ageing and is an excellent marker for age estimation. In this chapter, the structure and function of the mitochondrial genome, its inheritance pattern and transmission, application in forensic caseworks, data interpretation in mtDNA casework and population data basing, and future perspectives have been discussed.

Keywords

mtDNA · Heteroplasmy · Maternal inheritance · Hypervariable regions · CRS · rCRS

Mitochondria: Structure, Physiology, and Functions

Mitochondria are ovoid or sausage shaped double-membranous cytoplasmic cell organelle ranging 0.5–1.0 μM in diameter, operative towards the function of energy-transduction, imperative for various cellular metabolism. Its double-membranous wall comprises of an outer semipermeable layer that isolates the organelle from cytosol and an inner impermeable layer. The inner layer can be further bifurcated into an outer side facing and running parallel to outer membrane and an inner side which is invaginated into several folds called cristae that outlines the matrix (M-side) of mitochondria. Both the layers are separated by means of intermembrane space of

width about 60–80Å, also known as the C-side. At molecular level, both the layers have tail-to-tail arrangements of phospholipids bilayer along with embedded protein molecules. This bilayer assembly exists in liquid-crystalline state with one fraction of phospholipid molecule of hydrophobic (water-repellent) nature and another fraction being hydrophilic (water-attracting) nature (Demirel and Gerbaud 2019). Inner side of cristae enclosing the matrix is known to be the house of various enzymatic reaction and is embedded with majority of the electron transport chain (ETC) complexes, enzymatic components of citric acid cycle, and ATP synthase dimmers (McCarron et al. 2013; Glancy et al. 2020). Mitochondrial genome along with essential components required for its replication, transcription, and translation are present within the matrix region (Shoubridge 2004).

Functionally, mitochondria are involved in myriad of regulatory mechanisms that differs from organ-to-organ. Primarily the high pH of matrix ranging around 8 provides a transmembrane electrochemical gradient that helps in production of cellular energy in the form of ATP (adenosine triphosphate) (Kühlbrandt 2015). The crista lumens are the sites of electron carriers protein cytochrome c and ROS (reactive oxygen species) production, which are essential components of electron transport chain and are also involved in cell apoptosis (Plecitá-Hlavatá and Ježek 2016; Pearce et al. 2017). Mitochondria are also engaged in protein import and calcium influx for several mechanisms, such as tricarboxylic acid cycle, urea cycle, oxidative phosphorylation, and biosynthesis of heme, iron-sulfur cluster, cholesterol, glucose, etc. (Shoubridge 2004; Pearce et al. 2017; Spinelli and Haigis 2018; Demirel and Gerbaud 2019).

Mitochondrial Genome

Mitochondria possess a form of autonomously replicating genome originated from (eu)bacterial domain of life. The mitochondrial genomes (mt-genome) are structurally distinct from that of nuclear genome and are present in double-stranded circular conformity. Human mitochondrial DNA (mtDNA) is the covalently bonded double-stranded circular structure of 16,569 bp in length. Number of copies of mtDNA per cell ranges from 100 to 1000, depending on the cell-specific functions. Human mitochondrial proteome comprises of ~1500 proteins, majority of which are encoded by nuclear DNA while mt-genome expresses a subset of 13 polypeptides component of respiration chain, translated from 11 mRNAs along with 22 tRNAs, and 2rRNAs (Gray et al. 2001; Pearce et al. 2017) (Table 1). Based upon the G-C content gradient, the coding regions of mitochondria are classified as G nucleotide-rich heavy strand (H-strand) and C nucleotide-rich light strand (L-strand). Transcriptional promoters of these strands are located in the noncoding region of mtDNA, known as the displacement loop (D-loop) (Anderson et al. 1981; Taanman 1999). Graphical representation of mitochondrial genome is demonstrated in Fig. 1. D-loop or the control region is a metabolically active region of length 1.1 kb and is unsusceptible to the formation of polypeptide chain. D-loop consists of highly polymorphic

Table 1 Information of mt-genome, its encoding regions for 13 proteins, 2 rRNAs, and 22 tRNAs

Nucleotide position	mt Genome	Abbreviation	Description	Size of region	Strand transcribed
1–576, 16024–16569	Noncoding region	D-loop	Control region	1122	–
16024–16365	Hypervariable region	HV-1	Hypervariable region 1	341	–
73–340	Hypervariable region	HV-2	Hypervariable region 2	267	–
438–576	Hypervariable region	HV-3	Hypervariable region 3	138	–
577–647	tRNA	F	tRNA phenylalanine	71	H-strand
648–1601	rRNA	12S	12S rRNA	954	H-strand
1602–1670	tRNA	V	tRNA valine	69	H-strand
1671–3229	rRNA	16S	16S rRNA	1559	H-strand
3230–3304	tRNA	L1	tRNA leucine1	75	H-strand
3305–4263	NADH dehydrogenase subunits	ND1	NADH dehydrogenase 1	958	H-strand
4263–4331	tRNA	I	tRNA isoleucine	68	H-strand
4329–4400	tRNA	Q	tRNA glutamine	72	L-strand
4402–4469	tRNA	M	tRNA methionine	68	H-strand
4470–5511	NADH dehydrogenase subunits	ND2	NADH dehydrogenase 2	1042	H-strand
5512–5579	tRNA	W	tRNA tryptophan	68	H-strand
5587–5655	tRNA	A	tRNA alanine	69	L-strand
5657–5729	tRNA	N	tRNA asparagine	73	L-strand
5761–5826	tRNA	C	tRNA cysteine	66	L-strand
5826–5891	tRNA	Y	tRNA tyrosine	66	L-strand
5901–7445	Cytochrome oxidase subunits	COX I	Cytochrome c oxidase I	1545	H-strand
7445–7516	tRNA	S1	tRNA serine 1	72	L-strand
7518–7585	tRNA	D	tRNA aspartic acid	68	H-strand
7586–8294	Cytochrome oxidase subunits	COX II	Cytochrome c oxidase II	709	H-strand
8295–8364	tRNA	K	tRNA lysine	70	H-strand
8365–8572	ATP synthase subunits	ATP8	ATP synthase 8	208	H-strand
8527–9207	ATP synthase subunits	ATP6	ATP synthase 6	681	H-strand
9207–9990		COX III		784	H-strand

(continued)

Table 1 (continued)

Nucleotide position	mt Genome	Abbreviation	Description	Size of region	Strand transcribed
	Cytochrome oxidase subunits		Cytochrome c oxidase III		
9991–10058	tRNA	G	tRNA glycine	68	H-strand
10059–10404	NADH dehydrogenase subunits	ND3	NADH dehydrogenase 3	346	H-strand
10405–10469	tRNA	R	tRNA arginine	65	H-strand
10470–10766	NADH dehydrogenase subunits	ND4L	NADH dehydrogenase 4L	297	H-strand
10760–12137	NADH dehydrogenase subunits	ND4L	NADH dehydrogenase 4	1378	H-strand
12138–12206	tRNA	H	tRNA histidine	69	H-strand
12207–12265	tRNA	S2	tRNA serine 2	59	H-strand
12266–12336	tRNA	L2	tRNA leucine 2	71	H-strand
12337–14148	NADH dehydrogenase subunits	ND5	NADH dehydrogenase 5	1812	H-strand
14149–14673	NADH dehydrogenase subunits	ND6	NADH dehydrogenase 6	525	L-strand
14747–15887	Cytochrome b	Cyt b	Cytochrome b	1141	H-strand
15955–16023	tRNA	P	tRNA proline	69	L-strand

regions as compared to the rest of the genome, known as the hypervariable (HV) regions 1 and 2 (Horai and Hayasaka 1990).

Hypervariable region 1 (HV1) sequencing from nucleotide 16024–16365 and hypervariable region 2 (HV2) sequencing from nucleotide 73–340 are the regions of significance in terms of forensic interests. Relatively small stretch, highly polymorphic nature, and inter-person discriminatory power of these hypervariable regions add to their utility in forensic caseworks. Interpopulation variability in terms of mitochondrial genome has been widely reported in evolutionary and phylogenetic studies. In addition to these, hypervariable region 3 (HV3) with nucleotide 438–576 along with complete sequence of D-loop are found to be the sites of evolutionary mutational hotspots and are highly significant in determination of haplotype and ancestral origin of an individual (Hwa et al. 2012). Studies have revealed escalation of mtDNA mutation to a rate of 10–17 folds as compared to the nuclear DNA induced by exposure to oxidative damages as a result of chronic exposure to mutagenic products such as ROS formed from oxidative phosphorylation, nonfunctioning of defensive protein molecules, and incompetent genomic repair mechanisms. Accumulation of such oxidative damages leads to mitochondrial dysfunction or oxidative stress and is found to be associated with onset of senescence and ageing-related mechanisms (Reid 2018).

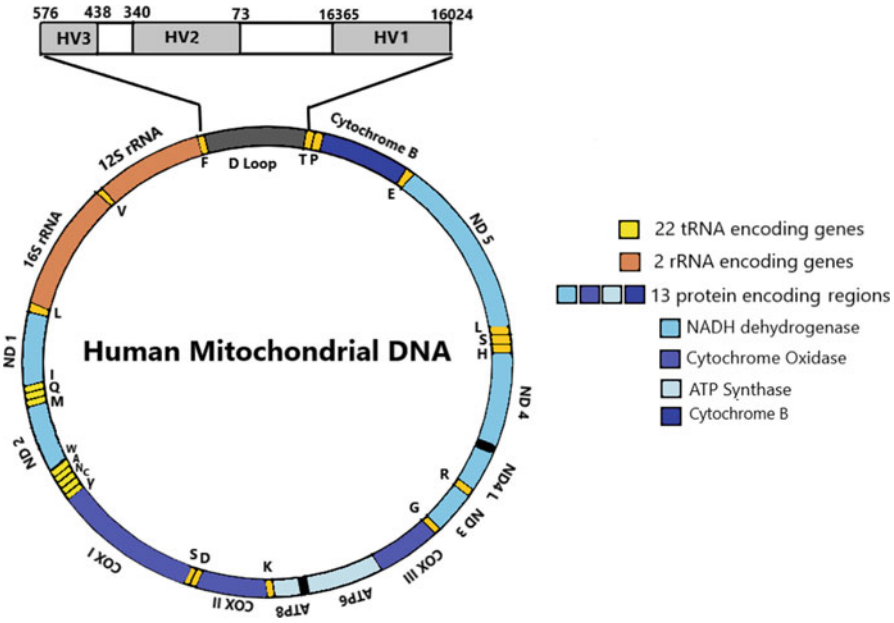


Fig. 1 Graphical representation of human mitochondrial genome

Inheritance of Mitochondrial Genome

In general terms, an individual keeps up only a single type of mitochondrial DNA (maternally inherited) regardless of the presence of several folds of mitochondrial DNA copies present in each cell, maintaining the condition of homoplasmy or identical genotype within entire mitochondrial genome (Fig. 2). Both the features of uniparental inheritance as well as homoplasmy are responsible for pertinence of mtDNA in forensic caseworks. However, the condition of heteroplasmy can raise complication in data interpretation. Heteroplasmy is the condition of existence of more than one genotype of mtDNA within an individual or tissue. This may arise due to pathogenic mutations resulting in transmission of both wild type as well as mutant type of genome. Such heteroplasmic genome may carry DNA molecules of varying length (length heteroplasmy) or different nucleotide sequence (point heteroplasmy) (Reid 2018).

Conventionally, the transmission of mitochondrial DNA was supposed to be carried out strictly along the matrilineal descent without undergoing any chromosomal recombination. However, events of biparental inheritance have also been documented in recent years (Schwartz and Vissing 2002; Luo et al. 2018) posing a challenge to the concept of central dogma and uniparental inheritance based premises of evolutionary and phylogenetic studies in forensic science. According to some theories, paternal mitochondria borne on sperms penetrate the cytoplasm of oocyte after fertilization and transiently coexists within the zygote along with bulk of

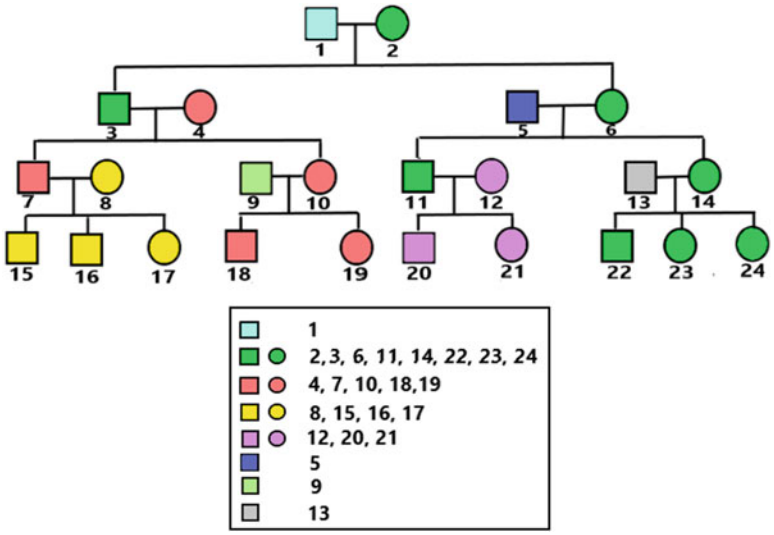


Fig. 2 Hypothetical pedigree illustrating the maternal inheritance of mtDNA. Squares represents male and circles represent females. Different colors represent unique type of mtDNA

maternal mitochondria. Later, these paternal mitochondria are eliminated during cell divisions by some mechanisms and are never transmitted to the offspring. Several hypotheses including two major hypotheses of "simple dilution model" and "active degradation model" have been laid to justify the mechanism for elimination of paternal mitochondria (Sato and Sato 2013). However, high-depth sequencing of mtDNA extending to several million folds have deplored the hypotheses of simple dilution. Besides, autophagal lysis of paternal mitochondria has also been reported in *C. elegans* (Sato and Sato 2012).

Despite such eliminating mechanisms and lysosomal pathways, chances of paternal mtDNA transmission via paternal leakage persist. Theoretically, stockpiling of small degree of male fertility-impairing mutations might be responsible for such genetic acclimatization that would necessitate paternal leakage (Vaught and Dowling 2018). Existence of such mechanisms demands analysis of paternal mtDNA transmission in cases with sporadic mitochondrial mutations and occurrence of heteroplasmy.

Application of Mitochondrial DNA Analysis in Forensic Caseworks

Mitochondrial DNA analysis has been employed to several forensic cases related to historical individual identification, genealogy, and phylogeny. Average mtDNA copy number ranging from hundreds to several thousands in somatic cells presents the possibility of extracting sufficient amount of DNA from degraded or compromised samples encountered in forensic caseworks where nuclear DNA cannot be recovered in sufficient quantities to be typed. mtDNA analysis is

considered to be useful in a range of biological materials such as hair, bones, teeth, whose nuclear DNA quantity may be low depending upon the circumstances. Manner of mtDNA inheritance is another beneficial characteristic useful in individual identification where no direct relative is available to use as reference sample. A distant relative, matrilineally related person to be identified can be used as reference to establish maternal link between the two for possible identification.

Human Identification from Degraded Compromised Samples

Skeletal remains are often the only source of DNA for human identification in cases of mass disasters and missing person identification. Human bodies found in such cases are generally exposed to harsh environments with high temperature and humidity, traumatic events, and long range of time span. Bones and teeth are the only source of DNA available after enduring such severe conditions. However, exposure to ultraviolet rays, humidity, low pH, and enzymatic reaction results in degradation of nuclear DNA in such samples. Linear structure of nuDNA and the chromatin structure may have a deteriorating impact on nuDNA but have no such relation with mtDNA (Foran 2006). High copy number per cell and ability of mtDNA to resist degradation are the grounds on which mitochondrial DNA profiling yields better results compared to nuclear DNA profiling analysis in case of hard tissue. Similarly, genomic content in hairs is comparatively low due to catabolic breakdown of nucleic acid in the course of keratinization process. In hair samples, retrieval of mtDNA is much greater due to its high copy number than nuclear DNA in terms of quantity and quality, even from the shafts of telogen or rootless hairs (Melton et al. 2012).

Skeletal remains of American soldiers excavated from tombs of Vietnam and returned to US government after Vietnam War in 1984 were identified by the application of mtDNA analysis where nuclear DNA analysis, HLA-DQ alpha, and VNTR analysis remain unsuccessful in positive identification (Holland et al. 1993).

mtDNA proved to be an efficient evidence in the trial of Tennessee Murder Case (1996) in a US courtroom for convicting 27-year-old Paul Ware with the charges of rape and murder. This was the first case of utilizing mtDNA as evidence in any courtroom. No exchange of biological fluids or evidence were found on the body of victim and suspect Ware, and only the circumstantial evidences presented against Ware were insignificant. However, during the autopsy, a red hair was found inside the throat of the victim. On reanalysis of the crime scene, several red hairs were found in the bed. One-one hair, each from the throat and bed were processed for mtDNA analysis. As a reference, mtDNA was also extracted from saliva of the suspect and blood of the victim. On comparison of the questioned sample with reference samples, both the questioned samples were found to have same source and their mtDNA sequence had an exact match with the suspect. The sequence from the samples did not match with the victim's source (Davis 1998).

A human identification case of highly decomposed body in Rio de Janeiro was resolved by implementation of MPS techniques on mtDNA fragments. The body was found in the seacoast of Rio de Janeiro in 2015. The investigation brought forth the alleged mother, but the advancing stage of decomposition of the body raised complication on processing routine STR profiling for its identification. Samples for DNA testing (both STR profiling and mtDNA sequencing) were acquired from the bone fragments of the body and buccal swabs of the alleged mother. STR analysis of bone samples generated partial profile due to sample degradation resulting in inconclusive results. However, sequence of mtDNA genome resulted in full profile of haplotype. Final haplotype generated from the bone of the body and buccal swab of alleged mother is represented in Table 2. Sequences generated from both the samples were fully concordant and thus matrilineage relation between the two cannot be excluded. As per the phylogeographic analysis based on online softwares EMPOP and HAPLOGREP, haplogroup of son/mother belonged to L3b1a (Bottino et al. 2021).

Age Estimation Using mtDNA Analysis

Several works over last two decades suggest that accumulation of mtDNA degradation and mutation contribute to physiological mechanisms that result in ageing and age-related diseases owing to enhanced oxidative damages caused by ROS generation as well as deteriorated repair mechanisms and characterized lesser replication extent in comparison with nuclear DNA (Fig. 3). Age-dependent mutations are frequently observed in postmitotic tissues that have high demands of energy. Clinical manifestations related with neurological, cardiac, hepatic, osteologic, and renal dysfunctions are common repercussion resulting from age-dependent mtDNA mutations. Accumulation of several mutations including deletion, duplication, and point mutation has been reported to have increased frequency with increasing age. Among these mutations, 4977-bp deletion is the

Table 2 Haplotype generated from the casework samples

Source of sample	Haplotype
Bone from the decomposed body	73G 263G 750G 1438G 2706G 3450T 4769G 5773A 6221C 7028T 8701G 8860G 9449T 9540C 10086G 10373A 10398G 10873C 11002G 11719A 12001T 12705T 13105G 13790G 13914A 13980C 14766T 15301A 15311G 15326G 15824G 16124C 16223T 16278T 16311C 16362C 16519C
Buccal swabs from alleged mother	73G 263G 750G 1438G 2706G 3450T 4769G 5773A 6221C 7028T 8701G 8860G 9449T 9540C 10086G 10373A 10398G 10873C 11002G 11719A 12001T 12705T 13105G 13790G 13914A 13980C 14766T 15301A 15311G 15326G 15824G 16124C 16223T 16278T 16311C 16362C 16519C

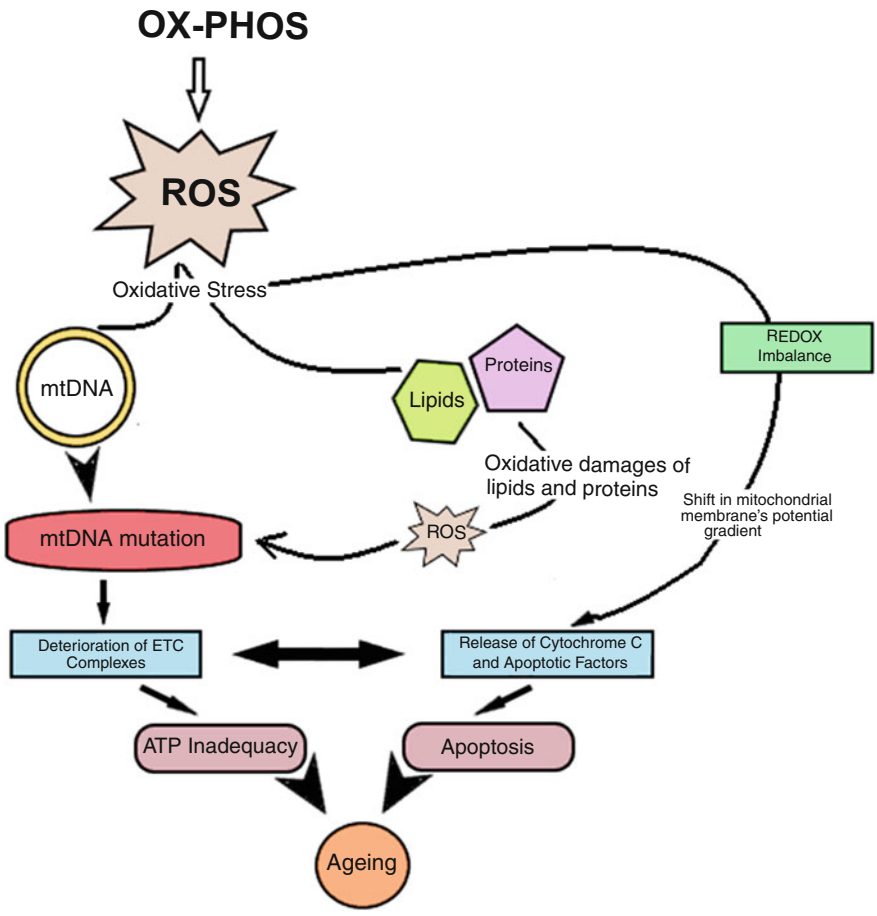


Fig. 3 Conceptual representation of various mechanisms of mtDNA mutation and its relation with the process of senescence or ageing

most common age-related mutation in mitochondrial genome and is employed for age estimation at time of death. Correlation between 4977 deletion and ageing has been successfully demonstrated in several studies and is found effective in discriminating elder person and young person in cases of mass disasters (Meissner et al. 1999; Trifunovic 2006). Besides this, 3715- and 6278-bp deletions in skin samples of ageing people have been reported to possess similar frequency as that of 4977-bp deletion (Eshaghian et al. 2006). Point mutation at 186-bp transiting from A to G at significant levels has been identified in skeletal tissue of ageing individuals. However, coexistence of mutant type (186G) along with wild type (186A) resulting in heteroplasmy possesses difficulty in deciphering mtDNA damages and their extent.

Genealogy Tracing

Maternal inheritance of genetic polymorphism across mitochondrial sequence without undergoing any kind of recombination accounts for the accumulation of these variants along the diverging maternal lineages in forms of phylogenetically related haplotypes. The group of such haplotypes that shares same genetic variants descendant from a common ancestor is known as haplogroups. Gradual mutations in the mitochondrial genome accumulated over time has led to the clustering of individuals into discrete haplogroups. Mitochondrial Eve (mt-Eve), also known as matrilineal most recent common ancestor (or mt-MRCA), is considered as the female biological ancestor of all humans inhabited in African subcontinent over 150,000–200,000 years before present (YBP). Mt-Eve diverged into seven major lineage specific groups – L0, L1, L2, L3, L4, L5, and L6 around 100,000 YBP. On facing harsh climatic changes and interstadial phases, a part of the population belonging to L3 group migrated from Africa to inhabitate other parts of the world resulting in generation of two macro-haplogroup – M defining the population that migrated to Asia and N went to Eurasia, Asia, and America. More than 30 subclades of haplogroup M is present in Asian continent that includes subclade A, B, C, D, G, and F. Haplogroup N served as the root for group R, that resulted in European haplogroups H, I, J, K, T, U, V, W, X around 45,000 YBP. HV, U, and JT are the major macro-haplogroups from European continent. Indian subcontinent and south-east Asian continents constitutes of M, N, and R. These subclades are further divided into many sub-haplogroups. At present, the global mtDNA haplogroup tree consists of more than 4000 discrete haplogroups (Mancuso et al. 2011; Chinnery and Gomez-Duran 2018).

MtDNA analysis for genealogical tracing was successfully employed to authenticate the identity of last Tsar Nicolas Romanov II and his family. Tsar Romanov II, his wife Tsarina Alexandra and their children Maria, Tatiana, Anastasia, Olga, and Alexei along with three servants and a doctor were killed during the Bolshevik Revolution of 1918. Nine sets of skeletal remains were excavated in 1991 in a mass grave with suspicion of them belonging to Romanov Family. Remains of Tsarina Alexandra and three children were identified by comparison of mtDNA sequence with the biological sources of a known maternal descendant – Prince Phillip, Duke of Edenburg. MtDNA sequence obtained from skeletal remains suspected to be that of Tsar matched with two of his maternal relatives with an uninterrupted maternal descendancy of Nicolas's maternal grandmother – Duke of Fife and Princess Xenia Cheremeteff Sfiri except at L16169 showing heteroplasmic state of C/T. For obtaining a complete authenticity of the remains, mtDNA analysis was performed on the excavated remains of Nicolas's brother – Duke of Russia Georgij Romanov (death in 1889). Comparison of mtDNA sequence from suspected Nicolas and his brother confirmed the presence of heteroplasmy (Gill et al. 1994). Remains of two children excavated in 2007, presumed to be of Tsar's two children, Anastasia (female, 18–23 years old) and Alexei (male, 10–14 years old), were subjected to mtDNA sequencing and its comparison to the mtDNA sequence generated from the remains of Alexandra and her three daughters. Thus, mtDNA sequencing played a significant role in identification of Romanov Family

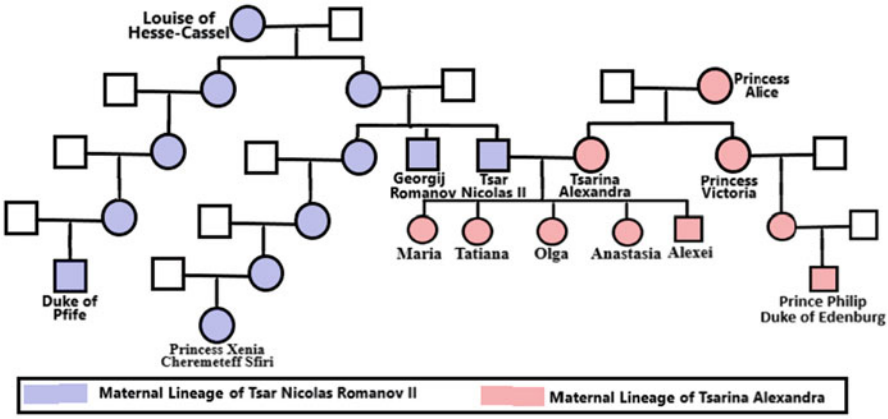


Fig. 4 Pedigree demonstrating the maternal lineage of Tsar Nicolas and Tsarina Alexandra showing maternal relationships of Tsar with Duke of Fife, Princess Xenia Cheremeteff Sfiri, Georgij Romanov, and maternal relationships of Tsarina Alexandra and her children with Prince Phillip

and tracing maternal descendancy (Coble et al. 2009; Rogaev et al. 2009). Pedigree chart demonstrating the matrilineage phylogeny of Tsar Nicholas and Tsarina Alexandra is given in Fig. 4.

Mitochondrial DNA Analytical Methodologies in Forensic Caseworks

MtDNA analysis for any forensic casework, either involving human identification or genealogical assessment, requires processing of the questioned sample (Q) collected from the scene of crime along with the processing of a known sample (K) collected from the maternal relative and their comparison. Contamination-free environment is the major requirement of mtDNA analysis, as high copy number of mtDNA than nuDNA within the sample makes it more susceptible to external contaminations. Figure 5 describes the complete procedure of mtDNA data generation, comparison, and interpretation. Procedure follows general step of DNA extraction, its quantitation, amplification of control region, sequencing of amplified product, and data interpretation.

Data Interpretation

Cambridge Reference Sequence(s)

Complete sequencing of mtDNA was determined for the first time in 1981 in MRC Laboratory, Cambridge, England (Anderson et al. 1981). The proposed sequence, known as "Anderson" sequence or Cambridge Reference Sequence (CRS)

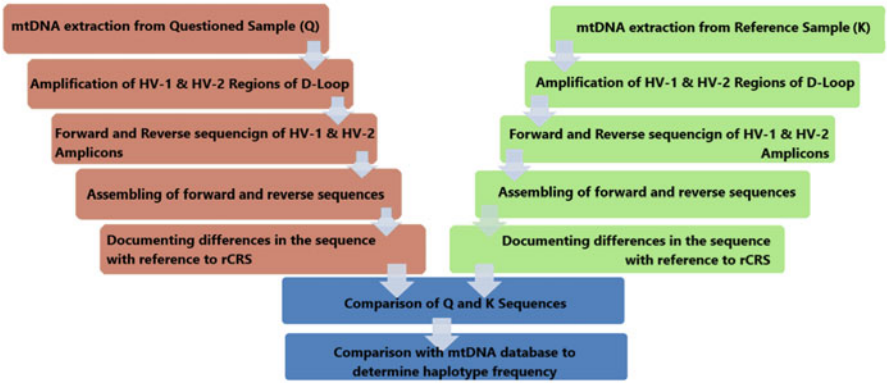


Fig. 5 Analytical process of evaluating mtDNA sample. Q represents the questioned sample that is collected from the crime scene and K represents the known or reference sample collected from the maternal relative of suspected individual

(GenBank accession: M63933), was considered as a reference sequence for comparison and interpretation of newly analyzed sequence over the years. Reports of any analyzed mtDNA sequence were generated in term of variation in the sequence as compared to the light strand (L-strand) of original Anderson sequence. Later in 1999, the source of DNA used for generating CRS was re-sequenced by Andrews et al. (1999) and came up with 18 rectified or corrected nucleotide positions including loss of one cytosine base at position 3106. The corrected sequence was designated as revised CRS (or rCRS) and acknowledged as new standard for comparison purposes. Differences observed between CRS and rCRS are listed in Table 3. Sequence comprised in rCRS belongs to an individual from European descent belonging to haplogroup H2a2 (Bandelt et al. 2014).

Several web-based bioinformatics software are available that provide a platform for management and analysis of mtDNA sequence conveniently. "mtDNAmanager" is one such interface that allows computational analysis of mtDNA control regions and identify its probable haplotype (Lee et al. 2008). BLAST or basic local alignment search tool helps in comparison and identification of mtDNA sequence and thereby is applicable for determining evolutionary sequence between the sequences. SeqScape, DNAStar'sSeqMan Pro, Variant Reporter are some of the popular web-based interfaces for mtDNA sequence analysis. GeneMarker HTS, Converge Forensic Analysis Software, mitoAnalyzer, mtDNA-Server facilitates mtDNA analysis based on next-generation sequencing (NGS) platforms with the ability to identify minor heteroplasmic variants, phylogenetically correct SNP and INDEL base calls, and correct haplogroup prediction.

Nomenclature of mtDNA Sequence

Reporting of mtDNA sequence demands a minimum data format in the form of difference with reference to rCRS. Any observed difference is noted in the format

Table 3 Nucleotide difference between original CRS and rCRS sequence

Position of nucleotide	mt Genome region	CRS sequence	rCRS corrections	Comments
263	D Loop	A	A	Rare polymorphism
311–315	D Loop	CCCCC	CCCCC	Rare polymorphism
750	12 rRNA	A	A	Rare polymorphism
1438	12 rRNA	A	A	Rare Polymorphism
3106–3107	16 rRNA	CC	C	Error
3423	ND1	G	T	Error
4769	ND2	A	A	Rare polymorphism
4985	ND2	G	A	Error
8860	ATP6	A	A	Rare polymorphism
9559	COX III	G	C	Error
11335	ND4	T	C	Error
13702	ND5	G	C	Error
14199	ND6	G	T	Error
14272	ND6	G	C	Error (bovine)
14365	ND6	G	C	Error (bovine)
14368	ND6	G	C	Error
14766	ND6	T	C	Error (HeLa)
15326	Cyt b	A	A	Rare polymorphism

of listing nucleotide position followed by the base present at that point. Presence of ambiguous base is marked by N. Heteroplasmy at any nucleotide is coded in the form of X/Y where X is the wild type base and Y is the mutant type base. In case of insertion of a base in a sequence as against rCRS, the nucleotide position is denoted by noting the site followed by a point and number of insertions followed by the base. Deletions are coded as "-" or "D" or "d" or "del" after the nucleotide position.

mtDNA Sequencing Result Interpretation

Comparison of questioned (Q) sample with known (K) samples for all 610 nucleotides (16024–16365, 73–340) is essential for conclusion of the results. On account of Q-K sequence comparison, the interpretation of mtDNA sequencing result can be categorized into three classes – exclusion, failure to exclusion, and inconclusive.

- **Exclusion:** Two or more nucleotide difference between questioned and known sample with no evidence of heteroplasmy.

- **Failure to exclusion:** (i) Fully concordant Q-K pair, (ii) one heteroplasmic base in both the samples at the same position, (iii) one heteroplasmic base in one sample, not observed in the other sample, with a common nucleotide present in both samples, (iv) two heteroplasmic bases at the same position in both samples, and (v) one heteroplasmic base at the same position, one base showing heteroplasmy in one sample but not in the other, with a common nucleotide in each.
- **Inconclusive:** Difference of one nucleotide between Q-K pair, with no signs of heteroplasmy and one heteroplasmic base at the same position, one different base at another position with no evidence of heteroplasmy.

Human Mitochondrial DNA Sequence Database

DNA database of any population is essential for estimation of expected haplotype frequency of the given samples. Collection of high-quality genetic information from a large number of maternally unrelated individuals offers reliable frequency estimation for a random match. Several databases of mtDNA sequences have been compiled globally that focuses on sequence information as well as phylogenetic information. Some of the popular databases include:

- **MITOMAP** – MITOMAP is a human mitochondrial genome database developed in 1996 that focuses on collecting information related with mtDNA polymorphisms and mutation in humans. It involves real-time sequencing method (Kogelnik et al. 1996).
- **EMPOP** – EDNAP Mitochondrial DNA Population Database (EMPOP) is a collaborative exercise of European DNA Profiling (EDNAP) body operational since 2004. The database aims at determining the uniformity of mtDNA analysis among different laboratories and identifying the possible source of errors (Parson and Dür 2007).
- **HmtDB** – It is an open-source genomic resource developed in 2005 that comprise of human mtDNA sequences with information related with population variability. The database is effective for population genetics as well as mitochondrially generated diseases (Attimonelli et al. 2005).
- **MitoVariome** – The database was created in 2009 for aiding the works on human evolution and variation. It is a cost-free accessible database in textual and graphical format beneficial for research works related with forensic science, tumors, ageing, and degenerative diseases (Lee et al. 2009).
- **AmtDB** – First database of ancient mitochondrial DNA reported in 2019. It is beneficial for tracking human past demographic events (Ehler et al. 2019).

EMPOP: Innovative and Forensically Significant Human mtDNA Database

In 2004, European DNA Profiling (EDNAP) group collaborative exercised on the development of online database consisting of mtDNA sequence data obtained from various populations known as EDNAP Mitochondrial DNA Population Database

(EMPOP). The database aims at determining the uniformity and concordance of mtDNA analysis among different laboratories, supplying of a uniform stage for data interpretation and nomenclature of mtDNA analysis procedure, identifying the possible source of errors and development of a secured IT-based platform for logistic data transport and storage. It involved the usage of PHRED – computer software for computing base calls and assigning them a quality score (Parson et al. 2004). Website of EMPOP – www.empop.org was launched in October 2006. Since then, the database continuously evolved in terms of quality control and feasible results, and has emerged as the globally known repository for mitotyping data. Software based on quasi-median network analysis-based software was developed to review mtDNA data table in order to identify the occurrence of any possible type of error including that of interpretation as well as transcription errors and keep a check on quality assurance. Condition of mixing of HVI and HVII, also known as "artificial recombination," was one of the common errors rectified with specific tools. Network analysis software package in the name of NETWORK have been introduced for routine evaluation of mtDNA sequence. It makes use of three filters designed for a specific function – EMPOPspeedy for removal of frequent mutation, EMPOPall for deletion of all mutations, and Unfiltered for no removal of mutations (Parson and Dür 2007). SAM, the string-based search software, was introduced in 2011 to harmonize EMPOP searches and rectify the issue of phylogenetic alignment difference in case of database searches that require rCRS comparison (Röck et al. 2011). The issue of alignments in casework mitotypes was resolved but, reporting and alignment ambiguity in population studies still remained an unsolved problem. Latest version of the database EMPOP 4 came up with modified string-based search software – SAM2 that comes with unbiased and harmonized database useful for not only the field of forensic but also for other genetic areas (Huber et al. 2018).

Modifications to the Current Methods of Mitochondrial DNA Analysis and Future Perspective

Screening Approaches for mtDNA Analysis

Methods of mitochondrial DNA analysis currently being utilized for obtaining full sequence information are time-consuming, expensive, and laborious. Consequently, these conventional methods demand utilization of screening assays in order to avoid analysis of complete sequence of the samples that could be excluded. Screening assays often includes physical screening, anthropological screening prior to performing mtDNA analysis in order to eliminate the unnecessary exhibits and concentrate on prime samples. MtDNA variation-based screening is employed to escape complete mtDNA sequencing across HVI and HVII regions. Sequence-specific oligonucleotide (SSO) probe, linear array typing, minisequencing, pyrosequencing, denaturing gradient gel electrophoresis (DGGE) are some of the commonly practiced screening assays performed prior to mtDNA sequencing in forensic caseworks (Butler 2011).

Currently Employed CE Technique Versus Next-Generation Sequencing Approaches

Sanger sequencing, also considered as the first-generation sequencing, was the earliest approach towards determination of nucleotide sequence in a DNA strand. The method involves incorporation of dideoxynucleotides triphosphates (ddNTPs) onto the amplified DNA molecules that results in production of DNA strands of varying lengths. Four types of DdNTPs namely ddATP, ddTTP, ddGTP, and ddCTP were utilized in the method. Sanger sequencing method is one of the widely employed methods for detection of mtDNA mutations. It can generate data of 25–1200 nucleotides. Also, the method is cost-effective and user-friendly. However, detection of low levels of heteroplasmy (<15%) is unachievable, indicating low degree of sensitivity of the sequencing method (Zhou et al. 2020).

Next-generation sequencing (NGS) technique is a class of advanced sequencing technique developed since 2005. These modern approaches are massively parallel sequencing techniques designed to provide enhanced sensitivity and high-throughput in detection of mtDNA mutations as well as heteroplasmy. Roche 454 based on pyrosequencing, Illumina "HiSeq" based on incorporation of fluorescently tagged nucleotides and ABI SOLiD based on ligation are some of the popular NGS platforms for mtDNA analysis. In spite of the rapid advancement, NGS platforms still need upgradation of several forensic-based softwares so as to be employed efficiently in the legal system. Conventional method of Sanger sequencing focusing upon the targeted mitochondrial regions, particularly HVI and HVII are used in most of the forensic science laboratories. In the recent years, the range of targeted sequence or regions have been extended to HVIII regions of D-loop resulting in sequencing of whole control regions (Sinha et al. 2020).

Implementation of whole genome sequencing (WGS) and whole exome sequencing (WES) data for mtDNA analysis can be useful in genomic screening, detection of mutation, as well as assessment of heteroplasmy. WGS and WES are applicable for identifying broad range of variants including single nucleotide polymorphisms (SNPs) (Parsons 2006), insertion/deletion (INDEL) variants (Diroma et al. 2020), structural polymorphisms (Yuan et al. 2020), copy number variants (CNVs) (Longchamps et al. 2020), pseudogenes (Cihlar et al. 2020), as well as heteroplasmy (Li et al. 2010; Cao et al. 2017), and therefore stands out as an effective technique for forensic human identification. In spite of being a time-consuming process, the technique of WGS is gaining attentions due to its advantages of cost-effectiveness and reduction in trial and experimental efforts (Duan et al. 2019).

Hybridization Capture-Based Target Enrichment Coupled with Massive Parallel Sequencing

Degraded biological samples are quite often encountered in forensic caseworks that possess problems in ideal extraction of DNA from such samples, thereby hampering the analytical procedure based on Sanger sequencing. Conventional hybridization

capture-based target enrichment combined with NGS technology can facilitate in detection of degraded and decomposed DNA and enhance their yield. Hybridization-based target enrichment is a DNA preparation step performed prior to DNA sequencing to enrich and sequence the part of genome containing the region of interest thereby allowing analysis of DNA fragments as small as 30 bp, often ranging below the PCR threshold. The process involves conversion of DNA fragments into a DNA library with the help of barcoded adapters, immortalization of DNA by PCR primers complementary to the adapter, and hybridization enrichment. Hybridization capture-based target enrichment coupled with massive parallel sequencing has been employed on a wide range of forensic samples such as simulated degraded DNA, telogenic hairs, chemical-exposed DNA, buried skeletal remains, and archeological samples (Young et al. 2019; Sinha et al. 2020).

Conclusion

Mitochondrial DNA analysis has been used for the past three decades in various forensic caseworks that involves human identification. Besides its favorable characteristics of high copy number and susceptibility to external degradations, the technical advancements in mtDNA analysis play an incredible role in enhancing its utility in analysis of variety of samples ranging from tiny degraded fragments of DNA to sequencing of entire mt-genome in short span of time. Over the past two decades, mtDNA analysis has been availed for forensic caseworks related with human identity such as mass disaster, missing-persons identification, decomposed skeletal remains identification, as well as maternal dispute and child swapping cases. Besides individual identification in forensic caseworks, the detailed analysis of mitochondrial genome is effectual in clinical, genealogical, and human evolutionary studies. In spite of several justified explanations for collecting detailed history of suspected offender related with his phylogeny or health conditions, many researchers contemplate it as unethical to unveil someone's genetic structure that reveals his personal information. Therefore, the advancing technologies related with mtDNA analysis must be mindfully utilized for expected purposes. Besides ethical complications, technical and structural drawbacks such as condition of heteroplasmy, paternal leakage, nuclear pseudogene, and sample mixture raise complication in analysis and interpretation of mitochondrial DNA sequences, which needs to be addressed by further technical advancement. Ability to recognize heteroplasmy and coherent explanation of biparental inheritance by recombination are some of the major challenges that must be focused in the interest of mtDNA analysis to emerge as an important alternative for forensic purposes.

References

Anderson S, Bankier AT, Barrell BG et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465. https://doi.org/10.1038/290457a0

Andrews RM, Kubacka I, Chinnery PF et al (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Attimonelli M, Accetturo M, Santamaria M et al (2005) HmtDB, a human mitochondrial genomic resource based on variability studies supporting population genetics and biomedical research. BMC Bioinform 6:1–9

Bandelt H-J, Kloss-Brandstätter A, Richards MB et al (2014) The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies. J Hum Genet 59:66–77

Bottino C, Silva R, Moura-Neto RS (2021) Resolving a human identification case for the Rio de Janeiro Police with massively parallel sequencing of mtDNA using a proposed pipeline. Genet Mol Res 20(1):GMR18757

Butler JM (2011) Advanced topics in forensic DNA typing: methodology. Academic

Cao Y, Zou KN, Huang JP et al (2017) Whole genome sequencing of human mtDNA based on ion torrent PGM™ platform. Fa Yi Xue Za Zhi 33:368–373

Chinnery PF, Gomez-Duran A (2018) Oldies but goldies mtDNA population variants and neurodegenerative diseases. Front Neurosci 12:682

Cihlar JC, Strobl C, Lagacé R et al (2020) Distinguishing mitochondrial DNA and NUMT sequences amplified with the precision ID mtDNA whole genome panel. Mitochondrion 55: 122–133

Coble MD, Loreille OM, Wadhams MJ et al (2009) Mystery solved: the identification of the two missing Romanov children using DNA analysis. PLoS One 4:e4838

Davis CL (1998) Mitochondrial DNA: State of Tennessee v. Paul Ware. Profiles DNA 1:6–7

Demirel Y, Gerbaud V (2019) Thermodynamics and biological systems. In: Nonequilibrium thermodynamics. Elsevier, pp 489–571

Diroma MA, Varvara AS, Attimonelli M et al (2020) Investigating human mitochondrial genomes in single cells. Genes (Basel) 11:534

Duan M, Chen L, Ge Q et al (2019) Evaluating heteroplasmic variations of the mitochondrial genome from whole genome sequencing data. Gene 699:145–154

Ehler E, Novotný J, Juras A et al (2019) AmtDB: a database of ancient human mitochondrial genomes. Nucleic Acids Res 47:D29–D32. https://doi.org/10.1093/nar/gky843

Eshaghian A, Vleugels RA, Canter JA et al (2006) Mitochondrial DNA deletions serve as biomarkers of aging in the skin, but are typically absent in nonmelanoma skin cancers. J Invest Dermatol 126:336–344

Foran DR (2006) Relative degradation of nuclear and mitochondrial DNA: an experimental approach. J Forensic Sci 51:766–770

Gill P, Ivanov PL, Kimpton C et al (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6:130–135

Glancy B, Kim Y, Katti P, Willingham TB (2020) The functional impact of mitochondrial structure across subcellular scales. Front Physiol 11:1462

Gray MW, Burger G, Ranz Lang B (2001) The origin and early evolution of mitochondria. Genome Biol 2:1018. https://doi.org/10.1186/gb-2001-2-6-reviews1018

Holland MM, Fisher DL, Mitchell LG et al (1993) Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. J Forensic Sci 38:542–553

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46:828

Huber N, Parson W, Dür A (2018) Next generation database search algorithm for forensic mitogenome analyses. Forensic Sci Int Genet 37:204–214

Hwa HL, Ko TM, Chen YC et al (2012) Sequence polymorphisms of mtDNA HV1, HV2 and HV3 regions in eight population groups living in Taiwan. Aust J Forensic Sci 44:243–252. https://doi.org/10.1080/00450618.2011.650208

Kogelnik AM, Lott MT, Brown MD et al (1996) MITOMAP: a human mitochondrial genome database. Nucleic Acids Res 24:177–179. https://doi.org/10.1093/nar/24.1.177

Kühlbrandt W (2015) Structure and function of mitochondrial membrane protein complexes. BMC Biol 13:1–11

Lee HY, Song I, Ha E, Cho SB, Yang WI, Shin KJ (2008) mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. BMC Bioinformatics 9:483. https://doi.org/10.1186/1471-2105-9-483

Lee YS, Kim W-Y, Ji M et al (2009) MitoVariome: a variome database of human mitochondrial DNA. In: BMC genomics. Springer, pp 1–5

Li M, Schönberg A, Schaefer M et al (2010) Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. Am J Hum Genet 87:237–249

Longchamps RJ, Castellani CA, Yang SY et al (2020) Evaluation of mitochondrial DNA copy number estimation techniques. PLoS One 15:e0228166

Luo S, Valencia CA, Zhang J et al (2018) Biparental inheritance of mitochondrial DNA in humans. Proc Natl Acad Sci 115:13039–13044

Mancuso M, Ienco EC, Simoncini C et al (2011) May mitochondrial eve and mitochondrial haplogroups play a role in neurodegeneration and alzheimer's disease? Int J Alzheimers Dis. 2011: 709061. https://doi.org/10.4061/2011/709061

McCarron JG, Wilson C, Sandison ME et al (2013) From structure to function: mitochondrial morphology, motion and shaping in vascular smooth muscle. J Vasc Res 50:357–371. https://doi.org/10.1159/000353883

Meissner C, von Wurmb N, Schimansky B, Oehmichen M (1999) Estimation of age at death based on quantitation of the 4977-bp deletion of human mitochondrial DNA in skeletal muscle. Forensic Sci Int 105:115–124

Melton T, Dimick G, Higgins B et al (2012) Mitochondrial DNA analysis of 114 hairs measuring less than 1 cm from a 19-year-old homicide. Investig Genet 3:1–5

Parson W, Dür A (2007) EMPOP – a forensic mtDNA database. Forensic Sci Int Genet 1:88–92. https://doi.org/10.1016/j.fsigen.2007.01.018

Parson W, Brandstätter A, Alonso A et al (2004) The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. Forensic Sci Int 139:215–226

Parsons TJ (2006) Mitochondrial DNA genome sequencing and SNP assay development for increased power of discrimination. US Army Research. Paper 104. http://digitalcommons.unl.edu/usarmyresearch/104

Pearce SF, Rebelo-Guiomar P, D'Souza AR et al (2017) Regulation of mammalian mitochondrial gene expression: recent advances. Trends Biochem Sci 42:625–639

Plecitá-Hlavatá L, Ježek P (2016) Integration of superoxide formation and cristae morphology for mitochondrial redox signaling. Int J Biochem Cell Biol 80:31–50

Reid R (2018) Applications of the mitochondrion in forensic DNA typing. In: DNA fingerprinting: advancements and future endeavors. Springer, pp 241–255

Röck A, Irwin J, Dür A et al (2011) SAM: string-based sequence search algorithm for mitochondrial DNA database queries. Forensic Sci Int Genet 5:126–132

Rogaev EI, Grigorenko AP, Moliaka YK et al (2009) Genomic identification in the historical case of the Nicholas II royal family. Proc Natl Acad Sci 106:5258–5263. https://doi.org/10.1073/pnas.0811190106

Sato M, Sato K (2012) Maternal inheritance of mitochondrial DNA: degradation of paternal mitochondria by allogeneic organelle autophagy, allophagy. Autophagy 8:424–425. https://doi.org/10.4161/auto.19243

Sato M, Sato K (2013) Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochim Biophys Acta, Mol Cell Res 1833:1979–1984

Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. N Engl J Med 347:576–580

Shoubridge EA (2004) Mitochondrial inheritance. In: Encyclopedia of biological chemistry. Elsevier, pp 716–719

Sinha M, Rana M, Kushwaha P (2020) Applications of mitochondrial DNA in forensic science. In: Forensic DNA typing: principles, applications and advancements. Springer, pp 329–343

Spinelli JB, Haigis MC (2018) The multifaceted contributions of mitochondria to cellular metabolism. Nat Cell Biol 20:745–754

Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta Bioenerg 1410:103–123

Trifunovic A (2006) Mitochondrial DNA and ageing. Biochim Biophys Acta Bioenerg 1757: 611–617

Vaught RC, Dowling DK (2018) Maternal inheritance of mitochondria: implications for male fertility? Reproduction 155:R159–R168

Young JM, Higgins D, Austin JJ (2019) Hybridization enrichment to improve forensic mitochondrial DNA analysis of highly degraded human remains. Front Ecol Evol 7:450. https://doi.org/10.3389/fevo.2019.00450

Yuan Y, Ju YS, Kim Y et al (2020) Comprehensive molecular characterization of mitochondrial genomes in human cancers. Nat Genet 52:342–352

Zhou K, Mo Q, Guo S et al (2020) A novel next-generation sequencing–based approach for concurrent detection of mitochondrial DNA copy number and mutation. J Mol Diagn 22: 1408–1418

Applications of NGS in DNA Analysis

21

Kelly M. Elkins, Hannah E. Berry, and Kashiya R. Reese

Contents

Abstract

DNA sequencing has led to the discovery of point mutations in DNA sequences that result in mutated, faulty proteins, problematic promoter signals, and genetic variations that have been determined to be the basis of genetic diseases and the basis for human identification. Traditional Sanger sequencing is limited to sequencing relatively short stretches of DNA (hundreds to approximately a thousand nucleotides) at a time. Pyrosequencing is a valuable tool for detecting point mutations and epigenetic patterns but also is limited in the length of sequence it can process in an experiment. In many ways, next-generation sequencing (NGS) has revolutionized DNA profiling. NGS methods have enabled the sequencing of whole genomes of organisms or large targeted portions of genomes in a single experiment. NGS has been employed for both human and

K. M. Elkins (✉) · H. E. Berry · K. R. Reese
TU Human Remains Identification Lab (THRIL), Department of Chemistry, Towson University, Towson, MD, USA
e-mail: kmelkins@towson.edu

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_19

non-human applications. This chapter will highlight examples of the breadth of DNA profiling applications including screening and identifying biological specimens for cases of human identification in forensic investigations, parentage testing, characterization of the microbiome of human body regions and body fluids, determining the molecular genetics of inherited and rare diseases for clinical diagnosis, determining sequence variants in cancer, plant biology, pathogen surveillance, infectious disease monitoring, enabling environmental studies of uncultured microbes using metagenomics, geolocation, archeology, studies of human genetic change and adaptation, and human genealogy research.

Keywords

NGS · MPS · Biofilms · Body fluids · Cancer · Pathogen · Infectious disease · Medical microbiology · Environmental microbiology · Plant microbiology · Evolutionary biology · Microbiome · Forensic biology · Human identification · Parentage testing · Cold case

Introduction

A Brief History of DNA Typing in Forensic Science

DNA profiling methods have changed the landscape of investigative science and the diagnosis and treatment of human disease, among other applications. The earliest DNA profiling methods using minisatellites emerged in the 1980s and were used to solve the cases of abduction, rape, and murder of two teen girls by Jeffreys and colleagues (Gill et al. 1985; Butler 2005). These cases, and those after them, harnessed the individualization capabilities of repeated DNA sequences in the genome. Repeated sequences are found throughout eukaryotic genomes. Mini- and microsatellite regions of DNA were found in thousands of regions in the genome and are characterized by DNA sequences that are repeated 5–50 times (Butler 2005). The minisatellites, which are also known as variable number of tandem repeats (VNTRs), are longer than microsatellites; minisatellite repeats range in length from 10 to over 100 base pairs (Butler 2005). In contrast, microsatellites, or simple sequence repeats (SSRs), have a repeat length of two to six base pairs and are scattered throughout genomes every 10,000 bases or so (Butler 2005). In DNA profiling, the number of the specific repeated DNA sequences were determined (Butler 2005). Restriction fragment length polymorphism (RFLP) of minisatellites was the DNA profiling method used to solve the first cases (Butler 2005). Now called traditional DNA typing methods, targeted polymerase chain reaction (PCR) followed by slab gel or capillary electrophoresis (CE) is employed to analyze mini- and microsatellite regions (Butler 2005).

For many years now, short tandem repeats (STRs) of 3 to 7 nucleotides and single nucleotide polymorphisms (SNPs) of single base variations have become common targets for genetic individualization of humans as well as pets, plants, bacteria, and viruses. Sanger DNA sequencing has been a mainstay for analyzing the sequence of

targeted 500–1000 base pair range regions of the genome since its introduction in the 1970s (Sanger et al. 1977). Several other DNA profiling methods have been developed for analyzing SNPs including reverse dot blot or linear arrays, genetic bit analysis, denaturing high performance liquid chromatography (HPLC), TaqMan PCR, mass spectrometry, fluorescence polarization, high-density (Affymetrix chip) arrays, electronic dot blot, molecular beacons, oligonucleotide ligation, PCR melt, pyrosequencing, allele-specific hybridization, and minisequencing (or SNaPshot) assays (Butler 2005). The newest method applied to DNA profiling is next-generation sequencing (NGS). NGS applications of DNA profiling will be the focus of this chapter.

Next-Generation Sequencing

NGS instruments and methodology are developed from Sanger sequencing, SNaPshot, pyrosequencing, and oligonucleotide ligation methods (Kircher and Kelso 2010). The massively parallel sequencing (MPS) NGS method was pioneered by 454 Life Sciences in the late 1990s and became commercially available in 2005 (Kircher and Kelso 2010). The NGS technology developed by Illumina and employed in its many instruments including the HiSeq, MiSeq, and iSeq is built on features of Sanger and SNaPshot sequencing by synthesis (SBS) methods (Kircher and Kelso 2010). In Sanger and SNaPshot sequencing, a dye-labeled terminator nucleotide base lacking the 3′-OH is used to detect and assign the base (Butler 2005; Kircher and Kelso 2010). Each nucleotide base is labeled with one of four different dyes with different emission maxima (Butler 2005; Kircher and Kelso 2010). In MPS using the Illumina method, the base is labeled with a dye and a temporary chain terminator (Kircher and Kelso 2010). After each base has been detected, the terminating 3′-O-azidomethyl group and dye is removed by enzymatic cleavage, and the next nucleotide is incorporated (Bruijns et al. 2018). The Ion Torrent method employed by the Thermo Fisher (formerly Life Technologies) NGS instruments also employs SBS but does not rely upon dye-labeled bases (Berkman et al. 2012; Bruijns et al. 2018). Rather, as each base is added, a semiconductor chip detects the hydrogen ion released by a pH sensor (Berkman et al. 2012, Bruijns et al. 2018). The pH change/signal is proportionally larger as more of the same base is added as in pyrosequencing (Bruijns et al. 2018). The SOLiD method is an NGS method that employs oligonucleotide ligation (Kircher and Kelso 2010; Berkman et al. 2012). A single-stranded DNA adaptor region for binding a primer conjugated to the target DNA sequence is attached to a bead deposited to a glass surface (Kircher and Kelso 2010). Fluorescently labeled eight base pair probes are introduced to the reaction (Kircher and Kelso 2010). The probe that is complementary to the target hybridizes and is attached to the probe by ligase (Kircher and Kelso 2010). The fluorescent probe is detected and cleaved and the cycle repeats (Kircher and Kelso 2010). After the complete length has been sequenced, the product is melted off and sequencing is repeated with a primer one base shorter than the last (Kircher and Kelso 2010). The sequence is computed by the emission spectra from several rounds of sequencing (Kircher and Kelso 2010).

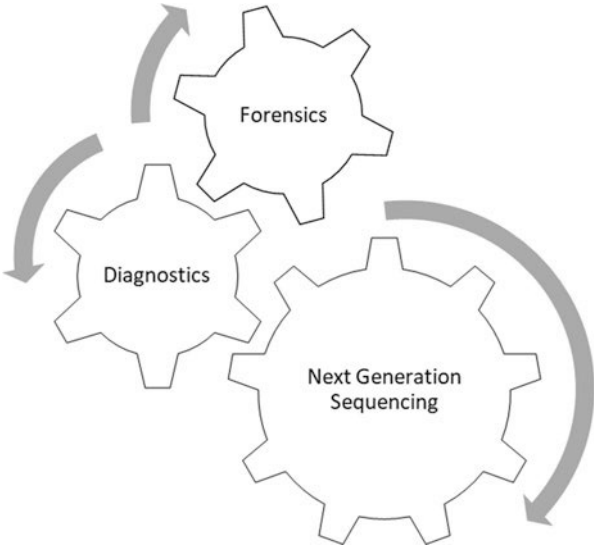


Fig. 1 NGS impact in forensics and diagnostics

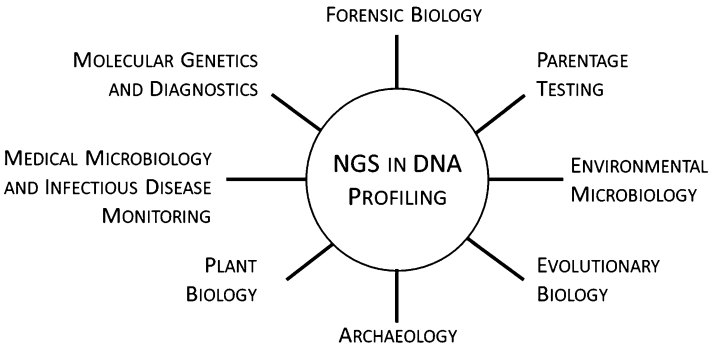


Fig. 2 Summary of forensic applications of DNA profiling using NGS

NGS DNA profiling has been applied to many fields including forensic biology, genealogy, archaeology, molecular diagnostics, molecular genetics, plant biology, forensic, medical and environmental microbiology, and biosurveillance (Figs. 1 and 2).

Applications of NGS in DNA Profiling

Molecular Genetics and Diagnostics

Although diagnostic molecular genetics originated with first-generation sequencing approaches, NGS has found numerous applications in this area. NGS has been used to determine the molecular genetics of inherited and rare diseases for clinical

diagnosis, determine sequence variants in cancer, perform cancer genome sequencing, and perform comparative transcriptome profiling. The use of NGS technologies in the identification of rare disease causing genes has improved over the years with over 230 genes being discovered (Boycott et al. 2014). With whole exome sequencing, the discovery of novel disease-causing genes has exploded since 2010 with decreased cost and increased accessibility (Boycott et al. 2014).

NGS has been used to profile minimal residual disease in hematological malignancies including acute myeloid leukemia and acute lymphoblastic leukemia, chronic lymphocytic leukemia, and multiple myeloma (Sánchez et al. 2019).

A new trial study utilizing NGS for screening patients with relapsed or refractory cancer for treatment using precision medicine was performed by the National Cancer Institute-Molecular Analysis for Therapy Choice (NCI-MATCH). The study analyzed formalin-fixed paraffin-embedded clinical specimens and cell lines with the Oncomine Cancer Panel assay. The overall sensitivity of the assay was 96.98% for 265 known mutations with 99.99% specificity and a limit of detection for each variant type of 2.8% for SNPs, 10.5% for insertion/deletions (indels), and 6.8% for large indels (gap $\geq$4 bp) (Lih et al. 2017).

Medical Microbiology

Early applications of NGS included medical microbiology, molecular genetics, and cancer biology applications. Medical microbiology is a subfield of microbiology focused on the prevention, diagnosis, and treatment of human disease. To understand which microbes cause disease, the microbiota of healthy and diseased individuals are characterized and compared. Disease-causing microbes include archaea, bacteria, viruses, fungi, and protozoa. Microbes can cause infection when ingested food or water is contaminated or colonized by foodborne pathogens. The contamination can be natural, accidental, or be caused by bioterrorism events. Microbes can also colonize areas with broken skin or accessible body surfaces or orifices and may form biofilms. Biofilms are communities of bacteria that can grow on almost every surface, forming architecturally complex communities (Joo and Otto 2012). NGS has been used to identify pathogens causing infectious disease and foodborne illness and for pathogen surveillance. Biofilms can be populated by bacteria including *Escherichia sp.* and *Vibrio sp.* that also cause foodborne infection (Joo and Otto 2012). In recent years, NGS has been applied in several studies to identify and characterize biofilms sampled from teeth and gums, wounds, transplants, and water sources from healthy and diseased individuals and sources as well as infectious disease monitoring.

The oral cavity is heavily colonized by microorganisms, including viruses, protozoa, fungi, archaea, and bacteria. Ninety-six percent of oral bacteria have been categorized into six major phyla that include Firmicutes, Actinobacteria, Bacteroidetes, Proteobacteria, Spirochaetes, and Fusobacteria, while the remaining 4% taxa belong to phyla Euryarchaeota, Chlamydia, Chloroflexi, SR1, Synergistetes, Tenericutes, and TM7. A distinct database, Human Oral Microbiome Database (HOMD), was launched in 2005 by the National Institute of Dental and

Craniofacial Research for maintaining the information of oral-derived cultivable and non-cultivable isolates. Various factors such as food habits, tobacco and alcohol consumption, stress, hormonal imbalance, puberty, oral hygiene, diabetes, and gingival inflammation influence the native bacterial community (Verma et al. 2018).

In a representative study, microorganisms found in biofilms on healthy children's gums or in plaque on their teeth were sequenced using the HOMINGS (Human Oral Microbe Identification using Next-Generation Sequencing) technique to probe the V3-V4 region in the 16S rRNA gene. Using the ProbeSeq database of probe sequences from the Human Oral Microbiome Database, the biofilms were determined to contain genera *Actinomyces*, *Fusobacterium*, *Neisseria*, and *Streptococcus* among others (Harris-Ricardo et al. 2019).

Periodontitis is an infection of the periodontium tissues in the mouth that surround and support the teeth. Factors such as diabetes, obesity, or tobacco use are known to increase the risks of periodontitis. A recent study using NGS reported upon microbial diversity in the periodontal pockets of individuals with healthy and diseased tissue. There were differences in microbial communities within and among individuals. Periodontitis was associated with a significantly higher level of microbe diversity than a healthy person. Based upon their oral health, patients were categorized into three groups: 23 of them were placed in the gingivitis group, 12 were placed in mild to moderate periodontitis group, and 1 was placed in the severe periodontitis group. All participants were given a periodontal treatment plan for at least 6 weeks before returning for a follow-up appointment. Samples were collected prior to and following treatment. The V1–V2 hypervariable regions of the 16S rRNA genes were amplified using PCR using barcoded primers. The PCR products were submitted and sequenced at a core facility using a Roche 454 GS FLX instrument. The researchers identified 87 genera of bacteria belonging to 12 different divisions and analyzed using QIIME. Interestingly, the sequences of pre- and posttreatment samples were similar in the same individual. With the use of NGS, it was determined that the regular treatment for periodontitis may not be as effective in eradicating microbes as widely thought. The treatment among a group that is affected more often had people show improvement from treatment, but the microbial diversity for the samples for both pre- and post-samples showed little to no change (Schwarzberg et al. 2014).

In another study, an average of 4.6 organisms were found among patients in their native knees prior to knee surgery, and 48 unique organisms were identified by NGS from all samples. All sterile water controls were negative for organisms. The results suggested there may be a native microbiome that exists in the knees of patients undergoing primary total knee arthroplasty (Torchia et al. 2020). NGS as a tool could be used in diagnosis or profiling of microbes in joint replacements.

NGS can be used to detect and identify bacterial infection of the organ deriving from the donor or the recipient that can cause disease and even rejection. In one study, patients who received lung transplants were observed using NGS methods to determine if preexisting donor bacteria had an effect on the recipient's prognosis after receiving the transplant (Liu et al. 2020). NGS was found to be more sensitive than bacterial culture for monitoring infection in lung transplant patients (Liu et al.

2020). It can also be used to detect organs that were transplanted and do not match the DNA profile of the other organs for human identification (Butler 2005).

NGS was used to analyze and identify biofilms present on infected suture sites in dogs. DNA was extracted from formalin-fixed, paraffin-embedded (FFPE) samples using a NucleoSpin Kit. In the study, NGS was performed using an Illumina MiSeq instrument. A fraction of the total reads were determined to be sequences from bacteria. Among the three dogs' wounds, there were two families of bacteria in common, Porphyromonadaceae and Fusobacteriaceae. Other bacterial families were present in two of the three dogs' wounds, such as Alteromonadaceae and Enterobacteriaceae. The analysis of bacteria in these biofilms may be valuable when assessing treatment options (König et al. 2014).

Forensic Microbiology and Biosurveillance

Forensic microbiology is the term used to describe the analysis of samples in cases of bioterrorism, biocrime, or the inadvertent release of a biological agent (Minogue et al. 2019). The identification of the agent is the most important step in the investigation of a biothreat. Thereafter, investigators will use the data to link the agent to a perpetrator. There are many examples of the use of NGS in the investigation of biocrimes such as the *Salmonella typhimurium* poisoning of salad bars or the intentional spread of HIV-infected blood which is considered attempted murder. The use of early NGS technologies in bioforensics became crucial in a case where *B. anthracis* spores were intentionally released into the public through the Amerithrax letters. Whereas other methods test only for a specific agent of interest, NGS can identify an agent with little to no prior information and is a distinct advantage over other PCR and culture-based methods (Minogue et al. 2019).

Biosurveillance, or detecting and collecting data from biothreats and diseases, has employed NGS methods in order to gain sequence information about viruses such as the Bass Congo virus and Lujo virus. Biosurveillance is essential for biothreat preparedness. It can be used to monitor infectious diseases such as Ebola in a population. NGS was used for biothreat biosurveillance in the most recent Ebola virus disease (EVD) outbreak. Researchers applied NGS by determining the Ebola genome and determining how EVD was being transmitted. In a case, Ebola virus transmission was suspected from an apparently healthy survivor to his female sexual partner. When it was found that EVD could be sexually transmitted, the CDC was able to change its community recommendations in an effort to decrease the spread of the virus by sexual transmission. After the outbreak, NGS was used to sequence the disease of many infected individuals to understand the stability of the Ebola genome. This was implemented in the Nigerian population to monitor the trace contact of the virus and allowed researchers to determine that there was a single trace contact rather than multiple large events that introduced the virus to the population (Minogue et al. 2019).

NGS can be used for diagnosis of infectious diseases. When other methods for a diagnosis of a leptospirosis case were unsuccessful, NGS was able to identify this disease. With the identification of the disease in a sample of spinal fluid from a

patient, an appropriate treatment plan could be developed. NGS is starting to be used in laboratory-developed tests (LDTs) to identify and detect diseases that are caused by microorganisms (Minogue et al. 2019).

Along with biosurveillance, NGS has applications within bioforensics. Bioforensics is the analysis of biological agents that are submitted as evidence for a case. NGS-based bioforensic applications include outbreak tracing and isolate characterization. NGS use in outbreak tracing can help to determine a source of the disease. The cholera outbreak in Haiti was found to be most likely caused by Nepal traveling to Haiti after the earthquake in 2010. NGS has also been used to identify genetically modified organisms (GMOs). One group was able to identify a modified bacterium in imported additives (Minogue et al. 2019).

NGS identified causality and traceability in a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* and determined how this organism was transmitted despite rigorous infection-control mechanisms and policies. Whole genome sequencing using NGS of the index and subsequent cases linked *K. pneumoniae* infections. The findings exhibited that the isolate was highly stable in a hospital environment which allowed for the bacteria to survive on the equipment used in the hospital. This was true even for the equipment that has been deep cleaned (Minogue et al. 2019).

Environmental Microbiology

Along with analyzing biofilms from teeth, joints and wounds, and pathogen surveillance, NGS has been used in environmental microbiology. NGS has enabled environmental studies of uncultured microbes using metagenomics and has been used to characterize the microbes in various environments. Samples including water, air, soil, and surfaces have been characterized with forensic DNA profiling applications.

NGS has been used to determine which bacterial genera are found in household kitchen and bathroom sinks which is applicable to crime scene location profiling. DNA was extracted from samples of biofilm accumulated in the drainpipes of the sinks. NGS was performed using Illumina MiSeq to sequence the DNA found in the drains. After analysis, it was determined that both the kitchen and bathroom had a great variety of bacteria present in the biofilm. The main phylum represented in both locations was Proteobacteria. In the kitchen, Bacteroides was the dominant genus and Flavobacteriaceae the most common bacterial family in the bathroom (Hemdan et al. 2019).

Characterizing microbial populations in a hospital environment is another example of environmental microbiology. NGS can be used for hospital surveillance to monitor and control infection and characterize the resistome (Comar et al. 2019). Surveillance reduces the occupational risk and aids in infection prevention (Giampaoli et al. 2018).

Another application of NGS has been the identification of biofilm communities in marine environments. A building up of biofilms and other organisms on structures

such as ships is referred to as "biofouling" (Rampadarath et al. 2017). Biofouling is not desirable for any method of maritime transport, as it increases fuel usage, cost, and maintenance of the vessel. In order to identify bacteria found in the biofilm, NGS was used. In Mauritius, fiberglass was utilized for biofilm accumulation (Rampadarath et al. 2017). After harvesting the bacterial community and extracting the DNA, the sequences were determined using Illumina technology and analyzed with the CLC Genomics workbench (Rampadarath et al. 2017). The resulting sequences represented 21 different bacterial groups, with the most common families including Acidithiobacillaceae and Actinomycetaceae. Determining which bacterial families are present in biofilms can be helpful to remediate watercraft to prevent fouling (Rampadarath et al. 2017). In an effort to reduce fouling on ships, antifouling coatings are put onto the body of the ship. There are two different types of coatings, biocidal and fouling-release (Winfield et al. 2018). NGS to identify biofilms has been used to evaluate the effectiveness of these coatings on preventing bacterial communities (Winfield et al. 2018). To test the effectiveness, the coatings were applied to wooden planks and placed into water at a marina (Winfield et al. 2018). The biofilm samples were then scraped off the wood and the DNA was extracted (Winfield et al. 2018). Both the 16S rRNA gene and the internal transcribed spacer (ITS) region were amplified prior to sequencing (Winfield et al. 2018). On the fouling-release coatings, there were 204 bacterial taxa identified, and 141 were identified on the biocidal coatings (Winfield et al. 2018). The decreased number of taxa present in the biofilm on biocidal coated wooden boards is most likely due to growth inhibition (Winfield et al. 2018). The biofilm constituents could be used to characterize the surface and past locations.

NGS using the Illumina platform was used to characterize the microbes on a mummy in the Slovak castle of Krásna Hôrkaand in Slovakia and the surrounding environment. The environmental sampling locations included the glass of the sarcophagus which contained the mummy, its inside air, and air from inside the chapel where the sarcophagus is located (Kraková et al. 2018).

A forensic application of environmental microbiology is geolocation (Giampaoli et al. 2018). A 2013 study of fungi using the 18S rRNA gene analyzed eleven samples collected from different soil environments including fields, forests, grasslands, and an urban park with different flora (Lilje et al. 2013). Nine GI matches resulted that were unique to the sampling area in their analysis of the NGS data (Lilje et al. 2013). In a NGS study of soils from farms and lakes that were fertilized or unfertilized and near and far from homes and a country road using an Illumina MiSeq, the proportion of DNA sequences from different taxa including metazoa, viridiplantae, protista, opisthokonta, fungi, chromista, and others varied in the six samples (Giampaoli et al. 2014). In another study, the fungal and bacterial microbiomes of indoor and outdoor environmental locations were found to differ (Hanson et al. 2016). In addition, indoor dust was found to be characterized by fungal and microbial communities that can vary (Weikl et al. 2016).

Forensic Biology

Eukaryotes like humans and animals are born with eukaryotic cells, but over time the skin surface, oral cavity, and gut are colonized by a tremendous diversity of bacteria, archaea, fungi, and viruses. In humans, the community formed by the complement of cells is called the human microbiome. The human microbiome is estimated to contain almost ten times as many cells as eukaryotic cells and weigh several pounds (Morgan and Huttenhower 2012). These communities differ among multicellular organisms. The microbiome also varies in human body fluids including urine, vaginal fluid, blood, and saliva; NGS technology employed in medical microbiology has been extended to human body fluid analysis (Elkins and Zeller 2022). Primarily, the 16S rRNA gene has been analyzed in the deep sequencing diversity studies (Elkins and Zeller 2022).

Advances in forensic biology have resulted in tremendous capabilities for human identification and solving crime. NGS can be used to generate investigative leads based upon the estimations of phenotype and biogeographical ancestry it can produce. Since 1993 when the first STR genetic loci were introduced for forensic DNA typing, DNA profiles have reflected differences in size, and the presence of variable numbers of the repeats in the population formed the basis of the method for human differentiation and identification (Butler 2005). Although the introduction of capillary electrophoresis and sequencing gels improved data analysis by allowing discrete sizing and single base resolution over the earliest approaches to DNA-based human identification, the DNA typing method masks sequence differences between individuals within the analyzed loci (Gettings et al. 2016). Gettings et al. (2016) sequenced 22 STR loci using NGS and reported upon these variations.

New commercial DNA typing kits are available for human identity testing and phenotype and biogeographical ancestry prediction using NGS. Upon obtaining the case sample, the DNA is extracted, quantified, and diluted or concentrated prior to library preparation. One such tool is the ForenSeq™ DNA Signature Prep Kit for library preparation that uses a modified Illumina MiSeq NGS instrument, the Verogen MiSeq FGx,® Forensic Genomics System for sequencing (Jäger et al. 2017). The kit has recently been evaluated, and developmental validation results have been published (Jäger et al. 2017). A single NGS run can yield DNA sequence information more than two hundred sites that were previously probed using dozens of independent assays. The ForenSeq™ kit provides data for a total of 231 DNA profiling markers including 7 X and 24 Y chromosome STRs, 27 autosomal STRs, 22 phenotyping single nucleotide polymorphisms (pSNPs), 56 ancestry informative SNPs (aiSNPs), and 94 identity informative SNPs (iiSNPs). Two SNPs are common to both ancestry and phenotype analysis (Jäger et al. 2017). Studies have reported the successful implementation of NGS and the ForenSeq™ kit for the analysis of human remains samples including hair, bone, and teeth (Edson 2019). The ForenSeq™ approach employs two PCR steps (Jäger et al. 2017). In the first, the targets are amplified and tagged (Jäger et al. 2017). In the second step, adaptors are added for the target to bind the flow cell in sequencing, and indexes are added to demultiplex the samples for analysis after the run (Jäger et al. 2017). The Verogen Universal

Analysis Software (UAS) demultiplexes the sequences and assigns them to each sample, presents the sequence data for each sample, identifies the traditional STR loci and SNP loci within the sequence data, and makes the STR number of repeat calls at each locus (Jäger et al. 2017) (Fig. 3).

Another NGS kit for forensic human identification applications is the Precision ID GlobalFiler™ NGS STR Panel v2 kit which can be used to prepare libraries for sequencing on the Thermo Fisher Ion Torrent series of instruments (Tao et al. 2019). The Precision ID kit simultaneously targets amelogenin, 31 autosomal STRs, and the Y-chromosome markers DYS391, SRY, and rs2032678 Y-indel (Tao et al. 2019). The Ion Chef™ robot can be used to prepare the libraries for sequencing, and the HID STR Genotyper Plugin can be used for sequence analysis. A third commercial NGS kit for human identification applications is the Promega PowerSeq™ system (van der Gaag et al. 2016). In a recent study, a population of 297 individuals from diverse groups from the Netherlands, Nepal, Bhutan, and Central African Pygmies were sequenced at 17 STRs and amelogenin using the prototype PowerSeq™ and sequenced on an Illumina MiSeq; concordance was found with the traditional CE approach (van der Gaag et al. 2016). The commercial options make NGS testing more accessible to forensic laboratories.

NGS DNA typing has been used to analyze human remains recovered from wars and massacres. Edson (2019) reported upon DNA typing of human remains of US service members lost in World War II, the Korean War, and wars in Southeast Asia using traditional DNA typing methods including AmpFlSTR® Yfiler™, AmpFlSTR® MiniFiler™, and PowerPlex® Fusion using capillary electrophoresis for fragment sizing and analysis and NGS using library preparation chemistry developed in house (Edson 2019). The libraries were sequenced on an Illumina MiSeq (Edson 2019). Elkins et al. (2021) have demonstrated the use of NGS using the ForenSeq™ kit for profiling historic remains. The team was able to predict the phenotype and biogeographical ancestry for five samples (Elkins et al. 2021).

In 2019, next-generation sequencing found its first use in an investigation resulting in an individual's conviction in a criminal case. The case, occurring in the Netherlands, was originally acquitted in 2016 until an appeal was permitted for the prosecutor. Electrophoresis technology used during the original trial yielded indeterminate results. The trial in 2019 allowed for the DNA to be tested using newer NGS technology. The MiSeq platform used for this case was able to generate a useful DNA profile. NGS aided in the conviction of the defendant (de Knijff 2020).

NGS applications in forensic biology also include differentiation of individuals who are monozygotic or identical twins using random variations or differences in DNA methylation due to environmental variation. After fertilization, a zygote can split in two, resulting in two embryos with the same genetic material. Because of

Fig. 3 Flow chart of NGS DNA profiling process

this, differentiation between monozygotic twins by STR fragment analysis is not possible as the twins present the same number of repeat genotype. Distinguishing these twins can be important when cases involving paternity tests and perpetrators at a crime scene arise. Due to the development of NGS, methods of detecting differences in identical twins have been discovered to aid in identification (Weber-Lehmann et al. 2014; Wang et al. 2015). When a suspect has an identical twin, it is necessary to eliminate the innocent one before a conviction. This was needed in a criminal case involving multiple rape charges and adultery (Yuan et al. 2020). Next-generation sequencing was used to sequence DNA from blood samples provided by the monozygotic twins. The NGS result identified an SNP at locus m.6903 for one of the individuals. This SNP was also detected in evidence from two of the crime scenes, so the perpetrator could be correctly identified (Yuan et al. 2020).

Along with cases involving crimes, next-generation sequencing can be used for paternity cases. In order to establish paternity, research has been performed using DNA in sperm cells of identical twins and in blood from one of the individual's children. Ultra-deep NGS was completed, and the sequences obtained were analyzed for somatic mutations using VarScan 2. Several SNPs were identified in the DNA of the child and the child's father. These SNPs were not found in the father's twin, showing that NGS can be useful in paternity testing applications to differentiate the DNA of monozygotic twins (Weber-Lehmann et al. 2014).

Mitochondrial DNA typing results can also provide investigative leads in cold cases as well as maternal ancestry information (Holland et al. 2018). Mitochondrial DNA profiling has been applied to ear cells which produced only a partial STR profile (Seo et al. 2002). Mitochondrial DNA profiling, owing to its high copy number, has proven useful for analyzing bone, teeth, and hair samples from human remains (Butler 2005). Mitochondrial DNA typing is used clinically to screen for human diseases such as diabetes and certain cancers and to screen for mutations in subpopulations of mitochondria. In 2010, NGS was used screen for the heteroplasmy, or the existence of multiple mtDNA types in an individual (Li et al. 2010).

The Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory has developed an approach for mtDNA NGS sequencing of DNA from degraded and skeletal samples using the Illumina MiSeq sequencer; haplotypes were determined by alignment (Marshall et al. 2017). The Nextera XT approach for mtDNA sequencing has been evaluated on control and non-probative case samples using a MiSeq (Peck et al. 2018). New commercial ForenSeq™ DNA typing kits have recently become available for mitochondrial testing in addition to autosomal DNA typing also using the MiSeq FGx instrument. One commercially available kit, the ForenSeq™ mtDNA Control Region Kit, is capable of providing sequencing information for the entire control region where the majority of the human identity information is found (Walichiewicz et al. 2019). The ForenSeq mtDNA consists of two primer sets of 125 primers overlapping by at least three base pairs that yield 18 short (<150 base pair) amplicons (Walichiewicz et al. 2019). A ForenSeq™ mtDNA whole genome kit has been released by Verogen to prepare libraries for sequencing the complete mitochondrial chromosome of samples of interest.

Mitochondrial DNA (mtDNA) genome sequencing has been used to explore variations in monozygotic twins. When compared to nuclear DNA, mtDNA has fewer mechanisms for correction of damaged DNA. This leads to more mutations within the mtDNA genome. In research involving six pairs of identical twins, it was found that four of the pairs exhibited an SNP at nucleotide 15,301. For each pair, one individual had a guanine at this position, and the other individual had an adenine base. Eleven point heteroplasmies (PHP) were also identified after sequencing in five of the groups. At the positions of the PHPs, heterogeneity of the bases existed at different levels (Wang et al. 2015).

NGS is also being used for genealogy by commercial companies for their 23andMe AncestryDNA® and My Heritage DNA Tests. DNA profiles of non-offenders are being used in searches to support forensic casework through the use of websites such as GEDmatch. GEDmatch was used to solve the case of the Golden State Killer (Selk 2018).

Plant Biology

NGS has found numerous applications in plant biology including screening and identifying specimens including food crops and seed stocks, diagnosing pesticide resistance, and characterizing illegal and legal plants used as drugs that may be analyzed in criminal investigations. NGS can be used to assess gene structure and expression, assign varietal SNPs, and perform transcriptome profiling (Berkman et al. 2012).

Although NGS technologies are improving, its use in plant biology is still very limited. A plant's heritable traits can be understood from its genome. Many cereal crops like rice, sorghum, and maize have been sequenced. Food crops are a potential target of bioterror events. Genomic sequencing has enhanced the ability to understand their physiology, with examples such as quantitative trait loci (QTL) analysis for genes associated with domestication and drought tolerance in rice, shoot fly resistance in sorghum, as well as disease resistance in maize (Berkman et al. 2012). These plants have a simple genome that is very easy to sequence using NGS technology (Berkman et al. 2012). More complex crops such as wheat are harder to sequence, but the same technologies used to sequence the cereal crops can be applied (Berkman et al. 2012).

Wheat has a very large and complex genome, estimated at 17 Gbp. The large size of the wheat genome is in part attributable to wheat being an allohexaploid, meaning that it contains three distinct diploid genomes that together function much like any diploid. NGS technologies are used to generate sequence tags that are expressed in genes which require no prior knowledge of a gene sequence for that organism. NGS can be applied for gene expression analysis, the structure of a loci, and sequence variation. An Affymetrix wheat GeneChip oligonucleotide array has been developed with over 61,127 probe sets representing 55,052 transcripts, enabling the generation of numerous high-quality gene expression data sets. NGS technologies also aid in the discovery of new SNPs which are the principal markers used in plant genetic

analysis. As in other fields, the large data sets from the Illumina sequencing platform provide the potential to discover very large numbers of genome wide SNPs. While wheat's large genome size and complexity hinder the use of NGS, advances both in sequencing technology and bioinformatics tools are making NGS applications in wheat increasingly feasible (Berkman et al. 2012).

In contrast to sequencing the plant itself, NGS has been used to probe and sequence the microbiomes of fungi found in controlled plants. Fungi were found in dispensary based *Cannabis* flowers from Amsterdam and Massachusetts. As the use of *Cannabis* has increased and now extends to more widespread medicinal use, there is a risk of illness due to pathogenic mold and bacteria. The internal transcribed spacer region 2 (ITS2) was sequenced, and toxigenic *Penicillium* and *Aspergillus* species were identified, including *P. citrinum* and *P. paxilli*, *P. commune*, *P. chrysogenum*, *P. corylophilum*, *A. terreus*, *A. niger*, *A. flavus*, *A. versicolor*, and *Eurotium repens*. In addition, the pathogenic species *Cryptococcus liquefaciens* was detected. DNA libraries were constructed using the New England Biolabs NEBNext Quick ligation module and the Illumina adaptor and sequenced on an Illumina MiSeq. The reads were aligned to reference sequences using MG-RAST and confirmed using One Codex (McKernan et al. 2015).

The opium poppy was recently sequenced by pyrosequencing technology to expand the availability of simple sequence repeat (SSR) genetic markers for the plant. 105 Mb of the genome were sequenced, and 166,724 contigs were assembled and reported increasing the genetic data for the plant sevenfold which is important for developing new diagnostic tools. The authors tested 53 genomic SSR markers against 37 opium poppy accessions and 7 *Papaver* species and found intermediate intraspecific polymorphism information content (PIC) values (Celik et al. 2014).

Evolutionary Biology

NGS has been used in applications in evolutionary biology including tracing human migration patterns and human adaptation by sequencing the mitochondrial DNA genome. The data support the global population deriving from two successive waves of human migration from Africa. The first migration from Africa was 85,000 years ago and saw movement to the Middle East and Asia, while the second migration was to Europe between 40,000 and 12,000 years ago. NGS is supplementing studies of fossils and detection of Y-chromosome and mitochondrial chromosome D-loop molecular markers using Sanger sequencing (Kundu and Ghosh 2015).

The MitoMap project maps the migration patterns using haplogroups and genetic human family tree with 10 Adams and 18 Eves (Kogelnik et al. 1996). With NGS, the non-recombining portion of the Y chromosome (NRY) data has grown exponentially. Whereas prior to NGS, the Y-chromosome phylogenetic tree was based on haplogroup assignment and consisted of a few hundred branches, it has grown in size and complexity to many thousands using NGS. New tools such as the Yleaf software tool perform effective NRY SNP calling and NRY haplogroup assignment from NGS data. (Ralf et al. 2018).

Conclusion

NGS technology has developed significantly over the last two decades. NGS has proven to be a versatile tool that can be applied to DNA analysis in diverse fields of study and for diverse research targets to solve many problems. NGS has been applied to determining human evolutionary patterns, the human microbiome, analysis of biofilms, variations in crop biology, human pathogen monitoring and surveillance, molecular determinants of human disease, cancer detection and profiling, geolocation, and sites for individualization in human, plant, and animal DNA forensics. The uses of NGS are only expected to grow in the future once the workflow is standardized and problems including large datasets, polyploid genomes, and lack of standardized methodology are overcome.

References

Berkman P, Lai K, Lorenc M, Edwards D (2012) Next-generation sequencing applications for wheat crop improvement. Am J Bot 99(2):365–371. https://doi.org/10.3732/ajb.1100309

Boycott KM, Dyment DA, Sawyer SL, Vanstone MR, Beaulieu CL (2014) Identification of genes for childhood heritable diseases. Annu Rev Med 65:19–31

Bruijns B, Tiggelaar R, Gardeniers H (2018) Massively parallel sequencing techniques for forensics: a review. Electrophoresis 39(21):2641–2654. https://doi.org/10.1002/elps.201800082

Butler J (2005) Forensic DNA typing, 2nd edn. Elsevier Academic Press, Burlington

Celik I, Gultekin V, Allmer J, Doganlar S, Frary A (2014) Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing. Mol Breed 34:323–334. https://doi.org/10.1007/s11032-014-0036-0

Comar M, D'Accolti M, Cason C, Soffritti I, Campisciano G, Lanzoni L, Bisi M, Volta A, Mazzacane S, Caselli E (2019) Introduction of NGS in environmental surveillance for healthcare-associated infection control. Microorganisms 7:708

de Knijff P (2020) Case study: how next generation sequencing resolved a difficult case, leading to the first criminal conviction of its kind. Verogen, 1–4. https://cdn2.hubspot.net/hubfs/6058606/Verogen-First-NGS-Court-Case-Study_Final_VD2019024_8.5x11-web.pdf?__hstc=238609695.bed74b81cf4041e42adad16833ab8584.1576870704888.1576870704888.1576870704888.1&__hssc=238609695.1.1576870704888. Accessed 27 Nov 2020

Edson SM (2019) Getting ahead: extraction of DNA from skeletonized cranial material and teeth. J Forensic Sci 64(6):1646–1657. https://doi.org/10.1111/1556-4029.14123

Elkins KM, Klavens A, Gorr K, Kollmann DD, Zeller CB (2021) Assessing the performance of next generation sequencing for determining sex, ancestry, and phenotypic characteristics of historic human remains. Forensic Science International: Genetics, submitted. (submitted May 31, 2021 so unfortunately this is still in review since it took us a long time to get submitted.)

Elkins KM, Zeller CB (2022) Next Generation Sequencing in Forensic Science: A Primer. Taylor & Francis/CRC Press, Boca Raton. https://www.routledge.com/Next-Generation-Sequencing-in-Forensic-Science-A-Primer/Elkins-Zeller/p/book/9781032072043. The paper was eventually published as a book chapter.

Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA, Guerrieri RA, Vallone PM (2016) Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Sci Int Genet 21:15–21. https://doi.org/10.1016/j.fsigen.2015.11.005

Giampaoli S, Berti A, Di Maggio RM, Pilli E, Valentini A, Valeriani F, Gianfranceschi G, Barni F, Ripani L, Spica VR (2014) The environmental biological signature: NGS profiling for forensic comparison of soils. Forensic Sci Int 240:41–47. https://doi.org/10.1016/j.forsciint.2014.02.028

Giampaoli S, Alessandrini F, Frajese GV, Guglielmi G, Tagliabracci A, Berti A (2018) Environmental microbiology: perspectives for legal and occupational medicine. Legal Med (Tokyo, Japan) 35:34–43. https://doi.org/10.1016/j.legalmed.2018.09.014

Gill P, Jeffreys A, Werrett D (1985) Forensic application of DNA 'fingerprints'. Nature 318:577–579. https://doi.org/10.1038/318577a0

Hanson B, Zhou Y, Bautista EJ, Urch B, Speck M, Silverman F, Muilenberg M, Phipatanakul W, Weinstock G, Sodergren E, Gold DR, Sordillo JE (2016) Characterization of the bacterial and fungal microbiome in indoor dust and outdoor air samples: a pilot study. Environ Sci Process Impacts 18(6):713–724. https://doi.org/10.1039/c5em00639b

Harris-Ricardo J, Fang F, Herrera-Herrera A, Fortich-Mesa N, Olier-Castillo D, Cavanzo-Rojas D, González-Quintero R (2019) Bacterial profile of the supragingival dental biofilm in children with deciduous and early mixed dentition using next generation sequencing (HOMINGS) technique. Enfermadades Infecciosas Microbiología Clínica 37(7):448–453. https://doi.org/10.1016/j.eimce.2019.04.005

Hemdan B, El-Liethy MA, ElMahdy MEI, El-Taweel GE (2019) Metagenomics analysis of bacterial structure communities within natural biofilm. Heliyon 5(8). https://doi.org/10.1016/j.heliyon.2019.e02271

Holland MM, Makova KD, McElhoe JA (2018) Deep-coverage MPS analysis of heteroplasmic variants within the mtGenome allows for frequent differentiation of maternal relatives. Genes 9 (3):124. https://doi.org/10.3390/genes9030124

Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G, Bruand J, Schlesinger F, Pond SJK, Varlaro J, Stephens KM, Holt CL (2017) Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. Forensic Sci Int Genet 28:52–70. https://doi.org/10.1016/j.fsigen.2017.01.011

Joo H-S, Otto M (2012) Molecular basis of in vivo biofilm formation by bacterial pathogens. Chem Biol 19(12):1503–1513. https://doi.org/10.1016/j.chembiol.2012.10.022

Kircher M, Kelso J (2010) High-throughput DNA sequencing-concepts and limitations. BioEssays 2(6):524–536. https://doi.org/10.1002/bies.200900181

Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC (1996) MITOMAP: a human mitochondrial genome database. Nucleic Acids Res 24(1):177–179. https://doi.org/10.1093/nar/24.1.177

König LM, Klopfleisch R, Höper D, Gruber AD (2014) Next generation sequencing analysis of biofilms from three dogs with postoperative surgical site infection. Int Scholarly Res Notices 2014:1–5. https://doi.org/10.1155/2014/282971

Kraková L, Šoltys K, Puškárová A, Bučková M, Jeszeová L, Kucharík M, Budiš J, Orovčík L, Szemes T, Pangallo D (2018) The microbiomes of a XVIII century mummy from the castle of Krásna Hôrka (Slovakia) and its surrounding environment. Environ Microbiol 20:3294–3308. https://doi.org/10.1111/1462-2920.14312

Kundu S, Ghosh SK (2015) Trend of different molecular markers in the last decades for studying human migrations. Gene 556(2):81–90. https://doi.org/10.1016/j.gene.2014.12.023

Li M, Schönberg A, Schaefer M, Nasidze I, Stoneking M (2010) Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. Am J Hum Genet 87 (2):237–249. https://doi.org/10.1016/j.ajhg.2010.07.014

Lih CJ, Harrington RD, Sims DJ, Harper KN, Bouk CH, Datta V, Yau J, Singh RR, Routbort MJ, Luthra R, Patel KP, Mantha GS, Krishnamurthy S, Ronski K, Walther Z, Finberg KE, Canosa S, Robinson H, Raymond A, Le LP, . . . Williams PM (2017) Analytical validation of the next-generation sequencing assay for a nationwide signal-finding clinical trial: molecular analysis for therapy choice clinical trial. J Mol Diagn 19(2):313–327. https://doi.org/10.1016/j.jmoldx.2016.10.007

Lilje L, Lillsaar T, Rätsep R, Simm J, Aaspõllu A (2013) Soil sample metagenome NGS data management for forensic investigation. Forensic Sci Int Genet Suppl Ser 4:e35–e36. https://doi.org/10.1016/j.fsigss.2013.10.017

Liu D, Zhang J, Wu B, Liu F, Ye S, Wang H, Lv J, Weng X, Chen Y, Han W, Chen J (2020) Impact of donor lung colonized bacteria detected by next-generation sequencing on early post-

transplant outcomes in lung transplant recipients. BMC Infect Dis 20:689. https://doi.org/10.1186/s12879-020-05393-w

Marshall C, Sturk-Andreaggi K, Daniels-Higginbotham J, Oliver RS, Barritt-Ross S, McMahon TP (2017) Performance evaluation of a mitogenome capture and Illumina sequencing protocol using non-probative, case-type skeletal samples: implications for the use of a positive control in a next-generation sequencing procedure. Forensic Sci Int Genet 31:198–206

McKernan K, Spangler J, Zhang L, Tadigotla V, Helbert Y, Foss T, Smith D (2015) *Cannabis* microbiome sequencing reveals several mycotoxic fungi native to dispensary grade *Cannabis* flowers. F1000Research 4:1422. https://doi.org/10.12688/f1000research.7507.2

Minogue TD, Koehler JW, Stefan CP, Conrad TA (2019) Next-generation sequencing for biodefense: biothreat detection, forensics, and the clinic. Clin Chem 65:383–392. https://doi.org/10.1373/clinchem.2016.266536

Morgan XC, Huttenhower C (2012) Chapter 12: Human microbiome analysis. PLoS Comput Biol 8 (12):e1002808. https://doi.org/10.1371/journal.pcbi.1002808

Peck MA, Sturk-Andreaggi K, Thomas JT, Oliver RS, Barritt-Ross S, Marshall C (2018) Developmental validation of a Nextera XT mitogenome Illumina MiSeq sequencing method for high-quality samples. Forensic Sci Int Genet 34:25–36. https://doi.org/10.1016/j.fsigen.2018.01.004

Ralf A, Montiel González D, Zhong K, Kayser M (2018) Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. Mol Biol Evol 35(5):1291–1294. https://doi.org/10.1093/molbev/msy032

Rampadarath S, Bandhoa K, Puchooa D, Jeewon R, Bal S (2017) Early bacterial film colonizers in the coastal waters of Mauritius. Electron J Biotechnol 29:13–21. https://doi.org/10.1016/j.ejbt.2017.06.006

Sánchez R, Ayala R, Martínez-López J (2019) Minimal residual disease monitoring with next-generation sequencing methodologies in hematological malignancies. Int J Mol Sci 20(11):2832. https://doi.org/10.3390/ijms20112832

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M (1977) Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265(5596):687–695

Schwarzberg K, Le R, Bharti B, Lindsay S, Casaburi G, Salvatore F, Saber MH, Alonaizan F, Slots J, Gottlieb RA, Caporaso JG, Kelley ST (2014) The personal human oral microbiome obscures the effects of treatment on periodontal disease. PLoS One 9(1):e86708. https://doi.org/10.1371/journal.pone.0086708

Selk A (2018) The ingenious and 'dystopian' DNA technique police used to hunt the 'Golden State Killer' suspect. The Washington Post. https://www.washingtonpost.com/news/true-crime/wp/2018/04/27/golden-state-killer-dna-website-gedmatch-was-used-to-identify-joseph-deangelo-as-suspect-police-say/. Accessed 25 Nov 2020

Seo Y, Uchiyama T, Matsuda H, Shimizu K, Takami Y, Nakayama T, Takahama K (2002) Mitochondrial DNA and STR typing of matter adhering to an earphone. J Forensic Sci 47(3):605–608

Tao R, Qi W, Chen C, Zhang J, Yang Z, Song W, Zhang S, Li C (2019) Pilot study for forensic evaluations of the Precision ID GlobalFiler™ NGS STR Panel v2 with the Ion S5™ system. Forensic Sci Int Genet 43:102147. https://doi.org/10.1016/j.fsigen.2019.102147

Torchia MT, Amakiri I, Werth P, Moschetti W (2020) Characterization of native knee microorganisms using next-generation sequencing in patients undergoing primary total knee arthroplasty. Knee 27(3):1113–1119. https://doi.org/10.1016/j.knee.2019.12.013

van der Gaag KJ, de Leeuw RH, Hoogenboom J, Patel J, Storts DR, Laros J, de Knijff P (2016) Massively parallel sequencing of short tandem repeats-population data and mixture analysis results for the PowerSeq™ system. Forensic Sci Int Genet 24:86–96. https://doi.org/10.1016/j.fsigen.2016.05.016

Verma D, Garg PK, Dubey AK (2018) Insights into the human oral microbiome. Arch Microbiol 200:525–540. https://doi.org/10.1007/s00203-018-1505-3

Walichiewicz P, Eagles J, Daulo A, Didier M, Edwards C, Fleming K, Han Y, Hill T, Li S, Rensfield A, Sa D, Husbands J, Holt C, Stephens K (2019). Performance evaluation of the

ForenSeq mtDNA control region solution. ISHI. https://verogen.com/wp-content/uploads/2019/11/Mito-ISHI-Poster_final.pdf

Wang Z, Zhang S, Bian Y, Li C (2015) Differentiating between monozygotic twins in forensics through next generation mtGenome sequencing. Forensic Sci Int Genet Suppl Ser 5:e58–e59. https://doi.org/10.1016/j.fsigss.2015.09.023

Weber-Lehmann J, Schilling E, Gradl G, Richter DC, Wiehler J, Rolf B (2014) Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. Forensic Sci Int Genet 9:42–46. https://doi.org/10.1016/j.fsigen.2013.10.015

Weikl F, Tischer C, Probst AJ, Heinrich J, Markevych I, Jochner S, Pritsch K (2016) Fungal and bacterial communities in indoor dust follow different environmental determinants. PLoS One 11 (4):e0154131. https://doi.org/10.1371/journal.pone.0154131

Winfield M, Downer A, Longyear J, Dale M, Barker G (2018) Comparative study of biofilm formation on biocidal antifouling and fouling-release coatings using next-generation DNA sequencing. Biofouling 34(4):464–477

Yuan L, Chen X, Liu Q, Song A, Zhang S, Wu Y, Bao G, Wei G, Lu J, Liu Z (2020) Identification of the perpetrator among identical twins using next-generation sequencing technology: a case report. Forensic Sci Int Genet 44:102167. https://doi.org/10.1016/j.fsigen.2019.102167

Using Multiple Chromosomal Marker Analysis Tools, for DNA Profiling in Human Identification: New, Evolving and Productive Approach

22

Rajendra V. E. Chilukuri

Contents

Abstract

DNA genotyping, which started in the 1990s to identify convicts, has become the most sought-after area in forensics. Human Identification (HID) by analyzing the fragment lengths of specific regions in DNA, such as Short Tandem Repeats (STRs), using Capillary Electrophoresis (CE) based method has changed the way forensic cases are resolved. What started with just four STR markers in the 1990s, analyses based on CE had seen many innovations, with now multiple markers systems used including autosomal STRs, Y- STRs, X- STR, and SNPs. With the use of multiple STR markers, case analysis and resolution improved tremendously, scientist now can process DNA that is highly degraded, minimal, contaminated, or decades old, with increased positive outcomes. With analysis of Short Nucleotide Polymorphisms (SNPs), it is now possible to estimate ancestry, predict phenotype, and determine relationships with confidence. Advanced systems are now available to resolve cases, with co-amplification of multiple markers, and multiple combination systems like (Autosomal + Sex Chromosomal markers or Autosomal + SNP markers or Autosomal + Sex Chromosomal + SNP

R. V. E. Chilukuri (✉)
SyberGENE Inc, San Jose, CA, USA
e-mail: cverajendra@global-forensics.com; cverajendra@gmail.com

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_64

Markers together) to analyze DNA. These systems promise to make analysis efficient, specific, faster, and cost effective. Implementation of Next Generation Sequencing in forensics, and high-throughput sequence data obtained from NGS makes analysis of haplotypes, isometric alleles, heteroplasmy, mixture samples, kinship, mono-zygotic twin analysis, visible traits, geographic ancestry, identity, all possible in a single run. In this chapter we discuss novel approaches taken in genotyping, the use of systems developed in multiplexing, co-amplification, and multiple marker interrogation in single analysis setup.

Keywords

Short Tandem Repeats · Single Nucleotide Polymorphism · Human Identification · Next Generation Sequencing

Introduction

STRs (short tandem repeat) are DNA repeat units, they are usually understood to be two to six nucleotide repeats, these repeats can be anywhere between six to several dozens in lengths, comprise of about 3% of the human genome (Lander et al. 2001), and vary between individuals. These STRs are interspersed in the human genome widely, albeit thought to be inactive with respect to gene expression. A large percentage of STRs are interspersed in the non-coding regions, it is believed that about 8% of them are located in the coding regions (MA et al. 2015). STRs have been instrumental in various other aspects such as having an effect of gene expression and in Human Identification (HID) when it comes to forensics (Sawaya et al. 2013; Chen et al. 2016). The use of STRs in forensic DNA analysis was initiated in the 1990s, starting with barely few markers for DNA profiling the numbers were further increased, with major countries or unions having their own measure of STR markers required for establishing DNA profiles (JM et al. 2004). The nomenclature of STRs is unique and it gives the details of the chromosomal specificity and the unique number associated with the STR, an example, D21S11 is given the name as D indicated DNA, 21 is the chromosome number on which the STR is located, S indicates that it is an STR and 11 is the unique number issued to the STR, most of the STRs are given similar addresses.

The advent of use of STR analysis in DNA profiling has started a wave of identifying the unique STR markers that will help in discrimination and human identification. Multiple verities of DNA markers are available in DNA profiling, however, using STR is the preferred option by almost all the laboratories now. The presence in abundance of microsatellite loci in the human genome was recognized by their characterization and by studies demonstrated successfully (Sawaya et al. 2013). The polymorphicity of STR loci makes them the candidate markers for commercially developing kits by various manufacturers, with more and more studies, research and development, manufacturers were able to develop kits that offer

laboratories options to generate powerful discriminatory profiles. These kits have become the main stay in human identification for laboratories.

Various novel approaches have been discussed in this chapter that have focused on development and use of combination kits on forensic samples and on population studies to build forensically relevant data. Use of kits with a combination of autosomal STRs and sex chromosome STRs has been a revelation, primarily in sexual assault cases and in paternity cases. Use of Y-chromosomal STR-based kits in combination with autosomal kits is quite effective and yields more information and removes the need for multiple tests and trying analyses with individual kits (Liu et al. 2017, 2020b), such combination kits will be of great application and a consorted effort by both the forensic laboratories-based research and manufacturers can help such systems, which help speed up the HID process.

The Evolution and Diversity STRs Markers for Human Identification

The set of markers that are evaluated and needed to conclude a case vary between laboratories, as DNA profiles across the world are generated by various public and private laboratories, for purposes such as human identity testing, DNA databasing, casework related to forensics, Identification of missing individuals, Identification of victims in mass disaster, for paternity and kinship analysis. The guidelines followed in carrying out these analyses have been primarily same and are developed by laboratories predominantly as per the Scientific Working Group on DNA Analysis Methods (SWGDAM). SWDGAM known as TWGDAM until the year 1999, works on a mandate "The responsibilities of SWGDAM are: (1) to recommend revisions, as necessary, to the Quality Assurance Standards for Forensic DNA Testing Laboratories and the Quality Assurance Standards for DNA Databasing Laboratories; (2) to serve as a forum to discuss, share, and evaluate forensic biology methods, protocols, training, and research to enhance forensic biology services; and (3) to recommend and conduct research to develop and/or validate forensic biology methods" (www.swagdam.com). The STR marker uniqueness has promoted them to be used in DNA profiling, the first advent of their usage, with three markers selected in 1994, by the Forensic Science Service (FSS, UK), had the quadruplex amplification system comprising of STRs: TH01, vWA, FES/FPS, and F13A1 (C et al. 1994; Kimpton et al. 1994). As these markers were eventually found to be not enough to generate credible profiles, in 1997 Federal Bureau of Investigation (FBI) had decided to include 13 autosomal STR loci to the list (MR et al. 1995; Budowle 1998) (Fig. 1). They have been enlisted as the core markers in the Combined DNA Index System (CODIS). CODIS is database that has been developed and comprises DNA profiles contributed by local, state, and federal forensic laboratories. The markers that were enlisted in the FSS that is, vWA and TH01 were also enlisted in the CODIS core set, FES/FPS and F13A01 were dropped from entry into CODIS owing to their

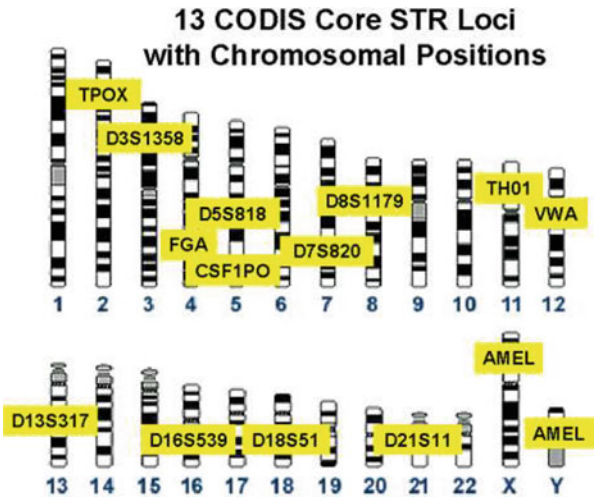


Fig. 1 The 13 core CODIS STR loci and their location on the autosomal chromosomes

Table 1 Includes the 13 core loci that were considered from October 1998 until December 31, 2016 and the 20 core loci that are now effective from January 1, 2017

CODIS loci up till 2016	CODIS loci listed from 2017
CSF1PO	CSF1PO
FGA	FGA
THO1	THO1
TPOX	TPOX
VWA	VWA
D3S1358	D3S1358
D5S818	D5S818
D7S820	D7S820
D8S1179	D8S1179
D13S317	D13S317
D16S539	D16S539
D18S51	D18S51
D21S11	D21S11
	D1S1656
	D2S441
	D2S1338
	D10S1248
	D12S391
	D19S433
	D22S1045

Fig. 1 Source: https://strbase.nist.gov/fbicore.htm

lower levels of polymorphism (CL et al. 2002). Further, the CODIS core marker list was upgraded to include seven more additional STRs in the year 2010, post their evaluation in the HID they were implemented for analysis since 2017, making it 20 core STR loci (Ge et al. 2012) (Table 1). The STR core loci used for genotyping vary, and are based on sufficiency of the accuracy in obtaining proper profiles for case works. The list of various agencies and the core loci is listed in Table 2.

Table 2 The list of core STR loci that are listed for DNA profiling by the US, Europe, UK, Germany, and Interpol. Amelogenin for sex determination is included as addition for all the listed core loci

U.S.A core loci	European Standard Set (ESS), Includes addition loci	United Kingdom core loci (Including SGM Plus Kit) as per DNA-17	Germany	Interpol
CSF1PO FGA THO1 TPOX VWA D3S1358 D5S818 D7S820 D8S1179 D13S317 D16S539 D18S51 D21S11 D1S1656 D2S441 D2S1338 D10S1248 D12S391 D19S433 D22S1045	FGA TH01 VWA D1S1656 D2S441 D3S1358 D8S1179 D10S1248 D12S391 D18S51 D21S11 D22S1045 D2S1338 D16S539 D19S433 SE33	D18S51 D21S11 TH01 D3S1358 D16S539 D2S1338 D1S1656 D10S1248 FGA D8S1179 vWA D22S1045 D19S433 D12S391 D2S441 SE33.	FGA TH01 SE33 VWA D3S1358 D8S1179 D18S51 D21S11	FGA TH01 VWA, D3S1358 D8S1179 D18S51 D21S11

The Sex Chromosomal STRs in Human Genotyping

The Y- STR: The human Y Chromosome with segments of it specific to males are considerably used in forensic DNA analysis (Roewer 2019). The Y- STR analysis comes quite handy when the routine autosomal DNA profiling turns inconclusive. A Y-chromosomal analysis is implemented for sexual assault cases predominantly (Roewer 2019; Zhang et al. 2020). The Y-Chromosomal Haplotype STRs are used to analyze paternal lineages of unidentified male trace donors, especially when the samples are a mixture of male DNA, as observed in sexual assault offences. The various benefits of Y- STR haplotyping applied in forensic analysis ensure (1) exclusion of males as suspects from crimes not committed by them, (2) help in identifying lineage of male perpetrators, (3) determine the involvement of multiple male perpetrators in a crime, and (4) generate leads in cases where the male suspect in unknown. Y- STR analysis is also a way to determine paternity in disputes involved with male offspring and various other kinship analysis (Zhang et al. 2020; Marth et al. 1999) such as including legal heir cases, missing person cases, and identification of disaster victims. Y Chromosome polymorphism is also credible in predicting bio-geographic ancestry paternally where trace and unknow male DNA is available and also in cases where the autosomal DNA genotyping is inconclusive (Roewer 2019; Marth et al. 1999) (Table 3).

Table 3 The list of Y and X chromosomal STRs that are listed suitable for DNA profiling

Y- STR markers widely used in forensic DNA analysis	X-XTR markers used in forensic DNA analysis (12 of the X- STR listed on strbase-b.nist.gov)
DYS19, DYS385a/b,S389I, DYS389II, DYS390, DYS391, DYS392,DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y-GATA-H4, DYS481, DYS533, DYS549, DYS570, DYS576, DYS643, DYS449, DYS460, DYS518, DYS627, DYF387S1a/b, DYS526a/b, DYS547, DYS612, DYS626, DYF399S1, DYF403S1a/b, DYF404S1.	DXS10074, DXS10079, DXS10101, DXS10103, DXS10134, DXS10135, DXS10146, DXS10148, DXS7132, DXS7423, DXS8378, HPRTB

Source: https://strbase-b.nist.gov/

X- STRs: The X Chromosome short-tandem repeats (X- STRs) analysis albeit less implemented has significant usage, in cases where the male sample is unavailable, yet the kinship has to be established, the X- STRs analysis gives significant leads (Butler 2012a). Also, the X- STR analysis is appreciably applied in population genetic studies, establishing the uniqueness and significance of the markers in a given population (R 2007). The efficiency of X- STR application is enhanced in complex cases when female cousins and or grandparent has to be analyzed to generate investigative leads, in cases of paternity testing with the involvement of close relatives, and in cases of maternity testing (Li et al. 2015). It can be well considered in the current field of forensics, where at least one female is involved use of X chromosomal STRs will reap benefit (Edwards et al. 1992). In cases with missing persons, assault incest, federal immigration issues, inclusion of the X- STR marker system is a credible way in maximizing the forensic arsenal with regard to STR kits (R 2007; Wyner et al. 2020). A list of commercially available X-STR kit as listed on the strbase-b.nist.gov is given in Table 3.

The Role of SNPs in Human Identification

Single Nucleotide Polymorphisms (SNPs), which comprise a single base alteration wherein an individual may have an “A” (Adenine), and a different individual may have “G” (Guanine) at a given locus. Similar to STRs, SNPs are highly effective tools in human identification (Marth et al. 1999; Kwok 2003). The microhaplotypes (multiple SNPs in a single DNA sequence) can compete with an STR locus in developing discrimination data and yet be less challenging to analyze by sequencing (Sobrino and Carracedo 2005). Some of the forensic biological samples are not of quality and are tough to obtain meaningful DNA profiles post analysis; often times these samples contain very little template DNA and are mostly degraded. Owing to these conditions technologies that would still help generate fruitful outcomes have been mooted. These solutions are sought to make possible genotyping from tough to work with samples, carried out primarily by focusing on the hypervariable regions

located in the control region of the mitochondrial DNA and thus by sequencing this region or by sequencing the whole human mitochondrial genome (MR et al. 1995). However, the mitochondrial DNA (mtDNA) typing as done in the present day by sequencing is considered laborious and time taking, the option is also economically not viable for some laboratories, leaving only few laboratories to choose mtDNA forensic analysis. It is also known that mtDNA delivers lesser power of discrimination as opposed to STR-based analysis (B et al. 2003). Single Nucleotide Polymorphisms (SNPs) are base substitutions, insertions, or deletions that occur at single positions in the genome of any organism (Kwok 2003; Børsting et al. 2014). The major advantage with using SNPs for analysis is them being smaller in length with sizes ranging somewhere from 60 to 80 bp (Kwok 2003; Butler 2012b). It is understood that approximately 85% of human variations are obtained from SNPs (Xing et al. 2009). SNPs offer great deal of hope for forensic DNA analysis as they are abundant in number, and with hands free solution available, their use can generate efficient data (Sobrino and Carracedo 2005; Butler 2012b). Single Nucleotide Polymorphisms usage is advantageous in situations such as disasters and in cases such as fire accidents, with highly degraded bodies and in cases of missing persons, in situations where the DNA is considerably degraded (B et al. 2003; Freire-Aradas et al. 2012). The degraded DNA is understandably smaller in fragments and is not suitable for STR-based analysis (Gettings et al. 2015). Another advantage is the low mutation rate observed with SNPs, this ensures stability of SNPs and hence are great for lineage-based investigations, inheritance-related cases and in missing person cases, and in cases with scenarios wherein direct DNA is not accessible (L et al. 2012; Budowle and van Daal 2018). Since single nucleotide polymorphism is biallelic at times they are considered less preferred in human identification (HID) (Budowle and van Daal 2018). SNP-based analysis is more favorable for studies where information is generated for Identity testing, with requirement of higher heterozygosity and minimal inbreeding coefficient. It is understood that for identity testing SNPs are believed to have highest heterozygosity, close to 50% heterozygosity for a biallelic system and have relatively and preferably low Fst values. This is so because it would mean that fewer SNPs are required to obtain higher level of discrimination power. Also, the use of SNPs would mean lesser population data sets are required for statistical assessments of forensic casework (Budowle and van Daal 2018). For lineage studies where the SNPs act as haplotype indicators for kinship studies, ancestry studies, and phenotypic studies, SNPs are believed to be better tools in establishing phenotypic characteristics like skin, hair, or eye color which help in serving as leads in cold cases or where visual witnesses are unavailable (L et al. 2012; Rahim et al. 2008; Lin and Musunuru 2018).

Systems with STR Markers and Their Variations

Ever since the first testing tool was developed and evaluated on the four markers, TH01, VWA, FES/FPS, and F13A1 by Kipton et al., the evolution of STR kits has been quite dynamic. With the addition of more markers to the process of evolution,

human identification has taken a turn for the better, with respect to achieving more elaborate, efficient, and accurate way of genotyping, with enlisting markers that were forensically relevant, the system was implemented by FSS in the beginning (Kimpton et al. 1994; Moretti et al. 2001; JM et al. 2004; Butler 2006). The use of capillary electrophoresis (CE) based genotyping in the areas of forensic analysis was first established in 1990s, it was used in the earlier days of application in toxicology studies, identification of chemical residues in gun shots, and so on. Initial works with STRs applied silver stained polyacrylamide gels for detection (Edwards et al. 1991), further fluorescence-based detection methods with gel electrophoresis were implemented (Goodwin et al. 2011). The advent of the presently considered capillary electrophoresis mode of detection with ABI instruments has revolutionized forensic DNA profiling (JM et al. 2004). Innovation being brought in with upgraded instrumentation has also prompted in developing reagent kits with STRs and reagent composition to enhance and achieve better profiling data and provide laboratories with options to use specific kits for specific case studies. As listed on the Strbase there are close to 60 variants of kits developed by different commercial manufacturers that are used in human DNA profiling (Shrivastava et al. 2018), albeit some of the kits are modified versions of the previous once by the manufacturers, they were developed only to help obtain better data as declared than their previous version. Most of these kits interrogate for only one specific type of variable being analyzed, be it autosomal STR, Y- STRs, X- STRs, mtDNA SNPs. For users it becomes inevitable to use different kits for different types of markers being analyzed. In complex cases such as for paternity, sexual assault, missing person, and disaster victims' cases there would be a need for multiple kits being used, this costs the laboratory considerably in terms of budget and also and the time invested. In some of the cases obtained data is not sufficient enough.

The list of major kits presently available are primarily to analyze autosomal STRs; some of the kits are designed to analyze clean samples, whereas others are modified to interrogate degraded samples, (e.g., Global filer vs Minifiler) these kits are further modified with inclusion or exclusion of STR markers based on their fragment length to facilitate for analysis of degraded samples. An abridged list is given in Table 4 indicating such kits, which are commercially available and predominantly used in most of the forensic laboratories. These kits are also designed to accommodate the dye chemistries feasible to be used on the available instrumentation at the laboratories.

The kits listed in Table 4 are predominantly used on major instruments and are validated under SWGDAM guidelines. The kits as indicated in the table are of various dye chemistries either five or six as listed, allowing for multiplexing, the kits offer laboratories with capabilities in generating HID data. Multiplexing is defined as using multiple dyes or tags and is a characteristic of image-based assays; multiplexing facilitates a way to report on multiple components (STRs in this case) of the cell simultaneously and eventually their influences on one another.

By implementing STR-based analysis with these reagents, a great number of cases can be resolved with either exclusion, which can be analyzed with various genetic mismatches between alleged father and child, or by inclusion due to high

Table 4 A list of commercially available STR kits that are commonly used for human identification at different laboratories. The listed above are just a few of many kits available for applications. The listed kits are used for analysis of single type pf markers, either autosomal. Y or X- STR only

A list of STR kits commonly used in laboratories				
SNo	Product Name	Manufacturer	No.of arkers	No of dyes
	For routine case work samples			
1	Power Plex Fusion System (Equivalent to SF25)	Promega	24	5 dye
2	PowerPlex Fusion 6C System (6 dye)	Promega	27	6 dye
3	AB Global Filer	Thermo	24	6 dye
4	AB Global Filer Express	Thermo	24	6 dye
5	VeriFiler™ Express PCR Amplification Kit	Thermo	24	6 dye
6	Investigator 24plex QS Kit	Qiagen	24	6 dye
7	VersaPlex™ 27PY System	Promega	**27**	6 dye
	For degraded samples			
6	AmpFLSTR™ Identifiler™ Plus (5 dye)	Thermo	15	5 dye
9	Investigator ESSplex SE QS Kit	Qiagen	16	5 dye
10	Investigator HDplex and Triplex Kits	Qiagen	12	5 dye
11	PowerPlex® ESX 16 and ESI 16 Fast Systems	Promega	16	5 dye
	Y chromosomal markers			
12	Yfiler Plus PCR Amplification Kit	Thermo	27	6 dye
13	PowerPlex Y23 System (5 dye system)	Promega	23	5 dye
14	Investigator Argus 12 Y	Qiagen	12	5 dye
	X chromosomal markers			
15	Investigator Argus X-12 QS	Qiagen	12	5 dye

likelihood of paternity. With kits like Y-Filer or Argus 12Y used for Y-chromosomal STR marker analysis, used in cases of a kinship analysis, relating to paternal lineage established the significance of such kits. The use of such kits in developing areas such as genetic geneaology. Wherein relatives of the male involved can be tested even with greater generation gap. The X Chromosome analysis kits such as the Argus 12 X hold prominence in sorting cases wherein the suspected father is not accessible for testing, in such cases grandmother sample can be used to determine relationship of father with child.

Use of Combination Kits

An appreciation to the amount of analysis required in certain scenarios can be perceived by going through cases like "A case of false mother included with 46 autosomal STR markers" (Li et al. 2015). The study was carried out to resolve the case using practically 7 different systems, 46 autosomal STRs were analyzed by

using 4 different systems, 20 X-chromosomal STRs, 40 SNP loci, and mitochondrial DNA (mtDNA) were applied to resolve the case. The study revealed that the putative mother and the boy had at least one allele common between them for the 46 tested autosomal STR loci, an X- STR profile, and the SNP markers analysis demonstrated variable genotypes at 13 X- STR loci and 5 SNP loci, this data helped to rule out the putative mother to be the actual mother. Noticeable differences in the mitochondrial profiles excluded the mother as a parent of the son because they have multiple differences. It was finally found that the putative mother is the sister of the biological father. Many such cases have established the effort and the need for the use of multiple kits to resolve cases, this not only indicates the need for having access to various resources, which one might not have, it also forces laboratories to fall short of critical forensic samples to run multiple analyses, and is not cost effective either. In the wake of these hindrances, usage of systems with combination analysis capabilities might be a better solution.

Although various solutions are available to get to the bottom of human identification, at the best possible way, innovations are still on to develop kits that can be multiplexed to analyze DNA samples. Majority of these efforts are to enhance the quality of information obtained in resolving cases, the speed at which the analysis can be delivered, and in being cost effective simultaneously. Use of combination of kits has been proved to be very effective and as observed in studies by Josephine Purps et al., adding a combination of another marker in analysis has enhanced the percentage of obtaining quality information by 21%, this was achieved by an analysis of the data by running profiles with autosomal STR kits (Purps et al. 2015). The said study analyzed around 2077 stains from sexual assault cases (287 cases), with the use of a total of 39 markers, which included 23Y- STR markers and 16 of autosomal markers. The results demonstrated that multiple male contributors were detected with thrice more accuracy, with added Y-chromosomal profiling, than just analyzing by applying autosomal STR profiling (Roewer 2019), the study proposed that almost one tenth of the cases that have remained inconclusive due to Y- STR analysis being not carried out would have had conclusive outcomes and proposed a broader barker analysis. This approach is a standard method however as demonstrated by many other studies, and well-reviewed (KN and M 2012; Jian et al. 2021).

With need-based innovations, driving the researchers, novel multiplexing systems were being developed, giving more informative datum, that was efficient and effective. Conventional kits have autosomal markers analyzed, with few Y- STR markers included, to have an access to sex determination and an insight into male component of DNA in the forensic sample. However, these minimal markers are not sufficient to give discrimination in most of the sexual assault cases or in cases involving multiple males and mixture samples. To overcome such road blocks and forced need of multiple analyses to be carried out, kits were being developed to deliver Y- STR information along with the autosomal marker analysis. One such study was reported by Shuanglin Li et al., which focused on the validation of the

system. The reagent system was developed to analyze 20 autosomal STRs and 4Y STRs, the study indicated that the system is quite efficient in resolving various types of forensic samples, including routine and also samples with low concentrations of female: male DNA mixtures (W et al. 2017). With the success of such kits developed in co-amplification and multiplexing, expansion in marker number was carried out further, as noticed with the kit developed which included capabilities to co-analyze for 10Y STRs and 15 autosomal STRs (Liu et al. 2017).

A similar system the "HomyGene19 + 14Y" was developed as a kit for HID analysis, the system had capabilities to analyze 18 autosomal STRs and 14 Y chromosomal STRs, validation studies on the kit demonstrated combined match probability of the kit at $2.39 \times 10–29$, this study showed that the kit would serve as a "robust, reliable, highly polymorphic, and informative forensic kit" (Dong et al. 2015). Another such system was reported wherein 19 autosomal STRs and 27Y STRs were co-analyzed with the use of a single kit (Liu et al. 2020a). The study reported that all the markers were amplified with great degree of robustness and with high specificity, appreciable tolerance to inhibitors, and with capability to generate DNA profiles from ambient concentration of DNA as obtained from forensics samples. These kits have added phenomenal opportunities by not only helping in solving tough cases, but also performing cost-effective population studies. Reports published to deliver information of population have demonstrated use of another co-amplification combination kit, the markers analyzed in the same were 16 autosomal (15 autosomal +1Amelogenin) and 19Y STRs (HM et al. 2017). The study analyzed the Guangdong Liannan Yao population in Chinese province to develop a genomic portrait of the same, analyzing the genetic polymorphism and evolutionary differentiation. The interesting observations made in this study describe here as one such example. A comparative study on the population exposed a relationship between the She ethnic population and the Liannan Yao population as compared to any other Chinese populations. A deeper Y- STR analysis revealed the presence of 102 unique haplotypes, 87 of the 102 were observed only in the Liannan Yao only once. A statistical pairwise Rst analysis with a multidimensional scaling plot revealed that the Liannan Yao inhabitants are closely linked to the Fujian She ethnic population. This study has established that the Liannan Yao population is unique within the Chinese population in their ethnicity and origin. The use of multiplex kits of this nature provides great deal of opportunities in human identification and population genetic studies (HM et al. 2017).

The combination kits that include autosomal STRs markers and the sex chromosome markers-based kits or a combination of autosomal STRs marker kits with SNPs or all the three types can serve a great bit in forensic studies as demonstrated above. In similar lines are the kits developed for application on Next Generation Sequencing-based DNA profiling in human identification.

Studies using combination of STR-SNP markers also gained prominence, systems developed for use on NGS platforms seem to have a very obvious future. One of the seminal studies on application of STR-SNP combination system performed

using CE was carried out by Tian Wei et al. (T et al. 2018). The study detailed the development and application of the system. The design of this initial kit was detailed as eight SNP-STR compound markers developed including rs11222421-D11S4463, rs12423685-D12ATA63, rs2325399-D6S1043, rs1276598-D6S474, rs16887642-D7S820, rs9531308-D13S317, rs188010-D17S974, and rs258112-D5S2800. As a part of the designed combination, six out of the eight SNP-STRs combinations were indicated to be loci on different chromosomes, the remaining two, rs1276598-D6S474 and rs2325399-D6S1043, were determined to be on the same chromosomal arm, 6q to be precise, but their physical and genetic distances were ensured to be far apart, so as to not overlap during analysis. This multiplex primer set was developed as a 5-dye chemistry, as described in the report the two allele specific primers for the SNPs for each of the SNP-STR group were labeled with 6-FAM and HEX or with TAMARA and ROX florescent dye sets at the 5′ end of the primers for CE-based detection. The set of eight SNP-STRs markers and the sex determination marker Amelogenin were prepared by allelic ladder size range and were attenuated to the remaining four dyes in order to develop the desired multiplex system, for detection by CE. Experimental conditions were optimized accordingly post primer designing and fluorescent labeling, a multiplex PCR system was developed with 9-plex fluorescent labeling. The eight SNP-STRs combination markers were observed to be amplified as desired (T et al. 2018). This study has shown that the SNP-STR haplotype can be detected fruitfully in the simulated imbalanced two DNA mixture, at a dilution 1:20. The forensic efficiency of this SNP-STRs combination was found to be higher as opposed to routine analysis of autosomal STRs. The study suggested that the SNP-STR markers would facilitate forensic analysts with credible tools, in analysis of DNA mixtures of any gender or any cellular source.

Studies that followed in using combination set of markers as reported by Qingzhen Zhang et al., looked at obtaining better outcomes using multiple kits. The study used multiple marker analysis-based kits and efficiently demonstrated that the use of 40 STRs and the 91 SNPs in combination had better outcome as opposed to using individual marker kit-based analysis that is, either 27 STRs and 91 SNPs or just the 40 STRs alone. The study also demonstrated that sensitivity as well as specificity of the combination 40 + 91 marker has yielded complete profile in full-sibling testing, with strong power to distinguish second-degree relatives from unrelated pairs, what is interesting is that the study implemented the use of CE-based kits along with the markers from the NGS system-based kits (Zhang et al. 2020).

Similar studies were also reported where a combination of SNPs and STRs analyzed together gave a much better outcome, as samples were analyzed, the study showed that when analyzed using mixture DNA, minor contributors were detected at dilutions of 1:500 which were not feasible at conventional approach on CE. The study also demonstrated that the likelihood ratio has substantially jumped from being at 2.86×103, when using STR kits alone to 7.14×107 when SNP-STR combination analysis was carried (Jian et al. 2021). The combination

of primers used here to co-amplify the targets has helped achieve better results, an indicative that combination analysis has yields more if designed properly, and in fact enhances the data obtained than individual marker analysis. Also, such studies are likely to save crucial time, sample shortage, and are also understandably cost effective.

Next Generation Sequencing-Based DNA Profiling Systems in Forensics

As understood, analysis of forensic samples on Capillary Electrophoresis (CE) yields information that is purely length based, termed as fragment analysis, indicated by the length of the STRs markers being analyzed (JM et al. 2004; Butler 1998). The difference and similarities in these fragment lengths between individuals help include or exclude as individuals being the contributors to the genetic material obtained from the scene of crime or in a given population (Tornar et al. 1994). For long although CE has served as gold standard but, only 60% of cases are brought to conclusion using this technology, The limitation of CE-based analysis ranges from lack of dimension to interrogate multiple genetic polymorphisms in one single reaction setup, also the resolution of markers in CE-based genotyping is a limitation, it is well perceived that CE does not deliver sufficient information with degraded samples, resulting in loss of useful genomic information, there are no reported and available manufactured reagents for mtDNA analysis on CE, and customized studies have shown that the data obtained is of low resolution. Very high discrepancies were observed in CE-based mixture DNA analysis and low resolution on such samples, rendering the study inconclusive (Chong and Syn 2021). While implementing CE-based genotyping, alleles of similar length but different sequences cannot be resolved and would lead to aberrant interpretations. Also, mutations observed in complex kinship cases mostly failed to be resolved using CE-based application (Cowell et al. 2015). The advent of Massively Parallel sequencing changed the face of genomics quite a bit, with numerous studies that were driven, even drugs and therapeutics are being developed with data obtained from Next Generation Sequencing, and the technology has become a revolution since its dawn. NGS has become a boon to biotechnology and biomedical research, pharmaceutical research, in personalized medicine (Irwin et al. 2011; Qin 2019; Giannopoulou et al. 2019; Morganti et al. 2020) (Fig. 3). The first sequencing-based forensic analysis was carried out on Swedish population, by analysing ten forensically significant marker using pyrosequencing based technology (Divne et al. 2010). Next Generation sequencing role in forensic research has become an inevitable reality (Fig. 2), with initial work on applying the technology in forensics has demonstrated the use of two platforms: the 454 GS Junior system and the Illumina Genome Analyzer, this study established that NGS technology is credible for application in genomic DNA as well

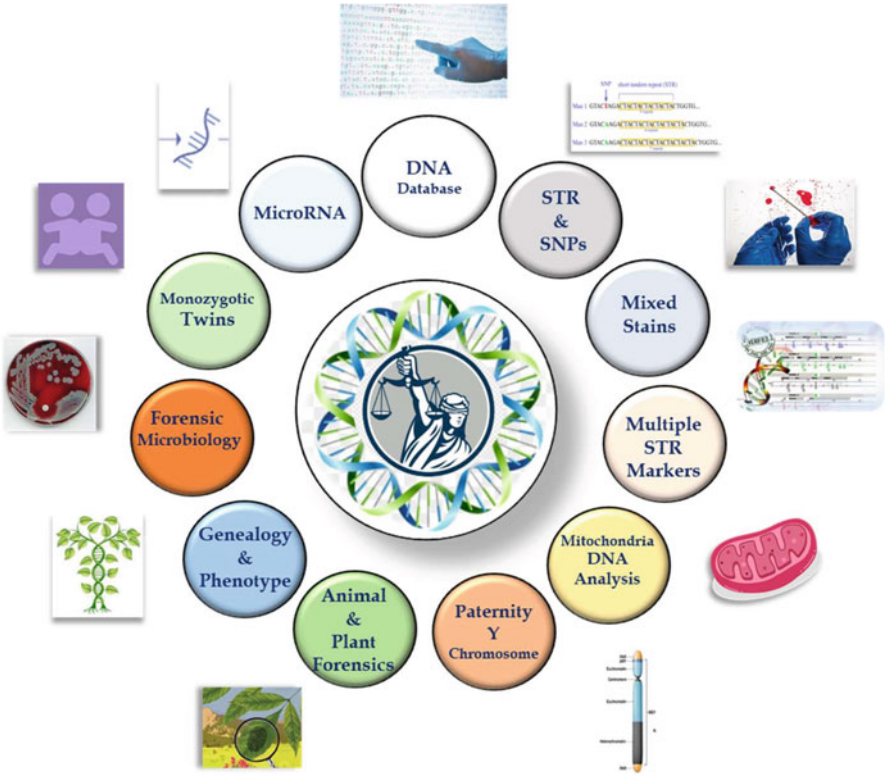


Fig. 2 The schematic representation of various areas of forensics that NGS can be applied and in elaborating the possibilities accurate and high-end data generation sequencing

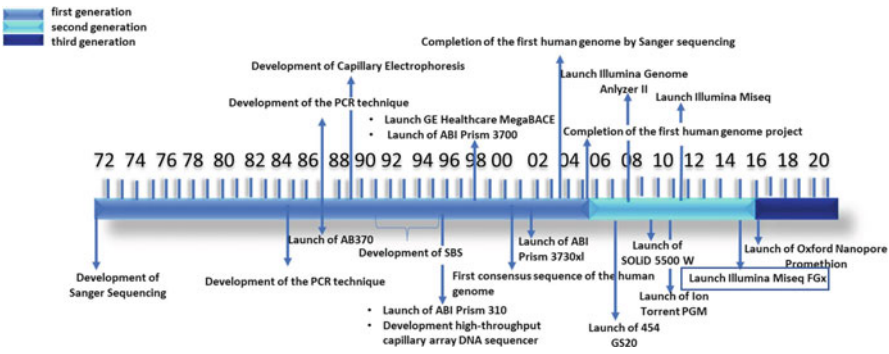


Fig. 3 The time line and the evolution of Next Generation Sequencing, and its eventual application into forensics (Box: with MiseqFGx developed specifically for forensic genomic application by illumina)

as mitochondrial DNA analysis, to obtain deep high throughput data for forensics (Figs. 3) (Irwin et al. 2011; Just et al. 2015; Bruijns et al. 2018). With implementation of the technology as demonstrated by works of Daniel M Bornman et al., analysis for the 13 CODIS STR markers as well as Amelogenin could accurately help in profiling for mixture as well as individual samples. This study has demonstrated that barely 18,500 reads were sufficient to genotype and to obtain the entire information with greater than 99% confidence for the analyzed 13 loci, the study also focused on the role of SNPs in significantly resolving the complication with mixture samples (Bornman et al. 2012).

Further with various platforms evolved that are now being implemented at various laboratories, the MiseqFGx, an NGS platform developed by Illumina^{R} is one of the first NGS platforms dedicated for forensic and human identification work flows, which is predominantly used and Ion Torrent which is promoted by Applied Biosystems (Jäger et al. 2017; Guo et al. 2017; Faccinetto et al. 2019). The advantage of using multiple marker analysis is widely explored with the NGS platform as well (England and Harbison 2020). As SNPs are efficient in biogeographical ancestry (BGA) prediction obtained by analysis of unknown crime scene samples, some critical SNPs can be analyzed for externally visible Traits (EVT) such as skin and eye color (England and Harbison 2020). The extreme possibility of multiplexing enhanced by the availability of sequence data renders this system as most versatile that is available, it is understood that such analysis as it covers all the forensically related markers in a single reaction makes it cost effective too (Ballard et al. 2020).

As understood the CE-based approach predominantly can be said as delivering qualitative data as in analog characteristics namely, peak color, size, shape, and height (Butler 1998; Tornar et al. 1994). NGS systems deliver precise digital data with regard to detailed read counts and also the important sequence information. With the capabilities of having digital data power with the settings in NGS, the sensitivity parameters can be tuned for any analysis there by increasing or decreasing coverage, this gives leverage to have unlimited dynamic range (Bruijns et al. 2018). The ability to generate digital read counts and high throughput sequencing renders high sensitivity for quantitative applications, these positive aspects help in scenarios such as minor DNA contributor detection in complex mixtures, that most often seem to be missed out on CE due to generation of partial profiles. As observed when performing heteroplasmy analysis on mtDNA, NGS high throughput sequencing can detect frequencies of minor variants at approximately 1% of the major as opposed to frequencies of the minor variants detected on CE at $>$10–20% (Just et al. 2015). With NGS as the samples are barcoded (wherein Indexes are attached to the samples giving them unique address as they are amplified) it enables multiplexing of the libraries, and facilitates for scalable throughput to degrees that are not feasible with CE-based approach (Tu et al. 2012).

With the use of multiple marker co-amplification, the high throughput nature and the capabilities to run multiple samples make NGS a much more dynamic system. The approach to sequencing and the capabilities to process samples at gigabytes of

data accommodate for deep throughput analysis and yield data that is much versatile than what is obtained on the CE systems. With NGS being used for forensic analysis quite a few exciting studies were reported to demonstrate the potential, one such is the sexting of an Egyptian mummy head, that is 4000 years old, and is suggested as a way of recovering the samples from degraded and damaged forensic specimens (Loreille et al. 2018). Study put forth the opinion that high throughput sequencing can be used to develop a way to process and analyze degraded DNA, by choosing mtDNA vs Nuclear DNA sequencing, the biological sexing of the mummy was carried out using shotgun sequencing and by application of the RX and RY method of sex determination. The study with variable data obtained from the nuclear vs mtDNA eventually with the information obtained has determined the sex of the mummy to be male (Loreille et al. 2018). The use of mtDNA analysis on NGS has also been established since long, studies from 2014 by King et al., have demonstrated the application of NGS (JL et al. 2014; Seo et al. 2015). The findings demonstrated that haplotypes were generated at a high depth of coverage showing minimal strand bias. The study established that the variations found in the mitochondrial genome sequence predominantly are observed within the coding region as opposed to the non-coding region. As the Haplotype, haplogroup variances were described with regard to the whole mtGenome and HVI/HVII. It was observed that an increased haplotype assessment, genetic diversity, credible random match probability, and also a better haplogroup assignment were achieved using MPS of the mtGenome, these observations enforced the application of MPS technologies in forensics.

With Next Generation Sequencing taking an extensive lead into the field of human Identification new strategies are being developed to revisit cold cases and drive them to conclusion. Extensive profile analyses are being carried out using public genetic genealogy databases like GED match to link to cold cases and potentially identify convicts. “Genetic genealogy is genetic relationship study performing DNA tests, by carrying out DNA profiling and testing in amalgamation with conventional genealogical methods, to build biological relationships between individuals” (Greytak et al. 2019). Such studies are only possible with high throughput data and deep end sequencing information available. With multiple marker analysis data from NGS-based forensic DNA analysis available, matching the same rightfully with public genomic data available could solve many cases. The same approach was taken to solve the case of the famous golden state killer case in USA, the approach was, with the analysis of crime scene DNA data that provides a partial match, forensic genealogists use that information to develop a family tree and zero in of determining the possible suspect (Wickenheiser 2019). Once the potential suspect is guesstimated from NGS data, with added advantage of autosomal STR + Y- STR + X- STR + iSNPs+ giSNPs+ phenotypic SNPs the next approach is to strategically acquire the suspect’s DNA directly and confirm the match. It is understood that Forensic Genetic Genealogy DNA analysis (FGGS) link was implemented in almost 200 cases solely in the year 2018 (Tillmar et al. 2021) The use of SNPs analysis from mtDNA and genomic DNA yields in-depth information, a new system developed approach uses almost >10,000 SNP markers to drive kinship analysis

from whole genome sequencing analysis (VEROGEN Inc., Kintelligence Kit), starting from a set of around 153,000 SNP markers post through filtration, the kit was developed to analyze over 10,000 SNPs that have capabilities to link kinship to even second- and third-generation siblings, such an analysis is impossible for data generated from CE. This approach can help in resolving cases such as those of missing persons, long pending violent crimes, and innocence projects, that help include or exclude people as possible biological sample contributors.

Thus, Next Generation Sequencing with versatile systems and solutions is making its stake as the next big approach in forensic DNA profiling and human identification. The systems developed to run on various dynamic NGS platforms are helping forensic biologists to get to the bottom of cases that hitherto were unresolved, thus opening up to a new horizon of DNA genotyping.

Conclusion

The advancements in forensic have been rapid to the point of identifying the markers that can help in discrimination and further adding on more markers to the list to enhance the process of human identification. With great impetus coming onto DNA-based forensic analysis in closing most of the cases, it has become imperative to carry out research on and developing reagent systems that would deliver the needful.

Restriction Fragment Length Polymorphism and Variable Number Tandem Repeats (VNTRs) were the modes of HID prior to STR application (RFLP) (Shiono 1996; Laber et al. 1994; Hochmeister et al. 1991; Blanco and Blanco 2017; Marwal et al. 2014; Panneerchelvam and Norazmi 2003). STR marker-based DNA analysis has been the mainstay in HID, extensive work was put in to develop and improve upon STR-based approaches since the advent of their use. With a humble but a very promising beginning with four markers STR-based DNA genotyping has seen advancement to almost becoming indispensable and the most reliable approach in HID. The active measures taken by various investigative and judicial systems have ensured development of enlisted markers such as those accepted by FBI, included in CODIS, and by European Network of Forensic Science Institutes (ENFSI), included in ESS, have been a guideline that most laboratories have adopted in STR-based DNA genotyping. These markers have been further picked by various manufacturers who have developed systems with great sensitivity and specificity in generating DNA profiles. Further, the manufacturers have also put in great deal of effort in creating systems that will help generate profiles from tough to work with samples such as degraded tissue, bones, and other such evidences.

Use of STR markers such as autosomal markers and X and Y sex chromosome-based markers has equipped forensic scientists with tools, to have diversity in divulging the DNA details of case work samples and deliver better outcomes. With increasing dependency on the DNA-based evidence data for homicides, sexual assault, burglary, missing person identification, determination of parentage in paternity and kinship cases, which began to substantiate eviction and conviction to a great

detail, laboratories have been burdened. This burdening of laboratories with increase in need of evidence analysis in-turn has increased costs. Using analysis systems that can interrogate for multiple types of markers such as combination kits can be a big time saver and will bring down costs on forensic laboratories.

Capillary Electrophoresis techniques have been the mainstay for two decades and above in forensics. It is well understood that the major chunk of cases are not resolved due to lack of sufficient information being available via CE-based approach. CE-based approach fails to give details on samples that may be isometric and yet have similar fragment length but have different sequences. Also, fragment analysis misses the mutations observed, more so if they are single base mutations, this adds to the drawback. Degraded samples are the primary kind of samples, that comprise the chunk of case work samples sent to laboratories. With CE-based approach a great deal of information is lost due to limitation in the sensitivity of the technology. Also, with CE-based approach sample concentration needed becomes a major limiting factor. Next Generation Sequencing based forensic analysis is a one-off alternative that is now available for laboratories to implement in overcoming the above-mentioned drawbacks with CE. Apart from obtaining sequence data that will take care of the Isometric allele aspect, the mutations can also be analyzed in samples. NGS-based forensic analysis being high throughput gives high coverage of the markers analyzed.

The robustness of the NGS technologies accommodates for use of combination marker analysis in a single run, the steps involved with normalization covers for the difference in sample concentrations to a great extent, thus ensuring unbiased read counts, based on sample availability. NGS offers great advantage in kinship and paternity analysis with availability of mitochondrial DNA analysis with ease. With systems available for mtDNA control region and whole genome SNP sequencing available, that are forensically relevant, NGS is the best available option. The NGS-based kinship analysis with large number of SNPs that can be analyzed, it is the only available approach to obtain links between the primary reference individual with not just immediate family members but also from second and third cousins too. A slow transition into applying NGS for DNA profiling along with CE-based approach definitely is the future of DNA forensics and will help cover up the loop holes from CE and will alleviate the stature in the way DNA-based human identification is carried out.

References

B B, MW A, MR W, R C (2003) Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 4:119–141. https://doi.org/10.1146/ANNUREV.GENOM.4.070802.110352

Ballard D, Winkler-Galicki J, Wesoły J (2020) Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. Int J Legal Med 1344(134):1291–1303. https://doi.org/10.1007/S00414-020-02294-0

Blanco A, Blanco G (2017) The genetic information (I). Med Biochem:465–492. https://doi.org/10.1016/B978-0-12-803550-4.00021-5

Bornman DM, Hester ME, Schuetter JM, Kasoji MD, Minard-Smith A, Barden CA, Nelson SC, Godbold GD, Baker CH, Yang B, Walther JE, Tornes IE, Yan PS, Rodriguez B, Bundschuh R, Dickens ML, Young BA, Faith SA (2012) Short-read, high-throughput sequencing technology for STR genotyping. Biotech Rapid Dispatch 2012:1

Børsting C, Pereira V, Andersen JD, Morling N (2014) Single nucleotide polymorphism. Wiley Encycl Forensic Sci:1–18. https://doi.org/10.1002/9780470061589.FSA1113

Bruijns B, Tiggelaar R, Gardeniers H (2018) Massively parallel sequencing techniques for forensics: a review. Electrophoresis 39:2642–2654. https://doi.org/10.1002/ELPS.201800082

Budowle B (1998) CODIS and PCR-based short tandem repeat loci: law enforcement tools. Promega

Budowle B, van Daal A (2018) Forensically relevant SNP classes. Biotechniques 44:603–610. https://doi.org/10.2144/000112806

Butler JM (1998) The use of capillary electrophoresis in genotyping STR loci. Methods Mol Biol 98:279–289. https://doi.org/10.1385/0-89603-443-7:279

Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. J Forensic Sci 51:253–265. https://doi.org/10.1111/J.1556-4029.2006.00046.X

Butler JM (2012a) X-chromosome analysis. Adv Top Forensic DNA Typing:457–472. https://doi.org/10.1016/B978-0-12-374513-2.00015-4

Butler JM (2012b) Single nucleotide polymorphisms and applications. Adv Top Forensic DNA Typing:347–369. https://doi.org/10.1016/B978-0-12-374513-2.00012-9

C K, D F, S W, M A, A U, J L, P G (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Legal Med 106:302–311. https://doi.org/10.1007/BF01224776

Chen HY, Ma SL, Huang W, Ji L, Leung VHK, Jiang H, Yao X, NLS T (2016) The mechanism of transactivation regulation due to polymorphic short tandem repeats (STRs) using IGF1 promoter as a model. Sci Reports 61(6):1–10. https://doi.org/10.1038/srep38225

Chong KWY, Syn CK-C (2021) Uncertainty in estimating the number of contributors from simulated DNA mixture profiles, with and without allele dropout, from Chinese, Malay, Indian, and Caucasian ethnic populations. Sci Reports 111 11:1–9. https://doi.org/10.1038/s41598-021-84580-4

CL H, M B, JM W, T N, KD L, PS W (2002) TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. J Forensic Sci 47:66–96

Cowell RG, Graversen T, Lauritzen SL, Mortera J (2015) Analysis of forensic DNA mixtures with artefacts. J R Stat Soc Ser C (Appl Stat) 64:1–48. https://doi.org/10.1111/RSSC.12071

Divne AM, Edlund H, Allen M (2010) Forensic analysis of autosomal STR markers using pyrosequencing. Forensic Sci Int Genet 4:122–129. https://doi.org/10.1016/J.FSIGEN.2009.07.004

Dong Y, Lin SS, Cao Y, Wu WW, Huang SQ, Zheng WG, Li FY, Ge BW, Guo YL, Zhou HG (2015) Development of a forensic multiplex amplification STR kit for 15 autosomal STR loci and 10 Y-STR loci. J Forensic Med 31. https://doi.org/10.3969/J.ISSN.1004-5619.2015.05.011

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253. https://doi.org/10.1016/0888-7543(92)90371-X

England R, Harbison S (2020) A review of the method and validation of the MiSeq FGx™ forensic genomics solution. WIREs Forensic Sci 2. https://doi.org/10.1002/WFS2.1351

Faccinetto C, Serventi P, Staiti N, Gentile F, Marino A (2019) Internal validation study of the next generation sequencing of Globalfiler™ PCR amplification kit for the ion torrent S5 sequencer. Forensic Sci Int Genet Suppl Ser 7:336–338. https://doi.org/10.1016/J.FSIGSS.2019.10.002

Freire-Aradas A, Fondevila M, Kriegel AK, Phillips C, Gill P, Prieto L, Schneider PM, Carracedo A, Lareu MV (2012) A new SNP assay for identification of highly degraded

human DNA. Forensic Sci Int Genet 6:341–349. https://doi.org/10.1016/J.FSIGEN.2011.07.010

Ge J, Eisenberg A, Budowle B (2012) Developing criteria and data to determine best options for expanding the core CODIS loci. Investig Genet 3:1. https://doi.org/10.1186/2041-2223-3-1

Gettings KB, Kiesler KM, Vallone PM (2015) Performance of a next generation sequencing SNP assay on degraded DNA. Forensic Sci Int Genet 19:1–9. https://doi.org/10.1016/J.FSIGEN.2015.04.010

Giannopoulou E, Katsila T, Mitropoulou C, Tsermpini E-E, Patrinos GP (2019) Integrating next-generation sequencing in the clinical pharmacogenomics workflow. Front Pharmacol 10:384. https://doi.org/10.3389/FPHAR.2019.00384

Goodwin W (William H), Linacre A, Hadi S (2011) An introduction to forensic. Genetics:198

Greytak EM, Moore CC, Armentrout SL (2019) Genetic genealogy for cold case and active investigations. Forensic Sci Int 299:103–113. https://doi.org/10.1016/J.FORSCIINT.2019.03.039

Guo F, Yu J, Zhang L, Li J (2017) Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ forensic genomics system. Forensic Sci Int Genet 31:135–148

HM G, C W, SY H, SH S, DJ X, YS W, CT L, MX Z (2017) Analysis of the 19 Y-STR and 16 X-STR loci system in the Han population of Shandong province. China Genet Mol Res 16. https://doi.org/10.4238/GMR16019573

Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, Dirnhofer R (1991) Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. J Forensic Sci 36:13189J

Irwin J, Just R, Scheible M, Loreille O (2011) Assessing the potential of next generation sequencing technologies for missing persons identification efforts. Forensic Sci Int Genet Suppl Ser 3:e447–e448. https://doi.org/10.1016/J.FSIGSS.2011.09.085

Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G, Bruand J, Schlesinger F, Pond SJK, Varlaro J, Stephens KM, Holt CL (2017) Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. Forensic Sci Int Genet 28:52–70. https://doi.org/10.1016/J.FSIGEN.2017.01.011

Jian H, Wang L, Lv M, Tan Y, Zhang R, Qu S, Wang J, Zha L, Zhang L, Liang W (2021) A novel SNP-STR system based on a capillary electrophoresis platform. Front Genet 12:96. https://doi.org/10.3389/FGENE.2021.636821

JL K, BL L, NM N, M S, SB S, X Z, DH W, CP D, W P, A S, B B (2014) High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Sci Int Genet 12:128–135. https://doi.org/10.1016/J.FSIGEN.2014.06.001

JM B, E B, F C, BR M (2004) Forensic DNA typing by capillary electrophoresis using the ABI prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 25:1397–1412. https://doi.org/10.1002/ELPS.200305822

Just RS, Irwin JA, Parson W (2015) Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. Forensic Sci Int Genet 18:131–139. https://doi.org/10.1016/J.FSIGEN.2015.05.003

Kimpton C, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, Gill P (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Legal Med 1066(106):302–311. https://doi.org/10.1007/BF01224776

KN B, M K (2012) Additional Y-STRs in forensics: why, which, and when. Forensic Sci Rev 24:63–78

Kwok P-Y (2003) Methods for genotyping single nucleotide polymorphisms. Annu Rev Genomics Hum Genet 2:235–258. https://doi.org/10.1146/ANNUREV.GENOM.2.1.235

L L, L W, QC F, YL Z, XL C, XD K, YM H, WF W, ZS Z (2012) A study on paternity testing with 96 autosomal SNPs. Zhonghua Yi Xue Yi Chuan Xue Za Zhi 29:28–33. https://doi.org/10.3760/CMA.J.ISSN.1003-9406.2012.01.008

Laber TL, Giese SA, Iverson JT, Liberty JA (1994) Validation studies on the forensic analysis of restriction fragment length polymorphism (RFLP) on LE agarose gels without ethidium bromide: effects of contaminants, sunlight, and the electrophoresis of varying quantities of deoxyribonucleic acid (DNA). J Forensic Sci 39:13649J

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Hong ML, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, De La Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Patrinos A, Morgan MJ (2001) Initial sequencing and analysis of the human genome. Nat 4096822(409):860–921. https://doi.org/10.1038/35057062

Li L, Lin Y, Liu Y, Zhu R, Zhao Z, Que T (2015) A case of false mother included with 46 autosomal STR markers. Investig Genet 6. https://doi.org/10.1186/S13323-015-0026-Y

Lin J, Musunuru K (2018) From genotype to phenotype. Circ Genomic Precis Med 11:1946. https://doi.org/10.1161/CIRCGEN.117.001946

Liu Y-S, Chen J-G, Mei T, Guo Y-X, Meng H-T, Li J-F, Wei Y-Y, Jin X-Y, Zhu B-F, Zhang L-P (2017) Genetic variation and forensic characteristic analysis of 25 STRs of a novel fluorescence co-amplification system in Chinese southern Shaanxi Han population. Oncotarget 8:55443. https://doi.org/10.18632/ONCOTARGET.19317

Liu F, Jia F, Sun F, Zhao B, Shen H (2020a) Validation of a multiplex amplification system of 19 autosomal STRs and 27 Y-STRs. Forensic Sci Res 5:292. https://doi.org/10.1080/20961790.2019.1665158

Liu F, Jia F, Sun F, Zhao B, Shen H (2020b) Validation of a multiplex amplification system of 19 autosomal STRs and 27 Y-STRs. Forensic Sci Res 5:292. https://doi.org/10.1080/20961790.2019.1665158

Loreille O, Ratnayake S, Bazinet AL, Stockwell TB, Sommer DD, Rohland N, Mallick S, Johnson PLF, Skoglund P, Onorato AJ, Bergman NH, Reich D, Irwin JA (2018) Biological sexing of a 4000-year-old egyptian mummy head to assess the potential of nuclear DNA recovery from the most damaged and limited forensic specimens. Genes 9:135. https://doi.org/10.3390/GENES9030135

MA B, E O, JS H-H (2015) Repetitive DNA in eukaryotic genomes. Chromosom Res 23:415–420. https://doi.org/10.1007/S10577-015-9499-Z

Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitziel NO, Hillier L, Kwok P-Y, Gish WR (1999) A general approach to single-nucleotide polymorphism discovery. Nat Genet 234(23): 452–456. https://doi.org/10.1038/70570

Marwal A, Sahu AK, Gaur RK (2014) Molecular markers: tool for genetic analysis. Anim Biotechnol Model Discov Transl:289–305. https://doi.org/10.1016/B978-0-12-416002-6.00016-X

Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B (2001) Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci 46: 15018J. https://doi.org/10.1520/JFS15018J

Morganti S, Tarantino P, Ferraro E, D'Amico P, Viale G, Trapani D, Duso BA, Curigliano G (2020) Role of next-generation sequencing technologies in personalized medicine. eHealth Agenda Heal Technol Futur P5:125–154. https://doi.org/10.1007/978-3-030-27994-3_8

MR W, JA D, D P, J R, B B (1995) Validation of mitochondrial DNA sequencing for forensic casework analysis. Int J Legal Med 108:68–74. https://doi.org/10.1007/BF01369907

Panneerchelvam S, Norazmi MN (2003) Forensic DNA profiling and database. Malays J Med Sci 10:20

Purps J, Geppert M, Nagy M, Roewer L (2015) Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases. Forensic Sci Int Genet 19:238–242. https://doi.org/10.1016/J.FSIGEN.2015.08.002

Qin D (2019) Next-generation sequencing and its clinical application. Cancer Biol Med 16(4): 10.20892/J.ISSN.2095-3941.2018.0055

R S (2007) X Chromosome markers: past, present and future. Forensic Sci Int Genet 1:93–99. https://doi.org/10.1016/J.FSIGEN.2007.03.003

Rahim NG, Harismendy O, Topol EJ, Frazer KA (2008) Genetic determinants of phenotypic diversity in humans. Genome Biol 9:215. https://doi.org/10.1186/GB-2008-9-4-215

Roewer L (2019) Y Chromosome short tandem repeats in forensics – sexing, profiling, and matching male DNA. WIREs Forensic Sci 1:e1336. https://doi.org/10.1002/wfs2.1336

Sawaya S, Bagshaw A, Buschiazzo E, Kumar P, Chowdhury S, Black MA, Gemmell N (2013) Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. PLoS One 8:e54710. https://doi.org/10.1371/JOURNAL.PONE.0054710

Seo SB, Zeng X, King JL, Larue BL, Assidi M, Al-Qahtani MH, Sajantila A, Budowle B (2015) Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform ion torrent™ PGM™. BMC Genomics 16. https://doi.org/10.1186/1471-2164-16-S1-S4

Shiono H (1996) Personal identification using DNA polymorphism the identification of forensic biological materials. Japanese J Leg Med 50:320–330

Shrivastava P, Dash HR, Kumawat RK, Srivastava A, Imam J (2018) STR typing and available kits. DNA Fingerprinting Adv Futur Endeavors:61–74. https://doi.org/10.1007/978-981-13-1583-1_4

Sobrino B, Carracedo A (2005) SNP typing in forensic genetics: a review. Methods Mol Biol:297

T W, F L, Y W, C P, C X, D H (2018) A novel multiplex assay of SNP-STR markers for forensic purpose. PLoS One 13. https://doi.org/10.1371/JOURNAL.PONE.0200700

Tillmar A, Fagerholm SA, Staaf J, Sjölund P, Ansell R (2021) Getting the conclusive lead with investigative genetic genealogy – a successful case study of a 16 year old double murder in Sweden. Forensic Sci Int Genet 53:102525. https://doi.org/10.1016/J.FSIGEN.2021.102525

Tornar A, Garofano L, Vecchio C, Pizzamiglio M, Albonici L, Manzari V (1994) The USE of capillary electrophoresis in forensic DNA. Analysis:390–392. https://doi.org/10.1007/978-3-642-78782-9_105

Tu J, Ge Q, Wang S, Wang L, Sun B, Yang Q, Bai Y, Lu Z (2012) Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis. BMC Genomics 131(13):1–9. https://doi.org/10.1186/1471-2164-13-43

W D, L C, H L, P Q, F L, J G, Y Z, B W, C L (2017) Developmental validation of the HomyGene19 +14Y system. Int J Legal Med 131:605–620. https://doi.org/10.1007/S00414-016-1505-2

Wickenheiser RA (2019) Forensic genealogy, bioethics and the Golden state killer case. Forensic Sci Int Synerg 1:114–125. https://doi.org/10.1016/J.FSISYN.2019.07.003

Wyner N, Barash M, McNevin D (2020) Forensic autosomal short tandem repeats and their potential association with phenotype. Front Genet 0:884. https://doi.org/10.3389/FGENE.2020.00884

Xing J, Watkins WS, Witherspoon DJ, Zhang Y, Guthery SL, Thara R, Mowry BJ, Bulayeva K, Weiss RB, Jorde LB (2009) Fine-scaled human genetic structure revealed by SNP microarrays. Genome Res 19:815–825. https://doi.org/10.1101/GR.085589.108

Zhang Q, Zhou Z, Wang L, Quan C, Liu Q, Tang Z, Liu L, Liu Y, Wang S (2020) Pairwise kinship testing with a combination of STR and SNP loci. Forensic Sci Int Genet 46:102265. https://doi.org/10.1016/J.FSIGEN.2020.102265

Part III

Varied Nature of Cases Solved by DNA Profiling

The Effect of Consanguineous Marriages in Solving DNA Cases

23

Noora R. Al-Snan, Fatima J. AlBuarki, and Samreen S. Sayed

Contents

Abstract

Up to date, there are few papers discussing the inbreeding marriages and how they positively affect the process of human identification. In this chapter, we discuss the crucial need of focusing upon the characteristics of STR electropherograms with the excess of homozygous loci found and how can we utilize this kind of DNA profiles in obtaining astonishing findings. We have recruited six buccal swabs from Bahraini volunteers in three generations of inbreeding to illustrate how the inbreeding marriages in first-cousin level can provide with excess number of homozygous loci followed by DNA processing using GlobalFiler Amplification Kit. Also, we have tested unknown suspects from caseworks related to terrorism to test our findings. The ratio of homozygous to heterozygous loci varied from 9.5–42.85% with some markers having more tendency to have homozygous loci than others. We have proposed a forensic DNA intelligence methodology as a guide to follow up to have a full implementation of these DNA profiles to obtain known profiles from the national DNA database. There is considerably high level of inbreeding in the Arab

N. R. Al-Snan (✉)
Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Kingdom of Bahrain, Manama, Bahrain
e-mail: nr.alsnan@interior.gov.bh

F. J. AlBuarki · S. S. Sayed
Forensic Science Laboratory, Directorate of Forensic Science, Ministry of Interior, Manama, Kingdom of Bahrain

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_20

and Muslim countries, with intrafamilial unions accounting for 20–50 + % of all marriages compared to other countries, as well as the emigrants who are residing in the Western areas of the world. These findings are well suited with support of national DNA database and criminal investigation units. Our approaches are recently applied in Biology and DNA Forensic Lab in the Kingdom of Bahrain and have maintained many successful outcomes in solving DNA cases and obtaining positive human identifications.

Keywords

Bahraini population · Consanguineous marriages · Inbreeding · Homozygous loci · DNA database · Forensic DNA intelligence

Introduction

Consanguineous Marriages

Consanguineous marriages are found among one-fifth of the world population mostly residing in the Middle East, West Asia, and North Africa, in addition to the emigrants from these communities now residing in North America, Europe, and Australia (Hamamy 2012). Consanguinity is unadvised in medical practices as well as in some religious conviction such as Islam, due to the reoccurrence of recessive genetic disorders such as neuromuscular disorders, metabolic disorders, osteopetrosis syndromes, and chondrodystrophia that might be lethal to the offspring (Overall and Nichols 2001; Teebi and Teebi 2005).

Nonetheless, there are still some cousin marriages at present day in most of the Arab countries, such as in tribes, conservative families, Bedouins, and in isolated subpopulations. There is considerably high level of inbreeding in the Arab countries, with intrafamilial unions accounting for 20–50 + % of all marriages compared to other countries (Al-Arrayed and Hamamy 2012).

Homozygosity

There are different proposals emphasizing the effect of inbreeding in the forensic evaluation, such as the need of using various genetical parameters: the ratio of coefficient of co-ancestry (F_{ST}) and the coefficient of inbreeding (F_{IS}) to assess DNA profiles in populations where high degree of consanguinity is present along with population substructure (Anwar and Taroni 2019).

While studying the Middle Eastern (ME) populations, it is identified a much larger number of long runs of homozygosity (ROH) in ME populations than in any other populations, which was most likely attributed to high levels of consanguineous marriages that significantly decreased both individual and population heterozygosity (Yang et al. 2014). Traditional methods for estimating the magnitude of inbreeding

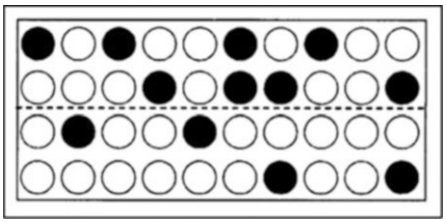


Fig. 1 This population is divided into two equal sizes. One group (top) comprises inbreeding marriages, and the other (bottom) comprises offspring from unrelated parents. (Adapted with CC license from Overall and Nichols 2001)

are generally functions of excess homozygosity found within multiple loci of STR electropherograms (Overall and Nichols 2001). In addition, excess homozygosity was not observed consistently at any one locus in the stated studies (Overall and Nichols 2001; Khubrani et al. 2019a, b). As shown in Fig. 1, the population is divided into two equal-sized groups. One group (top) comprises inbreeding, and the other (bottom) comprises progeny from unrelated parents (Overall and Nichols 2001).

With support of advanced multiplex STRs, the ability to discriminate between individuals has increased as the multiplex selection of markers reduces the risk of adventitious matches (Ludeman et al. 2018). We have previously published the population genetic studies and statistical characterization using 21 autosomal STRs found in the GlobalFiler PCR Amplification kit and the interpopulation diversity between Bahraini population and other populations in the region (Al-Snan et al. 2019). We have found out the crucial need of using more polymorphic multiplex locus such as SE33 to have higher discriminating power between individuals in highly genetic polymorphic populations. Also, it was revealed that the observed deviation from HWE (neglecting the Bonferroni's correction) could be a result of the diversity of the Bahraini population or caused by high polymorphism at the same loci investigated. This observation is likely to reflect the high level of endogamy with consanguinity rates in Bahrain (Al-Arrayed and Hamamy 2012; Al-Snan et al. 2019).

In this chapter, we have assessed the excess of homozygosity loci found in STR electropherograms presented within three generations of first cousins inbreeding families and the ability of utilizing these findings in solving DNA cases such as terrorism, theft, and human remains identification with the assistance of national DNA database.

Technical Updates

We have recruited two volunteered families, first with two married individuals (A, B) from first-cousin marriage along with one of their daughters (C) who married to her first cousin male (D) along with their son (E) and daughter (F), total of six buccal

swabs were assembled using cotton swabs (SceneSafe, UK). Second family consisted of two married individuals as first-degree cousin (G) and (H), and they have conceived one child (I).

Participants were communicated with the corresponding author and presented at the General Directorate of Criminal investigation and Forensic Science – Kingdom of Bahrain, to deliver their buccal swabs for the research after disclosing the informed consent. The age of the participants varied from 5 to 65 years old. Ethical review for conducting tests was obtained and approved by the Ministry of Interior (MOI), Kingdom of Bahrain. All participants provided informed consent prior to contribution their buccal swab samples. The research was performed in accordance with relevant guidelines/regulations.

Genomic DNA was extracted using EZ1 Advanced XL (Qiagen, Germany). Subsequently the extracted DNA samples were quantified using Quantifiler HP DNA Quantification kit (Thermo Fisher Scientific Company, Carlsbad, USA) in the 7500 Real-Time PCR System (Thermo Fisher Scientific Company, Carlsbad, USA) according to manufacturer's recommendation. About 1.0 ng of the extracted DNA was amplified using GlobalFiler PCR Amplification Kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to manufacturer's recommendation using 27 PCR cycles via microAmp optical 96-well reaction plate (Thermo Fisher Scientific Company, Carlsbad, USA) along with the provided positive control and low TE buffer as a negative control in a veriti 96-well thermal cycler (Thermo Fisher Scientific Company, Carlsbad, USA). A total of 24 loci were amplified, including 21 autosomal STR loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818, FGA, D12S391, D1S1656, D2S441, D10S1248, D22S1045, and SE33) and three gender determination loci (Amelogenin, Yindel and DYS391). The PCR products (1 μl) were separated by capillary electrophoresis in an ABI 3500xl Genetic Analyzer (Thermo Fisher Scientific Company, Carlsbad, USA) with reference to LIZ600 size standard v2 (Thermo Fisher Scientific, Inc., Waltham, MA, USA) in total of 9 μl of LIZ600 standard and Hi-Di formamide (Thermo Fisher Scientific, Inc., Waltham, MA, USA) master mix, using the standard injection parameters GeneMapper ID-X Software v1.4 (Thermo Fisher Scientific, Inc., Waltham, MA, USA), which was used for genotype assignment. Experiments were performed in duplicates in the Biology and DNA Forensic Laboratory, Ministry of Interior, Kingdom of Bahrain, which is accredited with CTS testing in regular basis using different quality standards protocols. RFU for reference samples were done using in-house validation for the GlobalFiler Amplification Kit to differentiate between the stochastic threshold and possible allele dropout (Al-Snan et al. 2019).

To validate and test the procedure with sample cases, we have analyzed two different cases, first case was terrorism act on police check point via Molotov bottles. Second case was warehouse containing real IEDs, positive with C4-RDX containing wires with tapes and devices. The collected samples were processed as the samples mentioned above, using GlobalFiler Amplification Kit, in 29 PCR cycles instead of 27. As for the statistical analysis, all the population genetic analysis and forensic

parameters applied can be found in the previous study (Al-Snan et al. 2019). As it is a case study, all the homozygous alleles were directly counted in the first and second generation to find out the ratio of homozygous loci over the heterozygous loci. In-house validation of the analytical threshold is 85 RFU for all the dyes.

Conclusions

The results obtained from the study indicated the following results as shown in Table 1; these results represented the first family as following: (A) grandfather (progeny of inbreeding) married to his cousin (B) who was progeny of inbreeding from her mother side and had daughter (C). Volunteer (C) inbred with her cousin (D) and had the following children (E) and (F) as shown in Fig. 2. Second family represented first-generation cousin marriage (G) and (H) with conceiving only one child (I) as shown in Table 2. DYS391 and Yindel were excluded from the table.

As shown in Table 1, the ratio of homozygous alleles in a family with three generations of inbreeding (first-cousin marriages) was between 19.04% reaching to 23.8% in the studied family.

Table 1 Results generated from six family members of three generation

Locus	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F
D3S1358	15,16	16,17	15,17	16,16	16,17	15,16
vWA	15,17	16,17	15,16	17,18	15,18	16,18
D16S539	11,13	9,12	9,13	11,12	9,12	11,13
CSF1PO	11,12	11,11	11,11	10,11	11,11	11,11
TPOX	8,11	8,10	8,8	8,9	8,8	8,9
AME	XY	XX	XX	XY	XY	XX
D8S1179	14,14	11,15	11,14	14,16	14,14	14,16
D21S11	29,30	30,31.2	29,30	29,32.2	30,32.2	29,29
D18S51	13,17	16,16	16,17	13,13	13,16	13,16
D2S441	11.3,15	11,11	11,11.3	11,11.3	11,11.3	11,11.3
D19S433	11,13.2	12,13.2	12,13.2	13.2,14	12,14	13.2,14
THO1	7,8	6.7	6,7	7,9	7,7	6,9
FGA	19,22	21,25	22,25	19,21	21,22	21,25
D22S1045	16,16	15,16	16,16	11,16	11,16	16,16
D5S818	12,13	9,11	11,13	9,12	11,12	11,12
D13S317	12,12	11,12	11,12	8,12	8,12	8,12
D7S820	10,11	10,10	10,10	8,10	10,10	9,10
SE33	24.2,25.2	14,15	14,24.2	14,30.2	24.2,30.2	14,30.2
D10S1248	14,15	13,16	15,16	14,15	15,16	15,16
D1S1656	13,16	14,16	14,16	12,16	12,16	12,14
D12S391	18,18	16,18	16,18	22,24	18,24	16.24
D2S1338	20,20	19,23	19,20	17,22	20,22	19,22
Number of Homozygous	**5**	**4**	**4**	**2**	**5**	**3**
Ratio of HOM/HET%*	**23.8%**	**19.04%**	**19.04%**	**9.5%**	**23.8%**	**14.28%**

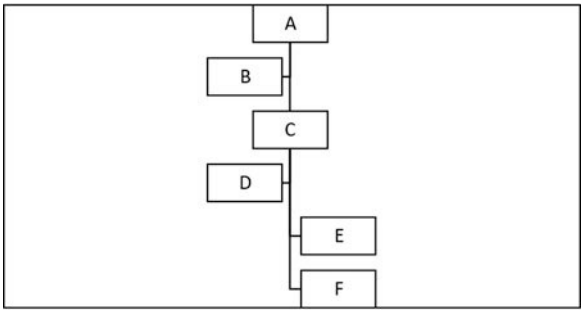


Fig. 2 Family pedigree showing the kinship relationship between the individuals

Table 2 Results generated from three family members of one generation

Locus	Sample G	Sample H	Sample I
D3S1358	15,18	14,17	14,15
vWA	16,17	18,18	17,18
D16S539	8,11	8,9	9,11
CSF1PO	10,13	10,11	10,11
TPOX	8,8	8,8	8,8
AME	XY	XX	XY
D8S1179	11,12	8,10	10,11
D21S11	29,32.2	28,34	28,32.2
D18S51	14,17	14,16	14,14
D2S441	10,11	11,11	10,11
D19S433	13,16.2	14,14.2	13,14.2
THO1	7,9	6,9	7,9
FGA	22,23	21,23	22,23
D22S1045	15,17	16,17	16,17
D5S818	7,13	12,13	7,12
D13S317	11,12	11,11	11,11
D7S820	8,10	8,10	8,10
SE33	18,19	29.7,31.2	18,31.2
D10S1248	14,15	13,14	14,14
D1S1656	15,16	15,15	15,15
D12S391	18,23	18,25	18,25
D2S1338	17,23	22,23	23,23
Number of Homozygous	**2**	**4**	**6**
Ratio of HOM/HET%*	**4.7%**	**19.04%**	**28.6%**

***HOM: Homozygous loci ; HET: Heterozygous loci**

Also, some of the markers had more tendency to have homozygous alleles such as CSF1PO (four times), D22S1045 (three times), D7S820 (three times), TPOX (two times), D8S1179 (two times), and D18S51 (two times) and the other investigated markers had heterozygous loci and few markers had one homozygous locus.

As shown in Table 2, the ratio of homozygous alleles in a family with one generation of inbreeding (first-cousin marriage) was 28.6% for the child, having the following homozygous loci: TPOX, D18S51, D13S317, D10S1248, D1S1656, and D2S1338. Tendency of these markers to have homozygous loci might depends on the marker location in the chromosomes and motifs sequence.

Further analysis must be conducted to show the correlation study between different loci and homozygosity. Actual cause is not described in the literature as the lack of consistency across loci suggests that null alleles might be not the general cause, and that demographic and stochastic factors are more likely to occur (Khubrani et al. 2019a). However, we have detected significant associations between the total quantity of homozygous alleles in both investigated families.

We have followed the same methods previously applied for two samples, Fig. 3 which is an electropherogram of touch DNA collected from cap of Molotov bottle and Fig. 4 which is an electropherogram of touch DNA collected from blue tape linked with wires and connectors in real IED case.

As shown in these Figs. 3 and 4, the number of homozygous loci in suspect of Molotov case was 7, in ratio of 33.33% (HM/HET), and the number of homozygous loci in suspect for blue tape in real IED was 9, in ratio of 42.85% (HM/HET).

Both cases indicated and gave predictions of having suspects from both parents are biologically related in terms of consanguineous marriage. When applying the homozygous loci in the National DNA database, the partial match profiles came to indicate the related sibling entered in the DNA database, thus leading to identifying the wanted suspects as shown in Table 3. These known matches gave positive forensic leads and aided the identification of the wanted suspects by collecting reference samples from the closely related individuals such as the siblings.

Inbreeding is highly practiced in most of the Arab and/or Muslim communities around the world for many reasons: Ease of communication, unique cultural cultures, and religious convictions have also allowed these population to marry within families. These consanguineous marriages have produced greater similarities between the individuals of these populations. These similarities between individuals should be addressed and considered in every evidence evaluation involving DNA profiles (Anwar and Taroni 2019). In this chapter, we have investigated the importance of using the excess of homozygous loci in the unknown DNA profiles in finding the most fit profiles of other siblings or family-related entry in the national DNA database. These findings support the human identification for different forensic cases to support and assist the interpretation of DNA results.

This procedure is practiced and qualified in Biology and DNA Forensic Laboratory in the Kingdom of Bahrain and presented to be exceedingly successful in most of the cases. We have defined the procedure to follow for such cases as follows (Fig. 5 for summary): First, to generate STR-DNA profiles of single source (preferably using any multiplex PCR system of 24 loci or above), then to study the unknown STR-DNA profiles and check if there are any known matches in the national DNA database. If there are no matches, enter the homozygous loci only found in the STR-DNA profiles. (Most preferably to have not less than 8 homozygous loci and you can also combine the unusual microvariants found in the DNA

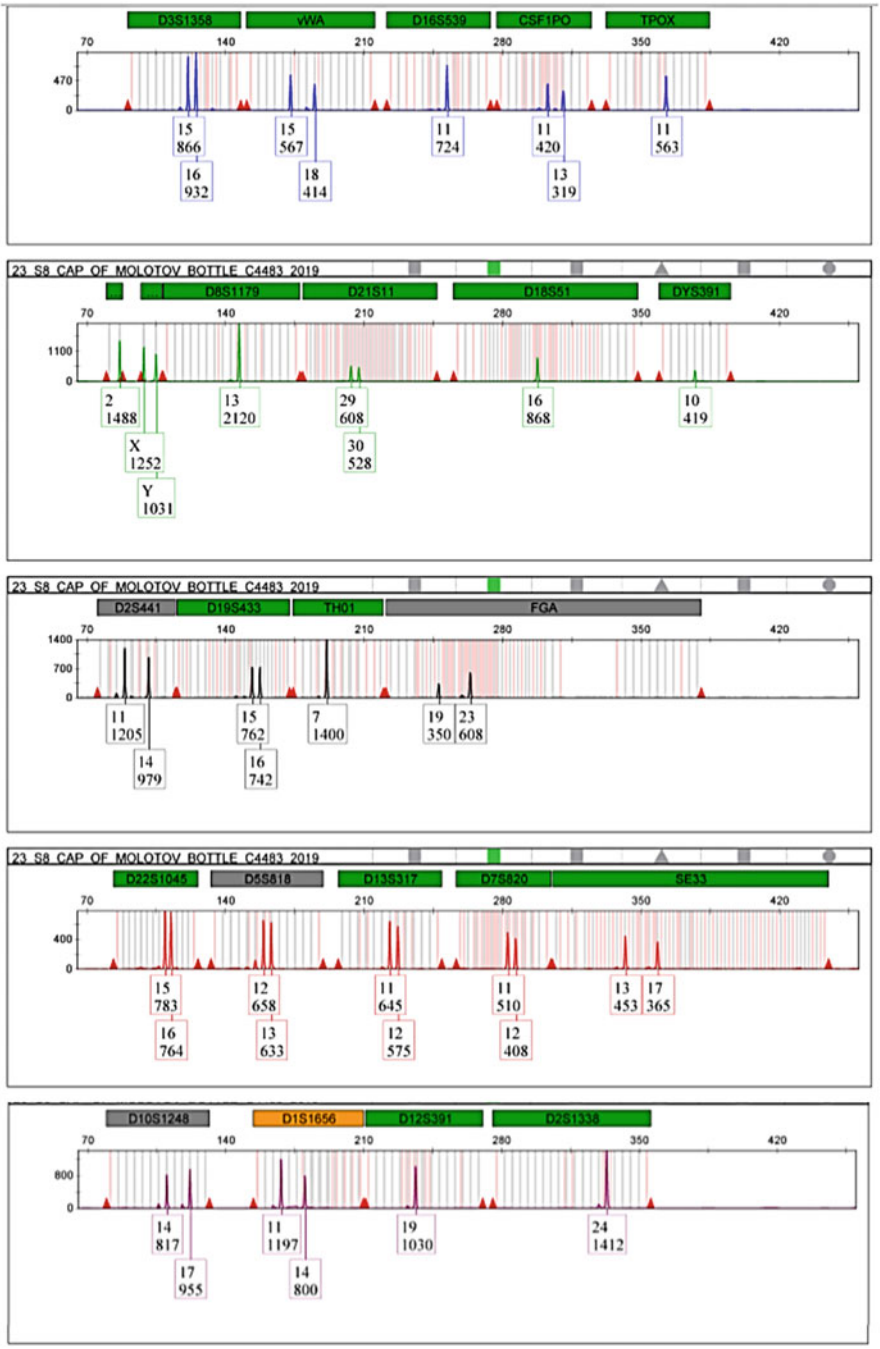


Fig. 3 Electropherogram of touch DNA collected from cap of Molotov bottle

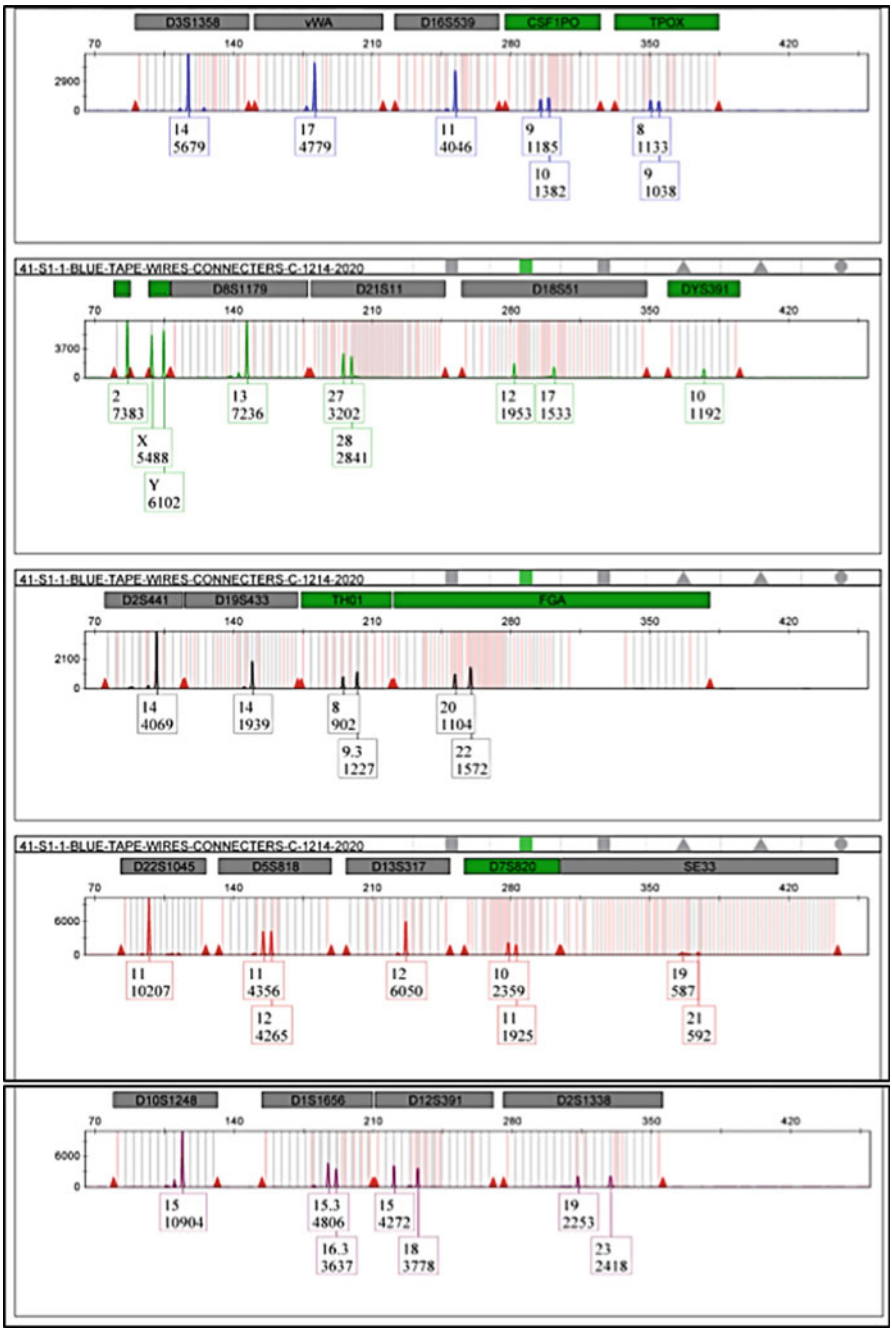


Fig. 4 Electropherogram of touch DNA collected from blue tape linked with wires and connectors in real IED case

Table 3 Partial match results from DNA database using the homozygous loci, the highlights indicate the homozygous loci

Locus	Molotov unknown	Known partial Match for Molotov	IED unknown	Known partial Match for IED
D3S1358	15,16	15,16	14,14	14,16
vWA	15,18	16,17	17,17	17,17
D16S539	11,11	11,11	11,11	11,12
CSF1PO	11,13	11,13	9,10	10,11
TPOX	11,11	8,11	8,9	8,9
AME	XY	XY	XY	XX
D8S1179	13,13	13,13	13,13	13,16
D21S11	29,30	29,29	27,28	27,29
D18S51	16,16	14,16	12,17	17,19
D2S441	11,14	10,14	14,14	11,14
D19S433	15,16	13,16	14,14	12,14
THO1	7,7	6,7	8,9.3	8,9.3
FGA	19,23	23,25	20,22	20,23
D22S1045	15,16	16,18	11,11	11,15
D5S818	12,13	11,12	11,12	11,11
D13S317	11,12	12,13	12,12	11,12
D7S820	11,12	10,14	10,11	9,11
SE33	13,17	13,17	19,21	18,21
D10S1248	14,17	14,17	15,15	15,15
D1S1656	11,14	13,14	15.3,28	15.3,16.3
D12S391	19,19	19,19	15,18	15,18
D2S1338	24,24	19,24	19,23	19,24

profiles such as 0.1, 0.2, 0.3… etc.). Depending on the number of DNA profiles entered in the national DNA database, the number of matches with the homozygous and unusual microvariants will be obtained.

After that, study the partial matches of the known DNA profiles, check the whole profiles if it is also similar with the unknown profile. The most fit known DNA profiles might be related with the unknown DNA profile in terms of sibling index, paternity, or maternity index. You might need to perform required statistics and analysis to have more accurate findings based on the known DNA profiles and the unknown profiles. If possible, perform YSTR DNA analysis for the known DNA profiles and the unknown DNA profiles using any YSTR amplification PCR kit. Report the findings for biosecurity check of all the related siblings/family of the known profiles. Finally, obtain all the reference samples with their informed consents for all the known profiles and perform similar STR-DNA analysis testing and find the unknown DNA profile.

These findings are well suited with support of DNA database and criminal investigation units. It can be applied in any DNA cases with unknown single DNA

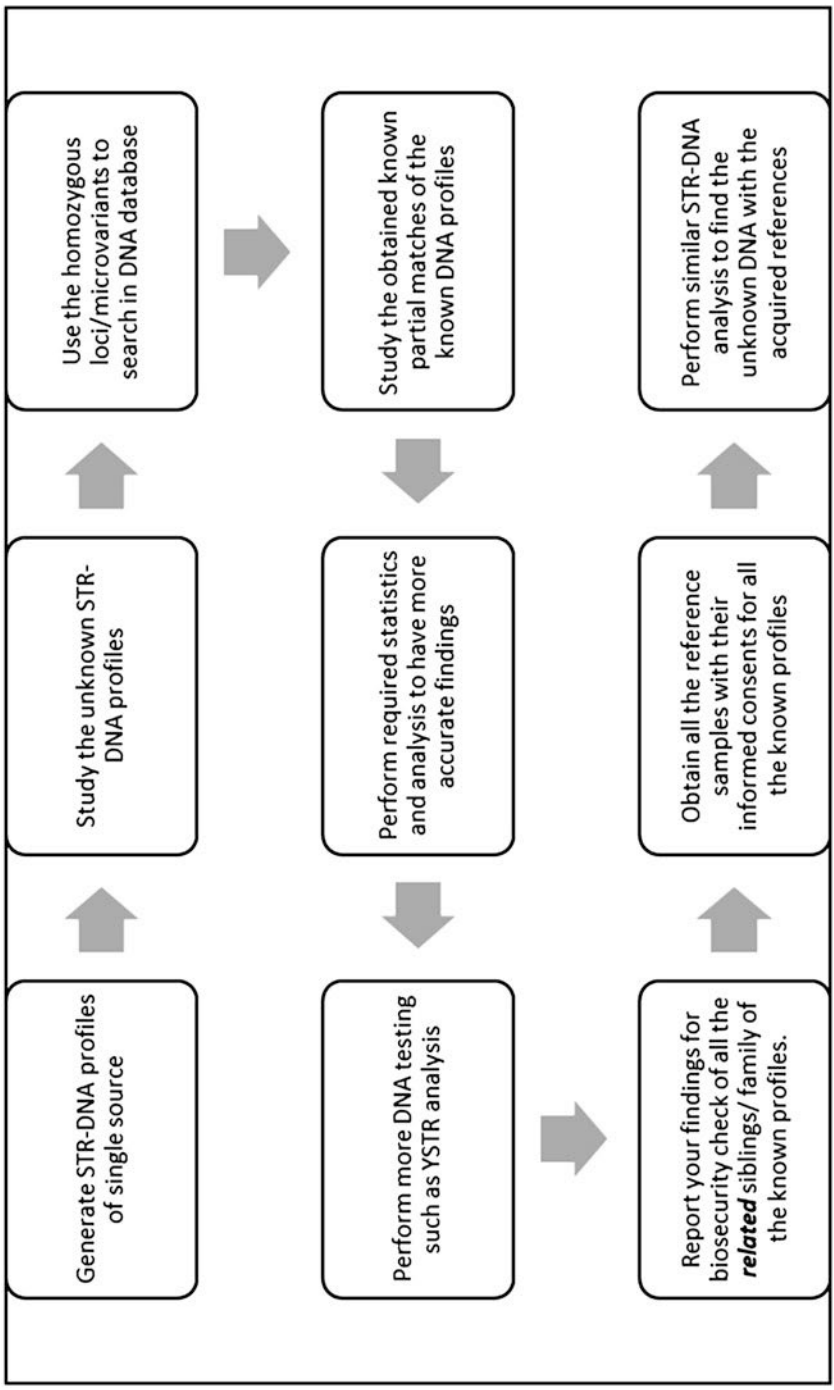


Fig. 5 Summary of the well-defined methodology for unknown source of STR electropherograms referred to consanguineous marriages

profiles and finds matches with the entered known DNA profiles who can be related or siblings of the unknown. Also, there are some studies indicating that environmental influence and genetic predisposition can influence the criminal behavior in members of same family such as twins and siblings (Raine 2002). As said, it is more likely to have convicts from same members of the family performing similar criminal act such as theft or terrorism and one of them has his/her DNA profile entered in the DNA database, which can support our methodology in obtaining a match for the unknown DNA profile using excess homozygous loci. Another proposal is to initiate YSTR database based upon surname and tribes' name which can assist in identifying the surname of the unknown YSTR-DNA suspect profile and can lead to important findings. In preserved and endogamous communities, such as the Arab and Muslim worldwide, these types of databases are crucial for speeding up the process of human identification and identifying the human remains in DVI cases. Extra caution should be taken to avoid falsely supporting inclusion of nondonors (Semaan et al. 2020).

There are limited studies on consanguineous marriages based on forensic aspect, and mainly they cover the structure of populations. Our chapter discusses the importance this point of view has in this kind of inbreeding and how can they solve and identify the unknown suspects based on the excess of homozygous loci as well as YSTR database for surname. These methodologies are applied in Biology and DNA Forensic Lab in the Kingdom of Bahrain and have maintained many successful findings. More detailed genetic databases should be developed for forensic DNA purposes, based on reference data from each of the appropriate subpopulations and not on random or combined samples. Further investigation would require testing our findings in a larger group of families to investigate the effect of inbreeding with the number of homozygous alleles.

References

Al-Arrayed S, Hamamy H (2012) The changing profile of consanguinity rates in Bahrain, 1990–2009. J Biosoc Sci 44(3):313–319

Al-Snan NR, Messaoudi S, Babu SR, Bakhiet M (2019) Population genetic data of the 21 autosomal STRs included in GlobalFiler kit of a population sample from the Kingdom of Bahrain. PLoS One 14(8):e0220620

Anwar I, Taroni F (2019) Genetic peopling of Pakistan: influence of consanguinity on population structure and forensic evaluation of traces. Forensic Sci Int 7:232

Hamamy H (2012) Consanguineous marriages. J Community Genet 3(3):185–192

Khubrani YM, Hallast P, Jobling MA, Wetton JH (2019a) Massively parallel sequencing of autosomal STRs and identity-informative SNPs highlights consanguinity in Saudi Arabia. Forensic Sci Int Genet 102164

Khubrani YM, Wetton JH, Jobling MA (2019b) Analysis of 21 autosomal STRs in Saudi Arabia reveals population structure and the influence of consanguinity. Forensic Sci Int Genet 39:97–102

Ludeman MJ, Zhong C, Mulero JJ, Lagacé RE, Hennessy LK, Short ML, Wang DY (2018) Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. Int J Legal Med 132(6):1555–1573

Overall A, Nichols R (2001) A method for distinguishing consanguinity and population substructure using multilocus genotype data. Mol Biol Evol 18(11):2048–2056

Raine A (2002) The biological basis of crime. Crime 43:74

Semaan M, Abbas S, Mansour I (2020) A mixed DNA profile controversy. J Forensic Res 11(453):2

Teebi AS, Teebi SA (2005) Genetic diversity among the Arabs. Public Health Genomics 8(1):21–26

Yang X, Al-Bustan S, Feng Q, Guo W, Ma Z, Marafie M, Jacob S, Al-Mulla F, Xu S (2014) The influence of admixture and consanguinity on population genetic diversity in Middle East. J Hum Genet 59(11):615–622

A Case Study on Murder Mystery Solved by DNA Typing

24

Subhasish Sahoo and Rashmita Samal

Contents

Abstract

DNA evidence plays the role of a powerful tool in linking a suspect to the crime. DNA typing in forensics is used for identification and paternity and immigration cases. It helps not only in identifying the culprit but also in exonerating the innocent. DNA typing and its continuous advances in forensic science and genetics have made these techniques almost foolproof to be applied in various complex cases. Here, we will discuss a murder case where a newlywed woman killed her husband on the first night but concealed it as death due to snake bite. Injection syringes used by her for injection of drugs were recovered from the backyard. The injection syringes contained a speck of dried blood, which came out to be a good source of DNA evidence. DNA was extracted from the dried blood, quantified, amplified, and sequenced on automated DNA sequencer. The short tandem repeat (STR) profile generated matched with the profiles of parents of the deceased at 21 autosomal STR loci and additionally 3 gender loci. DNA typing could confirm the alleged murder of the deceased by the newly wed woman and her accomplice by administration of drugs leading to cardiac arrest.

S. Sahoo (✉)
HOD DNA Division, State Forensic Science Laboratory, Bhubaneswar, Odisha, India

R. Samal
DNA Division, State Forensic Science Lab, Rasulgarh, Bhubaneswar, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_22

Keywords

Murder · DNA evidence · Injection syringe · Short tandem repeats · Autosomal DNA typing

Introduction

DNA typing has become the gold standard in forensic science since that first case 30 years ago. The advances in the 1970s paved the way for the detection of variation (polymorphism) in specific DNA sequences and shifted the study of human variation from the protein products of DNA to DNA itself (National Research Council (U.S.) 1992). As per the National Crime Records Bureau (NCRB) report, 28,918 murder cases were registered in the year 2019 (National Crime Records Bureau 2019). Advanced DNA typing techniques can ascertain the identity of accused in heinous murder/homicide cases. Evidences collected from the scene of crime containing even a trace amount of DNA evidence of accused/victim can lead to establishment of crime and identify the victim as well as accused. DNA evidence can be termed the world's best crime fighting technology and can be found from substances as tiny as drop of blood/semen/saliva, comb, toothbrush, cigarette butt, and strands of hairs, or even touch DNA can be sufficient enough to establish a crime and identify the culprits.

DNA is a powerful tool when biological evidence from crime scenes is collected and stored properly; forensically valuable DNA can be found on evidence that may be decades old. Therefore, cold cases that were previously thought unsolvable may contain valuable DNA evidence capable of identifying the perpetrator (Hagelberg et al. 1991). A 52-year-old cold case was solved by Seattle police, and mystery surrounding Mary Sullivan was resolved by the Boston Police Department almost 50 years after her death with the help of DNA technology (Bulman Philip 2014).

Linking the suspect to the crime is of utmost importance in a case of murder as shown in Fig. 1. A vital DNA evidence can be availability of suspect's DNA on victim's body/clothing/weapon of offence/any other object from the scene of crime or availability of victim's DNA on suspect's body/clothing/belongings. The list of evidences which can act as DNA evidence goes on as shown in Table 1. It can be swab from the blood-stained floor/blood-stained soil, blood left on a broken window, saliva found on the cigarette butt/beer bottle, skin cells on the wheels of a stolen car, sweat, skin cells from hat/mask/dirty laundry which would lead to establishing the crime scene; clothing of the deceased/its body parts and bone samples of skeletonized body would lead to establishing the victim/deceased; stones/bricks, knife/sickle, bullet, bullets recovered from the body of the deceased, etc. would lead to establishing the weapon of offence; clothing of the suspect with suspected blood-stains/rooted hairs found on the deceased body/in the crime scene, presence of saliva-stained cigarette butts/bear bottles, glasses at the crime scene, saliva from bite marks, fingernail scrapings, eyeglasses, toothpick, envelope seal, bottle, glass,

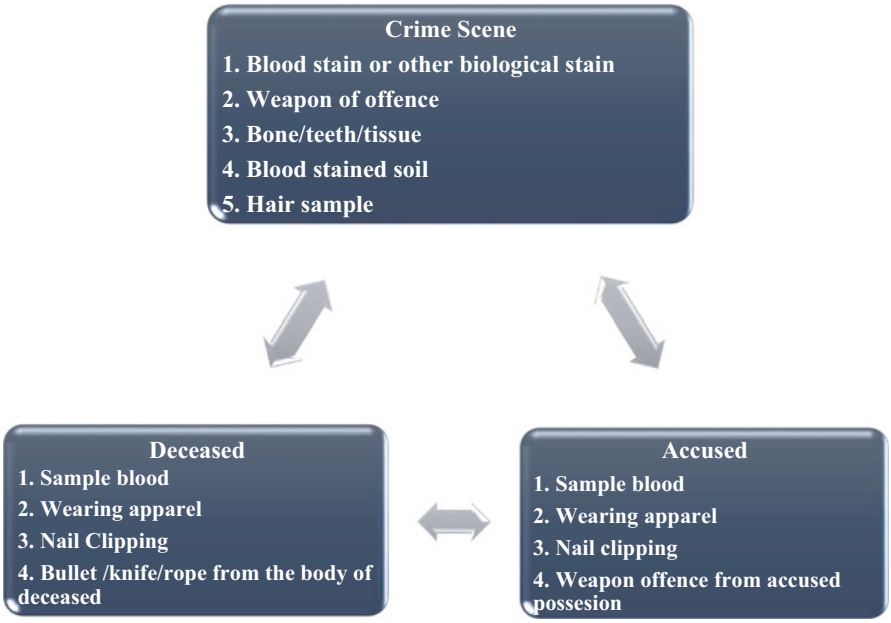


Fig. 1 Locard's exchange principle linking crime scene with victim and suspect

etc. would connect the suspect, thereby connecting the dots between suspect to victim/deceased to weapon of offence and scene of crime.

DNA evidences are often degraded, contaminated, or from multiple unknown sources. DNA typing technology sometimes cannot be repeated, because there is too little of sample and it often involves comparison among a wide range of alternatives in the population which requires statistical analysis of population frequencies. DNA evidence's contamination-free collection, preservation, and examination are vital in corroboration of occurrence of crime. Recent advances in DNA technology have enabled us to retrieve DNA profiles from as little as skin cells left behind when a criminal merely touches a surface. This improved sensitivity combined with new data analysis approaches has made it possible for investigators to identify and distinguish multiple individuals from the DNA in a mixed sample or compromised sample. And it's now even possible to run and analyze samples in less than 2 hours and nab the culprit as soon as possible.

Brief Case History

A young person died just on the first night after his marriage, in his own bedroom with puncture marks on his hand. His wife claimed that he was bitten by a snake, leading to death. In the beginning, a UD case was registered by the father of the

Table 1 Some of the possible evidence samples and collection procedure

Sl. no.	Sample	Location and condition	Type of exhibit	Methods of collection, preservation, and packaging
1	Blood	Crime scene (fresh/wet cloth/ damp/liquid spots)	Blood-stained soil, clothing, fabrics, victim's clothing, suspect's clothing, objects, etc.	Should be transferred on sterile gauze cloth/saline extract from the object using sterile gauze and then air-dried at room temperature and packed in paper packet
2	Dried blood stain, saliva stain, sweat stain, etc.	Crime scene (crust/stain/ spatters)	Unmovable surface, floor, concrete wall, carpet, wallpaper, wood, weapon, firearm, bullet small objects such as household utensils, stones, bricks, beer bottle, cigarette butt, vehicle upholstery, etc.	The dried stain should be moistened for 5–10 minutes with sterile saline solution/distilled water The moistened stain is collected with foam-tipped swab/FTA card/ gauze cloth; stained area is cutout; the item as such is air-dried and packed in paper packet
3	Bones/ teeth	Crime scene or place of recovery (wet/semidry/ dry)	Dead body/skeletal remains such as: a) Femur b) Molar teeth c) Tibia d) Humerus e) Radius and ulna f) Ribs g) Any other teeth	The bones or teeth should be cleaned to remove any tissue debris and allowed to dry completely and room temperature and packed in paper/cardboard parcels
4	Hair with root	Crime scene, weapon, clothing (dry or wet with blood, semen, saliva)	Scalp hair, body hair, Pubic hair, etc.	The hair sample should be collected with the help of tweezers/forceps with rubber tip in clean white paper/butter paper and packed in paper envelope

deceased, but later they came to know about their daughter-in-law having an alleged affair with a medicine store shopkeeper of her village. Suspecting foul play, the UD case was turned to murder case but by this time the dead body was disposed of. The investigating team took help of the DNA examination to corroborate the crime.

Burnt bone pieces of deceased were collected from cremation ground, and injection syringes and injection needles were collected after recovery from backyard of the house of the deceased along with different insulin and sedative strips. Further sample blood on FTA cards of the father and mother of the deceased were sent for DNA typing.

Description and Source of Samples

Burned bone pieces of the deceased collected from cremation ground (Exhibit 1); injection syringes 03 nos. Collected after recovery from the backyard of the house of the deceased (Exhibit 2) (Fig. 4); injection needles 08 nos. Collected after recovery from the backyard of the house of the deceased (Exhibit 3) (Figs. 2 and 3); many strips of medicines and sample blood of the father of the deceased on FTA card (Exhibit 4); and sample blood of the mother of the deceased on FTA card (Exhibit 5) were received in this case. All the samples were maintained in their recommended storage condition. Appropriate seal and label were checked properly prior to their receipt and examination.

Procedure

I. Exhibits 3, 4, and 5 were subjected to automated DNA extraction method to isolate DNA on Qiagen EZ1 Advance XL using EZ1 DNA Investigator Kit as per the protocol given by manufacturer.
II. DNA could not be obtained from the burned bone pieces of deceased (Exhibit 1), as they were completely burned and on physical examination found unsuitable for DNA extraction.
III. From physical examination, dried blood was observed on the back side plastic part of one of the injection needle (Exhibit 3) and was scrapped out and subjected to automated DNA extraction method.
IV. Sample blood of the father of the deceased on FTA card (Exhibit 4) and sample blood of the mother of the deceased on FTA card (Exhibit 5), subjected to automated DNA extraction method.

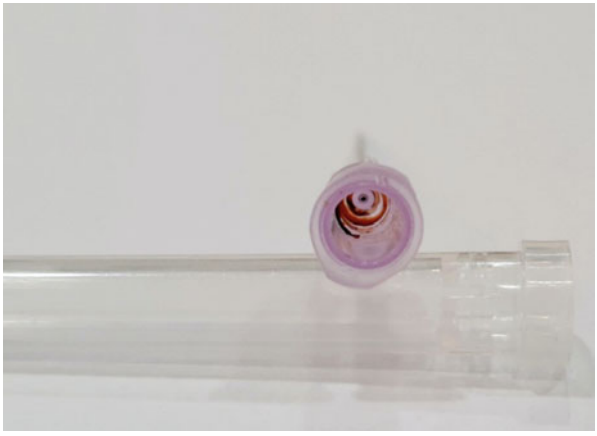


Fig. 2 Exhibit 3, needle with dried blood recovered from backyard of the house of the deceased

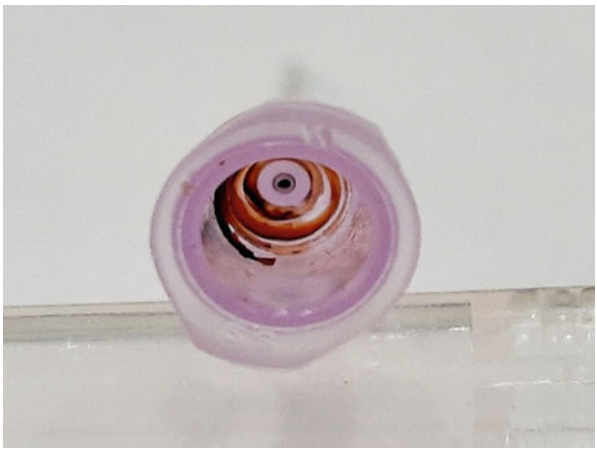


Fig. 3 Close view of needle with dried blood recovered from the backyard of the house of the deceased

Fig. 4 Exhibit 2, syringe recovered from backyard of the house of the deceased

V. Extracted DNA was quantified by 7500RT-PCR Applied Bio-system (ABI) using Quantifiler ®HP DNA Quantification Kit and was further diluted using TE buffer to obtain the optimum DNA concentration for multiplex PCR.
VI. Amplification was done by AmpFISTR®, Globalfiler® Kit, supplied by Applied Bio-System, to multiplex PCR reaction for co-amplification of 21 autosomal STR loci and additionally three gender loci on Veriti™ 96-Well Fast Thermal Cycler (ABI, USA) as per the protocol supplied by the manufacturer.
VII. For capillary electrophoresis/genotyping, the amplified products along with control samples, supplied by the Applied Bio-System, were run on the automated DNA sequencer model 3500 XL- Applied Bio-System (ABI). The sizing of fragments was carried out using Gene Mapper® IDX software with respect to Gene ScanTM 600 Liz TM size standard.
VIII. Appropriate positive and negative controls were used throughout the analysis, and subsequently data analysis was performed using the manufacturer's protocols.

Results and Observation

I. The results of the autosomal STR DNA typing of the exhibits are given in Table 2. One of the allele in the genetic profile of Exhibit 3 (dried blood remaining scrapped from needle) matches with one of the allele in each of the

Table 2 Comparison of autosomal DNA profiles generated from Exhibit 3 with 5 and 4

Sl. no.	Name of loci	Sample blood of the mother of the deceased on FTA card, Exhibit 5	Dried Blood remaining scrapped from needle, Exhibit 3	Sample blood of the father of the deceased on FTA card, Exhibit 4
1	D3S1358	15,17	14,17	14,17
2	VWA	16,17	17,17	16,17
3	D16S539	11,12	9,12	9,11
4	CSF1PO	11,11	11,11	11,11
5	TPOX	8,11	8,8	8,8
6	Yindel	–	2	2
7	AMEL	XX	XY	XY
8	D8S1179	15,15	14,15	14,14
9	D21S11	29,32.2	29,29	29,30
10	D18S51	17,19	14,19	14,17
11	DYS391	–	10	10
12	D2S441	11,14	14,14	11,14
13	D19S433	12,16	14,16	13,14
14	TH01	9,9	9,9	6,9
15	FGA	21,26	21,21	21,23
16	D22S1045	11,16	11,15	15,16
17	D5S818	12,13	11,13	10,11
18	D13S317	10,10	10,10	10,10
19	D7S820	8,13	8,11	11,12
20	SE 33	14,15	14,33.2	17,33.2
21	D10S1248	14,15	13,14	13,14
22	D1S1656	17,18	15,18	15,17
23	D12S391	18,20	18,18	17,18
24	D2S1338	20,20	18,20	18,20

autosomal STR loci of Exhibit 5 (sample blood of the mother of deceased on FTA card), and all the non-maternal alleles of each loci match with one of the allele of Exhibit 4 (sample blood of the father of the deceased on FTA card).

II. Amelogenin marker of Exhibit 3 shows "XY" suggesting the source of the sample to be a male individual. Further, consistent autosomal STR DNA profile was detected from Exhibit 4 and Exhibit 5.

III. From Exhibit 1 (burned bone pieces), DNA profile could not be generated as the bones were completely burned and were found unsuitable for DNA extraction.

IV. The DNA profile generated from the Exhibit 3 (dried blood remaining scrapped from needle) is matching with the DNA profiles generated from Exhibit 5 (sample blood of the mother of the deceased on FTA card) and Exhibit 4 (sample blood of the father of the deceased on FTA card). Therefore, the source of Exhibit 3 (dried blood remaining scrapped from the needle) is identified as from the deceased.

V. This suggests that the wife of the deceased injected the deceased with insulin and other drugs leading to cardiac arrest and death.

Conclusion

This is an unusual murder case, and trail of events suggest the newly married woman was having love affair with another man and was forced by her family members to marry this man. Her lover and the lady conspired to kill the man and seemed it to look like an accidental death due to snake bite. The lady got the injections from her lover and administered them to her husband, thereby leading to death of the person. The police made intensive search and found the used injection needles and vials and other drugs from the backyard of the house of the deceased, which were buried under the soil to conceal the crime. On further investigation and interrogation, the accused was apprehended and her compliance was also arrested. The result of forensic DNA examination of the dried blood scrapped from the injection needle found from the backyard of the deceased person's house matched 50% with that of the father and the other 50% with the mother of the deceased, thereby indicating that the deceased was injected with the drugs, confirming that the marks on the deceased body were of injection punctures and not of snake bite. This further confirmed the charge of murder in this case by administration of drugs by the newly wed lady to her husband.

Precautions

I. The chain of custody for the exhibits should be maintained properly, and storage of exhibits should be done in proper conditions as required.
II. The exhibits should be photographed and recorded accordingly for future reference.
III. All the plastic wares, glasswares, and working area should be sterile.
IV. Samples should be handled carefully, so that they do not get mixed up in the subsequent steps of experiment, and the sample tubes should be labeled properly and clearly before transferring the samples.
V. Internal laboratory record for case opening, DNA extraction, quantification, genotyping, and result interpretation should be maintained properly.
VI. The alleles need to be matched one by one among the obtained profiles.
VII. Other precautions as described by manufacturer protocol in the individual experiments for DNA extraction, quantification, multiplex PCR, STR genotyping, and result interpretation should be referred and followed.

References

Hagelberg E, Gray IC, Jeffreys AJ (1991) Identification of the skeletal remains of a murder victim by DNA analysis. Nature 352(6334):427–429. https://doi.org/10.1038/352427a0

National Crime Records Bureau (2019) Ministry of Home Affairs, Government of India

National Research Council (U.S.) (1992) DNA technology in forensic science. National Academy Press, Washington, DC

Philip B (2014) Solving cold cases with DNA: the Boston strangler case. NIJ J 273:48–51. Available at http://www.nij.gov/journals/273/pages/boston-strangler.aspx

Sexual Assault and Murder: When DNA Does Not Help Even Though It Is Present

25

Mark Benecke

Contents

Abstract

Two cases demonstrate that full reliance on biological stains ("CSI effect") may, in complicated or unlikely case scenarios, lead to either wrongful convictions or no prosecution at all. Case 1 was a homicide in a small town in which the short tandem repeat (STR) profile of the neighbor of the deceased was found on two fingers of the dead person. No motif for the crime was established. Transfer of DNA through newspapers was excluded even though the neighbors did share newspapers on a daily basis. We performed experiments and found an unexpected solution for the DNA findings. In Case 2, a woman reported that she was anally raped for about 1 hour in her bed while her parents and grandfather were in the same house. The police and DA's office of the larger city dismissed the case as improbable even though hematoma and anal fissures were documented in the hospital and the woman audio-recorded the event on her cell phone. On the oral, vaginal, and anal swabs, no or hardly any autosomal STRs of the man were found which led to a non-prosecution. On inspection of her pantyhose, we found clear signals from Y-STRs, though. We discuss structural problems that may occur even in highly standardized laboratories if wrong assumptions about – unlikely yet true – events enter the case files.

M. Benecke (✉)
Forensic Biology Unit, International Forensic Research and Consulting, Cologne, Germany
e-mail: forensic@benecke.com

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_23

Keywords

CSI effect · DNA profiling · STR · Homicide investigation · Rape · Sexual assault · Homicide · Forensic DNA typing · DNA transfer · Case work

Introduction

Due to a surge of series like CSI in the 1990s, and many that followed and flashily dealt with laboratory techniques in crime cases, the general expectation of the public and (where installed) juries in crime cases was that all crime scenes must be subjected to genetic fingerprint analysis – the "CSI effect." Soon, this became a "reverse CSI effect" (Reibe and Benecke 2010), i.e., "*while jurors may have come to expect, as a result of CSI-type shows, high-tech forensic testimony in criminal cases, and may inappropriately acquit when such evidence is lacking, these same jurors, as a result of these same CSI-type shows, often place too much weight on forensic evidence in cases where forensic evidence IS in fact produced by the prosecution, resulting in convictions in cases where the defendant probably should have been acquitted"* (Godsey and Alou 2011).

Two cases show that even in legal systems without a jury, high expectations on stains may lead nowhere if classical criminalistic thinking is not applied, even if DNA is present and seems to lead to obvious results (Benecke 2021b).

Case 1: Homicide

In 2002, a 65-year-old man who lived together with his female partner was convicted to a lifelong prison sentence for homicide II. In Germany, this usually means a stay in prison for at least 10 years or longer.

The now convict had no motive for killing his neighbor except an alleged smaller quarrel that had taken place a few years ago about an unknown matter. Some witnesses mentioned gossip that the two neighbors, both elderly men and brothers-in-law, did not speak to each other. Even though they shared the daily newspaper to save money. First, the later convict read it during daytime and then put it on the doorstep of his neighbor who picked it up in the afternoon and subsequently read it. During the trial, the gossip statements could not be corroborated; in contrast, it was found that many statements from villagers about general affairs between the inhabitants had been untrue.

The later convict had found his dead neighbor in the neighbor's living room/kitchen sofa. The dead person had been bound with zip ties on this wrists and ankles, lay on his face, and was beaten and stabbed to death. On the table, the equivalent of 300 US$ was found, and two drawers were pulled open. Our client, the finder of the dead man, then called the police. It remained unclear if a metal box with valuables was stolen from a cupboard. The finder remembered such a box from a former visit, the nephew of the deceased denied that such a box had existed. It was however established that jewelry of the dead man's deceased sister was missing. The object

used to batter the now dead man was not found. The front door was forcibly opened even though our client, the neighbor could have more easily walked over to the now dead man via a courtyard on the other side of the house.

In 2003, the trial was temporarily stopped after 39 trial days and the arrest warrant (custody) lifted. The inside areas of two fingers on the left hand of the deceased contained foreign DNA, and the judge decided – quote – to make this biological stain the deciding one. During questioning, the DNA expert (working for neither side but reporting for and to the court) said that the DNA on the fingers of the deceased may have been placed there due to the defendant "forcefully supporting his weight whilst binding the hands of the victim." Nine out of nine STR systems matched the defendant's DNA profile. In 2004, he was sentenced for malevolently killing his neighbor in his sleep.

The life partner and the medical doctor of the defendant said that the now sentenced man had severe back pain on the day of the murder, could not walk properly, and therefore could not have gone over to his neighbor much less apply force against him. This was confirmed in the doctor's written files in his office. On the cable binders, no DNA was found. A human hair found at the crime scene was STR typed and found to be neither matching the genetic fingerprint of the defendant nor that of the deceased.

In 2008, we decided to check upon the possibility of secondary transfer of DNA via the newspaper that was shared between the neighbors. This topic had already been dealt with during trial, but the DNA expert then had stated that the amount of DNA found on the two fingers of the deceased was too high: After 30 PCR cycles that had been used, signals from secondary transfer reached signal levels between undetectable up to maximally 400 relative fluorescence units (RFUs), whereas primary stains reached levels of 600 RFU and above. According to the expert and the later verdict, DNA on paper is also bound more rigidly there and can therefore be not transferred easily from paper to skin.

DNA signal thresholds depend of course on the laboratory procedures and decisions made out of context (Coble and Hill 2012; Butler 2014). Since to our knowledge, no systematic studies about DNA transfer via newspaper reading – either via skin cells in sweat after scratching one's head or regular skin lines/regular fingerprints – were available, we decided to replicate the situation with actual newspapers, actual readers, and the same STRs and machines (ABI Prism) used by the state laboratory. We found that freshly washed fingers did hardly transfer any DNA detectable by STRs to the fingers of the next reader, whereas unwashed fingers, especially after licking them, did lead to a transfer of higher amounts of DNA – mostly below 200 but sometimes, in single alleles, up to 800 RFUs (Fig. 1). Due to German laws and regulations and due to the absence of "Innocence Projects" like in the United States and England, the family had to pay all legal, laboratory, and expert costs.

A re-trial in Germany needs new, "never-used" evidence or stain groups; so our findings were technically considered to be of no juridical value: The matter of secondary DNA transfer was already dealt with in the first court decision and therefore "used up" procedurally, i.e., could not be presented again to request a new trial, even if our results seemed to be worth following scientifically.

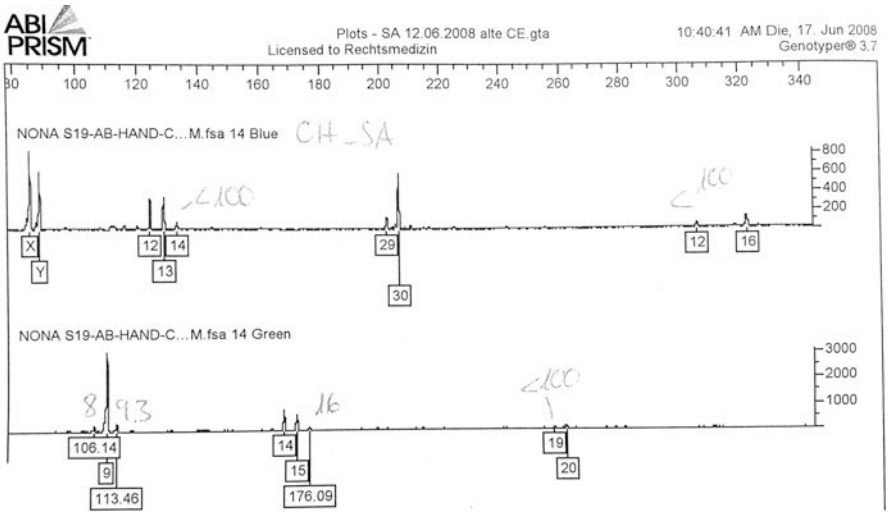


Fig. 1 Our experiment: STR profiles due to experimental transfer via newspaper reading. Some alleles drop out; others may reach high fluorescence levels

We decided that criminalistically, the case was too much dependent on the DNA evidence, while much else, including the broken front door and financial transactions in the family, allowed other ideas of what had happened. To our surprise, our client, the now prisoner, told us in prison – while his new lawyer briefly stepped outside to visit the visitor's restroom – that he, our client, had seen the murder, committed by his nephew, through the back window facing the yard. He "dared not to stir up this family matter," "fully trusted the police," and therefore waited until the nephew had left the scene of crime. Then, our client entered the scene and touched the hand of the deceased to see if he was alive. Our client dared not to speak to anyone about the incident which in light of his socialization under a dictatorial socialist regime with very strong, everyday presence of the secret service in all aspects of life is at least possible.

Baffled, we contacted the first lawyer of our client and asked him about this hard-to-believe statement which would easily explain the DNA of the convicted on the dead man's fingers: high stress, sweaty fingers, and primary contact. The lawyer confirmed that from the beginning, this statement was indeed the only one his client had made to him. He, the lawyer, had however requested that the accused remain silent at all times during the trial.

At age 77, in summer 2014, our client was pardoned by the state's governor. Juridically, the man remains the culprit, even though he immediately left prison and was not put on probation.

The focus on DNA during high times of the CSI effect led to a tunnel vision-like view that restricted the classical assessment of criminalistically relevant facts known in the case. It also led to a widespread yet in this case inadequate advice from lawyer to client: to remain silent and deny all (Sello 1911).

Case 2: Sexual Assault

In 2016, a 28-year-old woman reported to the police that she had been raped anally for around 1 hour in her bed in her parent's home by a friend. Ejaculation had not taken place, even though a used towel was found in the bathroom of which the woman suspected it might contain sperm cells. The woman reported that no industrial lubricant was used, but the man spit in his hands and then transferred the saliva to his penis.

She had allowed the man to stay over night in her room because he had been drunk, and both had visited the same party event. The man reported to the police that he slept in the woman's bed but did not touch her sexually. On the following morning, he went to work directly from the woman's (which is also her parent's) home. The woman said she did not call for help during the assault because she did not want to wake up and annoy her grandfather who also stayed in the same house. The woman recorded the event on her cell phone. On the sound file, she is heard saying that the man should stop; this utterance is repeated several times.

After the woman went to a local hospital the next day, four vaginal, two anal, and two oral swabs were taken. Town size is medium which means in Germany around 250,000 inhabitants. Forensic nurses do not exist in Germany even in larger cities, and scientific or medical forensic personnel of any kind was not involved at this stage. However, hematoma and scratches on the thorax, shoulders, and knees of the woman as well as lacerations in the anal area were documented photographically.

Since police did not suspect the man and found the woman's statement incoherent, the State (i.e., not the National) Bureau of Investigations did not receive clean comparison samples from the man. The woman's swabs were checked for human saliva (human α-amylase, dual monoclonal antibody quick strip test; detection limit 1μL human saliva). All swabs were positive for saliva, with the highest amount of saliva on the anal swab. Autosomal short tandem repeat (STR) analysis led to "minute amounts of male DNA exclusively in the anal swab"; all other DNA matched the woman's DNA profile. Neither underpants nor the towel were subjected to analysis. Y-STRs were not checked for unknown reasons even though good success rates even in spermatozoa negative digital (finger) and/or penile cases have been shown (Mcdonald et al. 2014).

The case was transferred to us by the woman in 2020 since no prosecution took place and the case had gone cold. We decided to focus on the slip that was worn by the woman before and after the possible intercourse and which had not been inspected yet. We rubbed different parts of the pantyhose thoroughly, subjected them to STR analysis, and found Y-chromosomal and autosomal STRs that matched a sample of the male's DNA (Fig. 2). The matches were found on the outside seams on the front and back of the woman's underpants as well as on the seams closer to the vaginal area. In total, 12 Y-STR loci matched to the men's DNA, i.e., a control stain that must have come from the man.

It is known that DNA may be transferred accidentally, especially at the waistband area (Breathnach et al. 2016; Murphy et al. 2019; Benecke 2021a). This is either

Fig. 2 Even in disputed or unlikely cases of sexual assault, it is helpful to check the seams of the clothing by tape lifting, cutting, or swabbing (Benecke 2021a, b). Here, we recovered clear Y-STR signals from the panties of the female complainant, whereas swabs from the body orifices during routine analysis in the state laboratory allegedly did not contain cells of the male offender

considered "background," or it may be used in cases in which the clothing was stored unfavorably, and the offender's DNA was indeed transferred by touch.

On the outside of garments, DNA – especially from skin flakes – may be present after sleeping in one bed. Since according to the man, no touching and no movement had taken place, and since he had not been sleeping in the woman's bed in the months before, we tried to check the laboratory reports of the state lab for amounts and levels of DNA. However, those reports were not submitted to us since the case was considered to be closed by the DA's office, i.e., the prosecution.

Our official request as Certified and Sworn In Experts for Biological Stains to receive the remaining swabs which were not needed any more (the case was "closed" from the state's side) was also not complied with. Therefore, the only way to check the DNA levels on the original swabs was made impossible. It is unknown if (and unlikely that) those levels were ever determined since the – wrong – assumption and wrong laboratory result had been that no DNA was present on the pantyhose.

The family also detected a labeling error in the files. The hospital had accidentally switched two swabs. This might explain why "no" saliva of the man was found in the anal swab.

As in Case 1, due to German laws, the family had to pay all legal, laboratory, and expert costs apart from the State Lab's work in this possible rape case. Since the sexual encounter, the woman had to drop her job and is now in treatment for posttraumatic stress disorder. She does not receive support under the Victim Assistance Act because she is not considered to be a victim but a faulty accuser.

Ultimately, it became impossible to determine the course of events on the relevant night, even though the DNA work was performed in a highly standardized, controlled, and certified state laboratory. The structural problem was that routine

procedures (here, autosomal STRs but no Y loci from swabs taken from the orifices of the complainant only) were applied too strictly. The rape case – its stains as well as the case work – should have been taken out of the routine line, and more items and stains should have been examined.

An obstacle in this case was that too less information between the routine laboratory, the police, and the family was exchanged. If this had been the case, samples like the underwear as well as the early laboratory reports that contained data about the relative amounts of DNA could have been more easily analyzed by an independent forensic biology case worker.

Another hurdle in Germany – for all cases that are initially thought to be based on faulty accusations – is that evidence, including swabs and clothing, will not be retained but destroyed soon. This is unfortunate since even very old cases can be worked on with DNA typing as long as the evidence is present (e.g., Louhelainen and Miller 2020).

Conclusion

Statements of victim and relatives might sometimes seem implausible but still contain leads that allow experimental follow-up using further methods, e.g., Y-STRs, peak height control experiments in the lab or in the field, or the simple check if the statement matches the evidence.

In contrast, "common sense" and logical plausibility can never be a concern during the stain examination stage. Thinking should be left to the court at a later stage. Otherwise, as demonstrated above, stains may get lost, become inaccessible, or lead to convenient but unsafe or untrue conclusions.

In other words, only *"when you have excluded the impossible (by experiments), whatever remains, however improbable, must be the truth"* (Doyle 1892).

References

Benecke M (2021a) Forensic DNA samples – collection and handling. In: Rogers R (ed) Encyclopedia of Medical Genomics and Proteomics, 2nd edn (in press). CRC Press/Taylor & Francis, Boca Raton

Benecke M (2021b) A routine rape case that became tricky (and educational), B47. In: Proceedings of the American Academy of Forensic Sciences (AAFS), 73rd meeting, p. 192

Breathnach M, Williams L, McKenna L, Moore E (2016) Probability of detection of DNA deposited by habitual wearer and/or the second individual who touched the garment. Forensic Sci Int Genet 20:53–60. https://doi.org/10.1016/j.fsigen.2015.10.001

Butler J (2014) Advanced topics in forensic DNA typing: interpretation. Academic, New York

Coble M, Hill B (2012) NIST applied genetics group: application of thresholds for interpretation. The Copenhagen forensic genetic summer school advanced topics in STR DNA analysis, June 27–28, 2012. https://strbase.nist.gov/training/8_Mixtures_CFGSS.pdf

Doyle A (1892) The Adventure of the Beryl Coronet. In: The Adventures of Sherlock Holmes. George Newnes, London

Godsey M, Alou M (2011) She blinded me with science: wrongful convictions and the 'Reverse CSI-Effect'. Texas Weleyan Law Rev 17:481–498

Louhelainen J, Miller D (2020) Forensic investigation of a shawl linked to the "Jack the Ripper" murders. J Forensic Sci 65:295–303. https://doi.org/10.1111/1556-4029.14038

Mcdonald A, Jones E, Lewis J, O'Rourke P (2014) Y-STR analysis of digital and/or penile penetration cases with no detected spermatozoa. Forensic Sci Int Genet 15:84–89. https://doi.org/10.1016/j.fsigen.2014.10.015

Murphy C, Kenna J, Flanagan L, Gorman M, Boland C, Ryan J (2019) A study on the background levels of male DNA on underpants worn by females. J Forensic Sci 65:399–405. https://doi.org/10.1111/1556-4029.14198

Reibe S, Benecke M (2010) Der reverse C.S.I.-Effekt. Wenn Spuren nicht beachtet werden. Kriminalistik 64:89–94; 174–179

Sello E (1911) Irrtümer der Strafjustiz und ihre Ursachen. R. von Decker's Verlag/G. Schenck, Königlicher Buchhändler, Berlin

DNA Profiling for Mass Disaster Victim Identification

26

E. V. Soniya and U. Suresh Kumar

Contents

Abstract

DNA profiling is the gold standard for human identification. This technique has been widely used successfully not only in forensic and crime investigations but also in identification of victims in mass disasters. Kerala, southern state of India, witnessed two major disasters in 2016 and 2017, which lead to the loss of many lives. Puttingal temple firework disaster that happened in 2016 resulted in dead bodies scattering all over the place and visual identification of the victims were impossible in many cases. DNA analysis identified 15 of the 17 reported missing persons by comparing DNA profiles obtained from the burned flesh and bone samples collected from the site with the DNA profiles of the relatives of the missing persons. When Ockhi cyclone caused devastation in 2017, unidentifiable decayed dead bodies, a total of 74, recovered from the sea were identified using

E. V. Soniya (✉) · U. Suresh Kumar
Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram, Kerala, India
e-mail: evsoniya@rgcb.res.in

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_24

DNA analysis by comparing with 536 blood samples collected from the relatives of the missing persons. Both these events highlighted the importance of DNA profiling in identification of mass disaster victims.

Keywords

DNA profiling · Disaster victim identification (DVI) · Short tandem repeat (STR) markers · Forensic

Introduction

DNA profiling using short tandem repeat (STR) markers have been successfully used for disaster victim identification (DVI) widely (Leclair et al. 2001; Brenner and Weir 2003; Lau et al. 2005; Tsokos et al. 2006; Nandineni et al. 2010; Hartman et al. 2011; Dhanardhono et al. 2013). It has emerged as a major tool not only for identification of victims but also as an aid in legal procedures, in case there is a need of evidence (Yudianto et al. 2009). Use of DNA for the identification of mass disaster victims dates back to 1990 when the victims of the fire on the Norwegian ferry "Scandinavian Star" was identified by DNA-based restriction fragment length polymorphism (RFLP) using variable number tandem repeat (vNTR) probes. Since then, DNA analysis has been a major tool in disaster victim identification (DVI) cases. Also, the technological advancements enabled faster and accurate identification of victims using DNA than traditional methods like visual, fingerprint, or odontology. DNA profiling was used in DVI cases like Spitsbergen civil aircraft disaster in 1996 (Olaisen et al. 1997), Swiss air flight 111 disaster (Leclair et al. 2001), France Airbus A320 crash (Ludes et al. 1994), Trenggalek shipwreck case (Dhanardhono et al. 2013), the 9/11 World Trade Center attack (Brenner and Weir 2003; Biesecker et al. 2005), Indian Ocean Tsunami (Lau et al. 2005; Bajaj 2005; Tsokos et al. 2006), Australian bushfire disaster (Hartman et al. 2011), Air India Express flight IX812 crash disaster (Nandineni et al. 2010), etc. successfully.

Background

It was Sir Alec Jeffreys who put forward the idea of DNA fingerprinting in 1984. Though humans share 99.9% of the genome, the 0.1% difference make an individual unique. DNA Fingerprinting was based on repeat sequences that occur throughout the genome in large numbers. These repeats, called mini or microsatellites depending on the length of the repeating unit, showed high polymorphism as the number of repeats individuals possesses may vary. These are genetically inherited too, making these regions ideal for relationship testing.

Earlier, DNA fingerprinting was done by restriction fragment length polymorphism (RFLP) method, but as the technological developments enhanced the detection methods, short tandem repeat (STR) analysis by capillary electrophoresis using

fluorescently labeled primers became the most sought-after method (Butler 2007). STR analysis detects repeat regions of 2–6 nucleotides. For STR DNA profiling, polymerase chain reaction (PCR) is used to amplify repeat regions or STR loci from small amounts of DNA. The relatively short PCR product sizes of STR (100 bp–500 bp) enable amplification from even degraded DNA (Butler et al. 2003). This is very important especially in forensic cases where the sources of DNA are exposed to harsh environmental conditions. Use of PCR, capillary electrophoresis, and fluorescent dyes enable PCR amplification of multiple STR loci in a single reaction. This is also very useful when the amount of DNA is very limited. STR analysis is now adapted all over the world as a standard method for DNA profiling.

Process

The complete process of STR analysis includes sample collection, storage and transport, preprocessing, DNA extraction, DNA quantification, PCR amplification of STR loci, capillary electrophoresis for STR allele separation, allele sizing, profile generation, and interpretation of the result.

Sample Collection, Storage, and Transport

Source of DNA for STR analysis can be any biological material of human origin. This can be blood, blood stains, saliva, bone, teeth, nail, skin, muscle, hair, or any part of a human. Personal items used by a person like comb, dress, etc. are also a potential source of DNA as these materials may have epithelial cells or hair from that person. Care should be taken while collection and storage to avoid contamination of DNA from other sources. The samples collected from a scene of crime or from a disaster site should be properly stored depending on the type of sample and distance to the analysis laboratory. DNA can be degraded over time when exposed to varying environmental conditions and storage of source DNA is important to decrease this degradation (Lee et al. 2012). Cold storage or preservatives can be used while collection and transport to the laboratory for analysis.

DNA Preprocessing, Extraction, and Quantification

Samples should be decontaminated if it seems to be soiled or contaminated with exogenous DNA. Hard tissues such as bones can be cleaned with sterile water or 10% bleach. Sodium hypochlorite or 70% ethanol also can be used. This step will minimize co-extraction of PCR inhibitors along with DNA. Efficiency of DNA extraction, i.e., amount of DNA, and removal of PCR inhibitors are two important aspects of DNA processing (Gilbert et al. 2006; von Wurmb-Schwark et al. 2008; Zimmermann et al. 2009). There are a number of different extraction methods and commercial DNA extraction kits available (Witt et al. 2012; Alonso 2013; Norén

et al. 2013). Organic DNA extraction, silica-based column, magnetic bead-based extraction are some of the methods available for DNA extraction. Choice of DNA extraction method or kit should be based on sample type, sample size, efficiency of extraction method, and yield (Verdon et al. 2011; kashkary et al. 2012; Rucinski et al. 2012). Extracted DNA should be checked for quality and quantity using appropriate methods (Nicklas et al. 2012; Goodwin 2017). Good quality DNA without inhibitors is required for efficient amplification of all the STR loci. Optimum amount of DNA is recommended for PCR amplification of STR loci; more amount can result in substrate inhibition, and less in insufficient amplification and dropout of loci.

PCR Amplification of STR Loci

Commercially available PCR amplification kits, which provide premixed primers and master mixes with buffers, dNTPs, and enzyme, simplify STR analysis by providing a uniform set of core STR loci that makes sharing and comparison of DNA profiles nationally and internationally possible (Butler and Hill 2012). In addition to the autosomal STR loci, Y-chromosomal and X-chromosomal STR loci can also be used in analysis when a lineage is needed to be traced or when there is lack of first-degree relatives for comparison (Gill et al. 2001; Diegoli 2015).

Capillary Electrophoresis and DNA Profiling

Capillary electrophoresis is done to determine the size of the PCR products in a genetic analyzer. Capillary electrophoresis requires only a minute amount of the PCR product and the high resolution separate even a single base difference. Allele sizing is achieved with software and DNA profiles are prepared.

Analysis

Depending on the type of case, analysis is done. In a normal paternity case with father, mother, and child, a child inherits 50% DNA from father and 50% from mother, for each locus one allele should match to the father and the other to the mother. In mass disaster cases, putative relative's DNA profiles are compared. In crime investigations, DNA profile obtained from crime scenes are compared with suspects DNA profiles or DNA database if available. Statistical analyses of the matches are conducted for confirmation.

Case Reports

Kerala, southern state of India, witnessed two major disasters in 2016 and 2017, which lead to the loss of hundreds of lives. DNA analysis was used for the identification of the victims of these two disasters by comparing the DNA profile

obtained from the postmortem samples with the DNA profile of the relatives of the reported missing persons, which otherwise was impossible due to the severity of the disasters.

Case 1

In 2016, just before the dawn of 10th of April at the Puttingal Temple, Paravur, Kollam district, firework celebrations went wrong resulting in a huge explosion. More than 100 people lost their lives and many more were injured. Dead bodies were scattered all over the place and identification of the victims were not possible in many cases.

Postmortem samples of victims which were unidentifiable and scattered pieces of burned flesh and bones collected from the site were brought to the DNA lab for DNA profiling. Nineteen persons were reported to be missing initially, of which 2 of them were found alive later. As there were no reference DNA profiles or established human DNA databases for comparison, samples of DNA were collected from close relatives of the missing persons after obtaining informed consent. All the reported missing persons were male and were nonrelatives. Reference samples collected were either the missing person's father, mother, children, or brothers. In some cases where these samples were not available, father's, brother's, or his son's sample were obtained for Y-chromosomal DNA profile matching.

Samples

A total of 169 samples, which included intact bones taken from unidentifiable postmortem samples, partially burnt bone and flesh samples recovered from the site, were brought to the DNA lab. Blood samples from close relatives of the missing persons were obtained on FTA cards at the DNA lab and were used directly in PCR.

DNA Extraction

DNA was extracted from flesh samples using NucleoSpin®Tissue Kit (Macherey-Nagel) following manufacturer's instructions. In brief, approximately 50–100 mg tissue was placed in a microcentrifuge tube, 180 μL Buffer T1 and 25 μL Proteinase K solution was added and incubated at 56 °C for about 3 h or overnight for complete lysis. After incubation, 200 μL buffer B3 was added and incubated at 70 °C for 10 min. To this, 210 μL of absolute ethanol was added and the mixture was transferred to NucleoSpin® Tissue Column for binding. The column was centrifuged at 11,000 g for 1 min, and washed with 500 μL buffer BW first and then with 600 μL buffer B5. DNA was eluted in 50 μL pre-warmed buffer PE.

DNA isolation from bone samples were performed using NucleoSpin® DNA Trace Kit and NucleoSpin® DNA Trace Bone Buffer Set (Macherey-Nagel) following manufacturer's instructions. Before lysis, bone samples were ground to a fine powder using liquid nitrogen. Seven milliliters of buffer T1, 2 mL of 0.5 M EDTA, and 100 μL Proteinase K solution were added and incubated at 56 °C overnight. Afterwards samples were incubated at 4 °C for 48 h. After incubation, 8 mL of buffer B3 was added and incubated at 70 °C for 10 min. To this, 8.4 mL of absolute ethanol

was added and the mixture was transferred to NucleoSpin® Trace F Column for binding. The column was centrifuged at 3000 g for 3 min, and washed with 3 mL buffer BW first and then with 3 mL buffer B5 twice. DNA was eluted in 40 μL pre-warmed buffer BE. DNA quality and approximate quantity was checked using agarose gel electrophoresis.

DNA Profiling

Short tandem repeat (STR) markers have been the gold standard for DNA profiling for years now, especially in the field of forensics. Commercially available autosomal PCR amplification kits (Life Technologies), GlobalFiler™ in case of extracted DNA and GlobalFiler™ Express in case of FTA cards, were utilized to amplify the sex-determining marker amelogenin, Y STR locus DYS391, Y insertion/deletion (Y indel) locus, and 21 autosomal STR loci which included CSF1PO, D12S391, D10S1248, D13S317, D16S539, D18S51, D19S433, D1S1656, D21S11, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D22S1045, FGA, SE33, TH01, TPOX, and vWA. Since all the missing persons were male, Yfiler™ Plus PCR Amplification Kit (Life Technologies), which amplified Y-chromosome specific STR markers DYS438, DYS627, DYS458, DYS437, DYS391, DYS449, DYS392, DYS635 (Y GATA C4), DYS19, DYS390, DYS439, DYS456, DYS393, DYS387S1 a/b, DYS576, DYS460, DYS533, DYS389 I/II, DYS570, DYS385 a/b, DYS481, YGATA H4, DYS518, DYS448, was also used for analysis wherever appropriate.

Amplification was performed on the GeneAmp™ PCR System 9700 Thermal Cycler (Applied Biosystems) following manufacturer's instructions. The amplified PCR products were size-separated by capillary electrophoresis in an Applied Biosystems 3500 Genetic Analyzer (Life Technologies). GeneScan™ – 600 LIZ®, an internal lane size standard, was included with every sample for sizing, and an allelic ladder was used for allele designation by the GeneMapper™ ID-X Software v1.4 (Life Technologies).

Matching of DNA Profiles

DNA profile obtained from victim samples were compared self-to-self first for identifying unique profiles, and then these unique DNA profiles were compared with the DNA profile of the relatives of the missing persons for matching.

Results

In disaster victim identification situations, DNA profiling has become an important tool for identification along with other methods like dental, visual, etc. (Bajaj 2005). As the technological advancements made it cost effective and easily available, it is now considered one of the main tools for identification.

Of the 169 postmortem samples, 8 samples did not yield any DNA profile, both autosomal and Y-chromosomal. Twenty-three samples gave partial profiles, of which 21 samples gave more than 80% alleles with GlobalFiler™ and more than 90% alleles with Yfiler™ Plus. The two profiles with less than 80% alleles were repeated, but were ambiguous and thus not used for further analysis (Table 1). Self-to-self match resulted in 31 unique profiles from the available 159 DNA profiles.

These 31 profiles were compared with the DNA profiles of the relatives of the 17 missing persons reported. When a direct-to-match, i.e., reference sample of the victim or a database for comparison, is not available, kinship analysis is the only way to identify a person. In this case, the victims were all males and were non-relatives, making the comparison/match using Y-chromosomal DNA more plausible (Jobling et al. 1997; Ge et al. 2011; Ballantyne et al. 2014). STR markers used for comparison/matching autosomal, Y-chromosomal, or both are depicted in Fig. 1.

Of the 17 missing persons reported, 15 matches were found in the 31 unique DNA profiles obtained (Tables 2 and 3, Fig. 2). There was no match for the other two. The 16 unmatched DNA profiles may be of those whose bodies were identified by other means and so no reference samples were made available.

Case 2

In 2017 November 30, cyclone Ockhi hit the southern districts of Kerala along with Kanyakumari district of Tamil Nadu causing havoc and destruction. Many lost their lives, majority were fishermen, who went out for fishing in the Arabian Sea and Indian Ocean. During the month of December, decayed bodies were fished out by the search team, which were impossible to identify. Postmortem samples taken from the decayed bodies, mainly sternum or femur, were brought to the DNA lab for DNA profiling. A total of 74 unidentifiable bodies were found. Since the devastation spread a large area, the missing persons account was also large, which was in thousands. In this case also, all the missing persons were male. A total of 536 blood samples were collected on FTA cards from the relatives of the missing persons after obtaining necessary consent.

Both autosomal and Y-chromosomal markers were used where appropriate. In some cases, additional X-chromosome DNA markers were also used for confirmation. Commercially available Investigator® Argus X-12 QS Kit (Qiagen), which amplifies 12 X-STR loci in 4 presumed linkage groups (LG) (LG1: DXS8378–DXS10135–DXS10148; LG2: DXS7132–DXS10074–DXS10079; LG3: DXS10101–DXS10103–HPRTB; LG4: DXS7423–DXS10134–DXS10146) were used for amplification of X-chromosome DNA markers. DNA size standard 550 (BTO) was used for allele sizing.

Table 1 Number and type of samples received with success rate of DNA profiling

Type of sample	Numbers received	Complete profile obtained (%)	Partial profile obtained (%)	Not worked/not usable (%)
Case 1				
Burnt flesh	54	32 (60%)	14 (25%)	8 (15%)
Bone	115	106 (92%)	7 (6%)	2 (2%)
Case 2				
Sternum	32	32 (100%)	–	–
Femur	42	42 (100%)	–	–

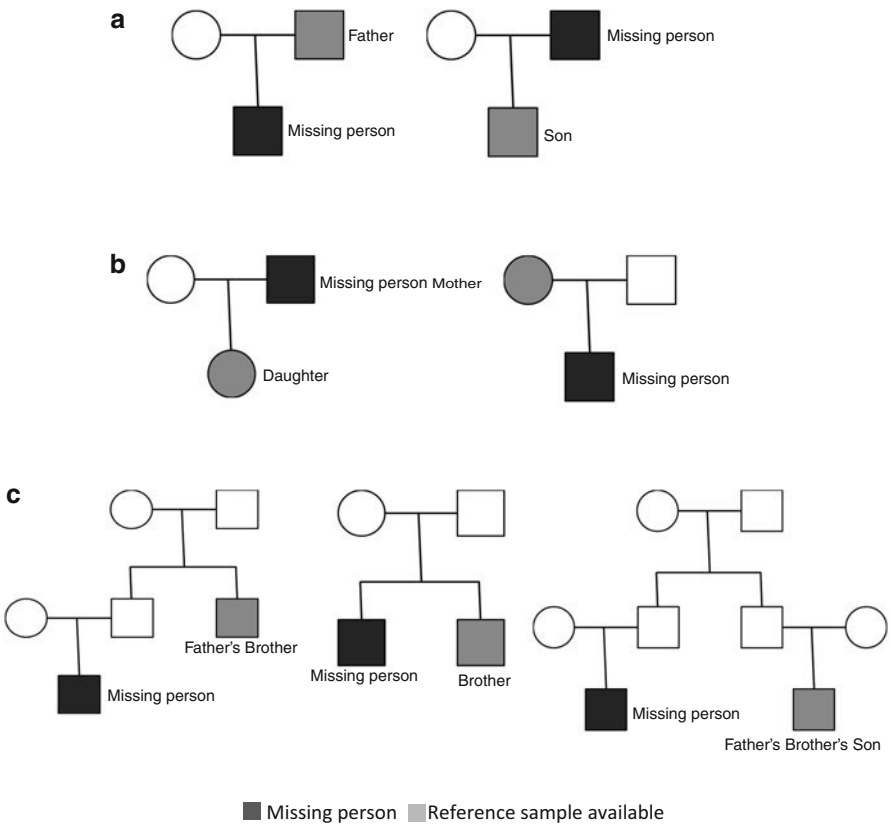


Fig. 1 STR markers used for comparison/matching, autosomal or Y-chromosomal or both; (**a**) Matching done using both autosomal and Y-chromosomal STR markers; (**b**) Matching done using autosomal STR markers alone; (**c**) Matching done using Y-chromosomal STR markers alone

DNA extraction and DNA profiling procedures were as in previous case. Complete DNA profiles were obtained for all the samples. Even though the bodies were started to degrade, samples used for DNA profiling was either sternum or femur bone, which were intact and were good source of DNA. The DNA profiles were then compared with 536 DNA profiles of reference samples provided by the relatives of the missing persons. All the 74 bodies were identified by matching with the reference samples.

Discussion

Both these cases highlight the importance of DNA profiling in disaster victim identification. DNA analysis is considered the gold standard for the identification of human remains and may be the only available method, when other methods like

Table 2 A representative allele sharing table, where the PM sample is compared and matched to the missing person's son in Y-chromosomal markers and autosomal markers

Y marker	PM sample	Reference sample (son)	Autosomal marker	PM sample	Reference sample (son)
DYS576	**19**	**19**	D3S1358	**15**, 17	**15**
DYS389I	**12**	**12**	vWA	14, **15**	**15**, 16
DYS635	**23**	**23**	D16S539	**11**, 12	**11**, 12
DYS389II	**28**	**28**	CSF1PO	**11**	**11**, 12
DYS627	**21**	**21**	TPOX	**8**, 11	**8**, 10
DYS460	**11**	**11**	D8S1179	12, **15**	**15**
DYS458	**15**	**15**	D21S11	**30**, 32.2	**30**, 33.2
DYS19	**14**	**14**	D18S51	12, **19**	13, **19**
YGATAH4	**11**	**11**	D2S441	**11**, 11.3	**11**
DYS448	**19**	**19**	D19S433	**13**	**13**
DYS391	**10**	**10**	TH01	**7**, 9	**7**, 9
DYS456	**16**	**16**	FGA	**20**, 21	**20**, 23
DYS390	**22**	**22**	D22S1045	14, **15**	11, **15**
DYS438	**11**	**11**	D5S818	**11**	**11**
DYS392	**14**	**14**	D13S317	10, **13**	**13**, 13
DYS518	**39**	**39**	D7S820	8, **10**	**10**, 10
DYS570	**16**	**16**	SE33	**18**, 20	**18**, 33.2
DYS437	**15**	**15**	D10S1248	14, **15**	**15**, 15
DYS385	**14**	**14**	D1S1656	**16**	14, **16**
DYS449	**27**	**27**	D12S391	17, **22**	21, **22**
DYS393	**11**	**11**	D2S1338	**16**, 18	**16**, 22
DYS439	**12**	**12**			
DYS481	**23**	**23**			
DYF387S1	**35, 39**	**35, 39**			
DYS533	**12**	**12**			

dental records or fingerprints are not available or irrecoverable. Forensic DNA typing can be used to identify biological samples even when the samples are fragmented and the DNA is degraded (Biesecker et al. 2005). Environmental conditions determine the quality of the DNA that can be typed from these samples. In the first case, mostly burnt samples were the source of DNA, but the quick response of the collection team and laboratory ensured no further degradation and thus complete profiles were obtained from most of the samples. This shows the importance of the time taken from disaster to sample collection and analysis.

In the case of Trenggalek shipwreck case (Dhanardhono et al. 2013), which was also considered an open disaster just like the two case studies, the samples were exposed to different environmental conditions such as high humidity and temperature that accelerated the process of degradation. The environmental conditions to which the samples were exposed had a crucial influence on the investigation and thus in identification. Of the 103 remains found, only 74 yielded DNA and full profiles

Table 3 A representative allele sharing table of Y-chromosomal markers, where the PM sample is compared and matched to the missing person's father's brother. Shared alleles are in bold

Y marker	PM sample	Reference sample (father's brother)
DYS389I	**13**	**13**
DYS635	**25**	**25**
DYS389II	**29**	**29**
DYS627	**21**	**21**
DYS460	**10**	**10**
DYS458	**16**	**16**
DYS19	**14**	**14**
YGATAH4	**12**	**12**
DYS448	**19**	**19**
DYS391	**10**	**10**
DYS456	**17**	**17**
DYS390	**23**	**23**
DYS438	**11**	**11**
DYS392	**10**	**10**
DYS518	**39**	**39**
DYS570	**19**	**19**
DYS437	**15**	**15**
DYS385	**13, 20**	**13, 20**
DYS449	**34**	**34**
DYS393	**15**	**15**
DYS439	**12**	**12**
DYS481	**28**	**28**
DYF387S1	**36, 39**	**36, 39**
DYS533	**11**	**11**

were obtained from 46 samples only. The low success rate was attributed to the extreme environmental conditions such as temperature, high salinity of the sea water, and high humidity that the samples were exposed to in the Indian Ocean. The preferred samples there was Achilles' tendons, whereas we preferred sternum or femur for samples recovered from sea or high salinity water.

In cases where the samples are fragmented like in the firecracker blast case, it will become necessary to sort and reassociate disarticulated remains (Puerto and Tuller 2017) using DNA profiles by self-to-self match and make unique DNA profiles before matching. In the firecracker blast case, 31 unique profiles were obtained from 159 recovered samples representing 31 unique individuals. In disasters such as an air crash or bomb blast, this becomes the first process after initial DNA profiling, as the victim's body parts can be injured and separated. From the Swissair Flight 111 disaster site (Leclair et al. 2001), 1277 samples were recovered and 228 unique genotypes were derived consistent with the flight manifest, thus ensuring the possibility for everyone's identification. In the Taoyuan Airbus crash accident (Hsu et al. 1999), of the 685 fragments of human remains analyzed for DNA, 202 unique profiles were obtained accounting for all the victims.

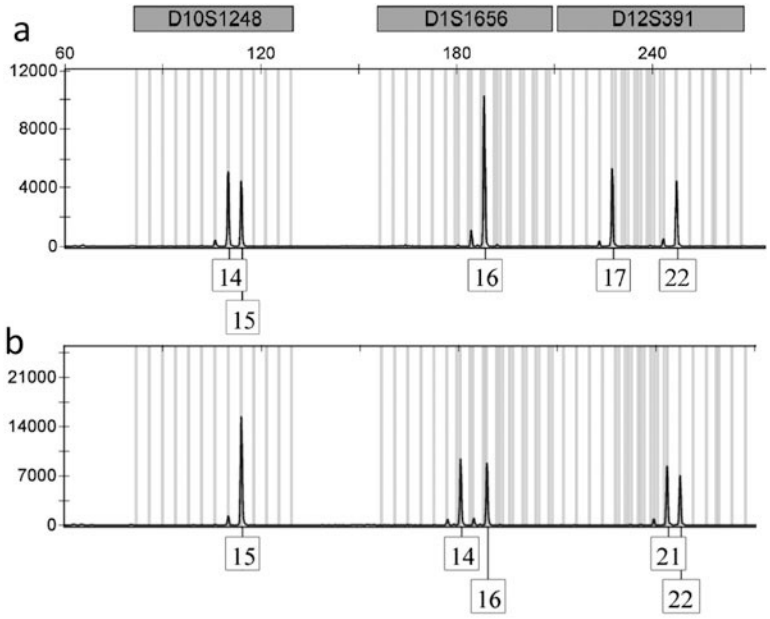


Fig. 2 A representative image of electropherogram showing alleles: (**a**) PM sample and (**b**) Reference sample (son)

It is also important to understand the scale of the disaster as to how many are missing, how many are the actual victims, number of victims recovered, condition of the remains recovered, and degree of fragmentation or mixing, if any (Puerto and Tuller 2017). A disaster is considered an open disaster if the number or identity of the victims involved is not known. The two case studies discussed were open disasters thus making the identification process a challenge. Disasters caused by bush fire, tsunami, blast, etc. fall in to this category. Air crashes are mostly considered a closed disaster event as flight manifests are readily available and victims are known.

Identification of victims involve the matching of DNA profiles obtained from body remains and postmortem samples with reference samples, which can either be direct match samples or familial samples. Direct match is possible when there is a database of DNA profiles available or from DNA separated from personal belongings of the victims or hair samples from combs, etc. The possibility of having DNA database is not very common. In some countries, offender's DNA database is available, which can be helpful only in some cases. In majority of the cases, a readily available database is absent. When DNA can be obtained from personal items, they are compared with the DNA profiles obtained from samples from the site for a full match for identification. But there are limitations for obtaining good DNA profile of a victim as the personal items can get contaminated easily with other family members.

When a direct match is not possible, familial comparison is the most appropriate method. Samples from relatives of the possible victims were used for familial analysis. Selection of relatives as reference samples should also be dealt with carefully as first-degree relatives are preferred (Prinz et al. 2007). The large number of reference samples in the Ockhi cyclone disaster case made the matching cumbersome, though with positive result as all the bodies were matched. Sometimes when more than one member of a family is involved in a disaster, which can occur mainly in air crash accidents, living relatives' samples are used to identify any one of the family members and then this profile is used to identify other family members as in case of the Taoyuan Airbus crash accident (Hsu et al. 1999). In the two case studies, the missing persons were neither related nor first-degree relatives, thus ruling out this possibility for identification.

Conclusion

By using a proper and established protocol, process of disaster victim identification, using DNA profiling along with traditional methods like fingerprints and dental records, can be managed efficiently even when the disaster is of humungous nature. It is important to identify the victims of a disaster considering the certainty of a person living or dead is required for legal purposes related to compensation, insurance, inheritance, and others. New technologies and methods have made DNA analysis the most feasible and easiest method for identification of mass disaster victims, requiring only a short time for analysis. DNA profiling also enable identification from even a small amount of tissue. It can be concluded that DNA profiling has significantly revolutionized the process of identification in mass disaster events.

References

Alonso A (2013) DNA extraction and quantification. In: Encyclopedia of forensic sciences, 2nd edn. Elsevier Ltd, Waltham, pp 214–218

Bajaj A (2005) Disaster victim identification: tsunami. Br Dent J 198(8):504–505

Ballantyne KN, Ralf A, Aboukhalid R, Achakzai NM, Anjos MJ, Ayub Q, Balažic J, Ballantyne J, Ballard DJ, Berger B, Bobillo C (2014) Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. Hum Mutat 35(8):1021–1032

Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, Budowle B, Butler JM, Carmody G, Conneally PM, Duceman B (2005) DNA identifications after the 9/11 World Trade Center attack. Science 310(5751):1122–1123

Brenner CH, Weir BS (2003) Issues and strategies in the DNA identification of World Trade Center victims. Theor Popul Biol 63(3):173–178

Butler JM (2007) Short tandem repeat typing technologies used in human identity testing. Biotechniques 43(4):Sii–SSv

Butler JM, Hill CR (2012) Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Forensic Science Review 24(1):15–26

Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 48(5):1054–1064

Dhanardhono T, Wulandari N, Bhima SKL, Ahmad HJ, Widodo PT (2013) DNA profiling of disaster victim identification in trenggalek shipwreck case. Forensic Sci Int Genet Suppl Ser 4 (1):e5–e6

Diegoli TM (2015) Forensic typing of short tandem repeat markers on the X and Y chromosomes. Forensic Sci Int Genet 18:140–151

Ge J, Budowle B, Chakraborty R (2011) Choosing relatives for DNA identification of missing persons. J Forensic Sci 56:S23–S28

Gilbert MTP, Hansen AJ, Willerslev E, Turner-Walker G, Collins M (2006) Insights into the processes behind the contamination of degraded human teeth and bone samples with exogenous sources of DNA. Int J Osteoarchaeol 16(2):156–164

Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. Int J Legal Med 114(6):305–309

Goodwin WH (2017) The use of forensic DNA analysis in humanitarian forensic action: the development of a set of international standards. Forensic Sci Int 278:221–227

Hartman D, Drummer O, Eckhoff C, Scheffer JW, Stringer P (2011) The contribution of DNA to the disaster victim identification (DVI) effort. Forensic Sci Int 205(1–3):52–58

Hsu CM, Huang NE, Tsai LC, Kao LG, Chao CH, Linacre A et al (1999) Identification of victims of the 1998 Taoyuan Airbus crash accident using DNA analysis. Int J Legal Med 113:43–46

Jobling MA, Pandya ARPITA, Tyler-Smith CHRIS (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110(3):118–124

Kashkary L, Kemp C, Shaw KJ, Greenway GM, Haswell SJ (2012) Improved DNA extraction efficiency from low level cell numbers using a silica monolith based micro fluidic device. Anal Chim Acta 750:127–131

Lau G, Tan WF, Tan PH (2005) After the Indian Ocean tsunami: Singapore's contribution to the international disaster victim identification effort in Thailand. Ann Acad Med Singapore 34(5): 341

Leclair B, Frégeau CJ, Bowen KL, Fourney RM (2001) Enhanced kinship analysis and STR-based DNA typing for human identification in mass fatality incidents: the Swissair flight 111 disaster. J Forensic Sci 49(5):939. JFS2003311-15

Lee SB, Clabaugh KC, Silva B, Odigie KO, Coble MD, Loreille O, Scheible M, Fourney RM, Stevens J, Carmody GR, Parsons TJ (2012) Assessing a novel room temperature DNA storage medium for forensic biological samples. Forensic Sci Int Genet 6(1):31–40

Ludes B, Tracqui A, Pfitzinger H, Kintz P, Levy F, Disteldorf M, Hutt JM, Kaess B, Haag R, Memheld B, Kaempf C (1994) Medico-legal investigations of the Airbus A320 crash upon Mount Ste-Odile, France. J Forensic Sci 39(5):1147–1152

Nandineni MR, Prasad SP, Goud CV, Negi DS, Nagaraju J, Gowrishankar J (2010) DNA-based identification of victims of the Mangalore air crash of May 2010. Curr Sci (00113891) 99 (3):341–342

Nicklas JA, Noreault-Conti T, Buel E (2012) Development of a real-time method to detect DNA degradation in forensic samples. J Forensic Sci 57(2):466–471

Norén L, Hedell R, Ansell R, Hedman J (2013) Purification of crime scene DNA extracts using centrifugal filter devices. Investig Genet 4(1):1–8

Olaisen B, Stenersen M, Mevåg B (1997) Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster. Nat Genet 15(4):402–405

Prinz M, Carracedo A, Mayr WR, Morling N, Parsons TJ, Sajantila A, Scheithauer R, Schmitter H, Schneider PM (2007) DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). Forensic Sci Int Genet 1(1):3–12

Puerto MS, Tuller H (2017) Large-scale forensic investigations into the missing: challenges and considerations. Forensic Sci Int 279:219–228

Rucinski C, Malaver AL, Yunis EJ, Yunis JJ (2012) Comparison of two methods for isolating DNA from human skeletal remains for STR analysis. J Forensic Sci 57(3):706–712

Tsokos M, Lessig R, Grundmann C, Benthaus S, Peschel O (2006) Experiences in tsunami victim identification. Int J Legal Med 120(3):185–187

Verdon TJ, Mitchell RJ, van Oorschot RA (2011) Evaluating the efficiency of DNA extraction methods from different substrates. Forensic Sci Int Genet Suppl Ser 3(1):e93–e94

von Wurmb-Schwark N, Heinrich A, Freudenberg M, Gebühr M, Schwark T (2008) The impact of DNA contamination of bone samples in forensic case analysis and anthropological research. Legal Med 10(3):125–130

Witt S, Neumann J, Zierdt H, Gébel G, Röscheisen C (2012) Establishing a novel automated magnetic bead-based method for the extraction of DNA from a variety of forensic samples. Forensic Sci Int Genet 6(5):539–547

Yudianto A, Sosiawan A, Kusuma SE (2009) IDENTIFICATION OF THE VICTIMS OF KM SENOPATI NUSANTARA USING DNA PROFILING: a case report of disaster victim identification. Folia Med Indonesiana 45(3):225

Zimmermann P, Vollack K, Haak B, Bretthauer M, Jelinski A, Kugler M, Loidl J, Pflug W (2009) Adaptation and evaluation of the PrepFiler™ DNA extraction technology in an automated forensic DNA analysis process with emphasis on DNA yield, inhibitor removal and contamination security. Forensic Sci Int Genet Suppl Ser 2(1):62–63

Potential Use of DNA Profiling in Solving Terrorism Cases

27

Noora R. Al-Snan

Contents

Abstract

Terrorism cases are rarely reviewed in academic literature, due to the significant numbers of explosive (IED) cases in the worldwide; we have reviewed our forensic investigations to explain our findings. The concept of ***Intelligence Forensics DNA*** was developed, and it is defined as the human cells deposited in low copy number in challenging areas within the evidence which leads to significant results. Processing of IEDs in the most efficient way should be done through a dedicated investigative forensics team. With the collaborative effort of the team, touch DNA can be collected and processed to obtain readable and accepted results for the Law Courts. In this study, a total number of four IED cases have been investigated (i.e., directionally focused charges (DFC), directional focused fragmentation charge (DFFC), and explosively formed penetrator/projectiles (EFP)). Also, a total number of five samples that were positive with RDX-C4 explosives have been investigated. DNA recovery utilized different collection methods, such as swabbing, tape lifting, wiping, and direct cutting of

N. R. Al-Snan (✉)
Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Kingdom of Bahrain, Manama, Bahrain
e-mail: nr.alsnan@interior.gov.bh

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_27

samples. Samples were extracted and purified using magnetic beads chemistry and quantified. Low copy DNA extracts were subjected to different concentration steps. DNA extracts were amplified and processed for the detection to obtain reliable results. This chapter will be very useful and informative to assist the forensic community worldwide as terrorists do not respect geographical boundaries nor ethnicities of the victims and the use of DNA profiling technology is the most suitable way to identify the terrorists. Continued efforts must be done to re-evaluate standard operating protocols with empirical studies.

Keywords

Forensics · IED · Intelligence forensics DNA · Kingdom of Bahrain · DNA recovery · Terrorism · DFC · DFFC · EFP · RDX-C4

Introduction

Forensics DNA Intelligence

Most of the traces in crime scenes are comprised of blood spots, semen and sperm secretions, tool marks, trace evidence, and human remain. With the advances in forensic technology, this evidence is increasingly recovered within crime scenes, using the correct equipment, processes, and expertise (Magalhães et al. 2015), as long as the "visible stains" are present, a simple cotton swab can be used to collect and the exhibit processed to solve the mystery.

The terrorism cases have caused many casualties, increasing fear in the population and sometimes death among citizens and first responding police officers. In assessing these forensic cases, the scientist requires a different approach and intellectual way of thinking; one must enter the criminal's mind and visualize the way the terrorist bomb maker assembled the IED and how they deployed it.

The traditional "simple swabbing" procedure is no longer effective. Many cases were misreported neither negative nor simply had extremely high contributors in DNA mixture because of the excessive swabbing area. Improvised explosive devices, known as IED, are used by terrorists worldwide. Touch DNA from bomb assemblers usually yields low success rate due to the inability locate the areas where DNA has been left (Daly et al. 2012).

However, as was learned from Locard's exchange principle "every contact leaves a trace" (Petherick et al. 2009), trace in the guise of DNA usually must be there, somewhere, and somehow, in the form of hidden touch DNA.

Maintaining this golden principle, hundreds of terrorism cases were processed. The investigations revealed that most suspects who were identified using DNA evidence had significantly more serious criminal histories than those identified through traditional property crime investigations (Ritter 2008).

We have defined the ***Forensics DNA Intelligence*** as human cells deposited in low copy number in challenging areas within the evidence which has led to

Fig. 1 Some of processed IEDs of different terrorism cases

significant results (Al Snan and Ghayyath 2019). This concept was achieved by processing a huge number of terrorism cases with different types of samples duly collected which started from year 2015 and continued to date as shown in Fig. 1.

Forensic Investigative Team

Before dealing with securing evidence in terrorist cases, dedicated trained and professional forensics investigative teams (Fig. 2) must be established with the following skills:

- Crime scene examiner
- DNA expert
- Fingerprint expert

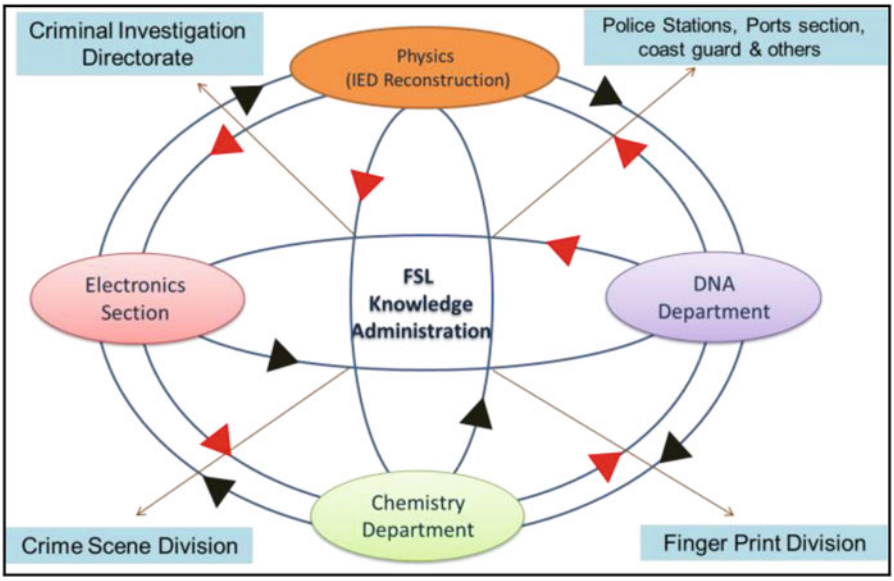


Fig. 2 Diagram of forensics investigative team workflow (FSL: Forensic Science Laboratory)

- Ballistics expert
- Forensic electrical engineer
- Forensic photographer

Each forensic specialist has a significant role in identifying the forensic leads which would assist in the identification of how IEDs and components manufactured or modified and how they are implanted (Al Snan and Ghayyath 2019). To proceed with the forensic collection, several steps must be carried out to achieve the maximum recovery of results:

- Obtain good laboratory practice (GLP), to protect the evidence from contamination or cross-contamination.
- Initiate safety first: Ensure the IED and/or components (improvised detonators) are made safe and modified devices by Explosive Ordnance Disposal (EOD) expertise.
- Detailed description is completed including overall photos of the IED.
- Separate the components of the IED (i.e., wires, devices, detonators, etc.).
- Unseal locked/ closed devices.
- Take detailed photos after opening.

Recovery of Touch DNA from DNA Contaminated with RDX-C4 Explosives

Royal demolition explosive (RDX) is the organic compound with the formula (O2NNCH2)3. It is a white solid material without smell or taste, widely used as an

explosive. Chemically, it is classified as a nitramide and like HMX, an octogen (Field 2017). It is more energetic explosive than TNT, and it was used widely in World War II (Chong et al. 2020).

It is often used in mixtures with other explosives and plasticizers or phlegmatizers (desensitizers) (Al and Colbach 1969). RDX is very stable at room temperature.

It burns rather than explodes. It detonates only with a detonator, being unaffected even by small arms fire. Thus, it is considered one of the most energetic military explosives, and an explosion can only be initiated by a shock wave from a detonator (Field 2017).

RDX is the main component of Composition-4 (C4), which is a plastic explosive used in military operations. C4 is composed of RDX (91%), dioctyl sebacate (5.3%), polyisobutylene (2.1%), and mineral/motor oil (1.6%) (Al and Colbach 1969); the ratios of composition depends on the source of manufacturing. As RDX is the main component of C4, C4 has the explosive nature of RDX and has been utilized due to its malleable nature, which allows it to be molded into any desired shape and redirects the direction of the resulting explosion (Manual 1984).

Terrorists-related crimes have increased over the past 15 years and have become the most common activity around the world besides theft and sexual assault cases (Vision of Humanity 2019).

In every year, terrorists have improved their tools and skills, which emphasizes on the importance of having similar skill level of the law enforcements to reduce the challenges upon examining these types of cases. One of the terrorism activities is to plant remotely controlled explosive devices such as IEDs in different areas to ensure maximum damage to the communities (Gibbons et al. 2003).

The effect of explosive RDX-C4 is very massive and can cause many causalities and fatalities among civilians and policemen. It can penetrate through metals and buildings.

Terrorists do not respect geographical boundaries nor ethnicities of the victims and the uses of DNA profiling technology are the most suitable way to identify the terrorists and keep an end to their violence.

The use of DNA in a criminal investigation such as an incident involving a remotely detonated IED may provide a strong association between a terrorist and the materials used to construct or detonate the device (Rampant 2017).

As the suspected person came into direct contact with the IEDs in terms of manufacturing, assembly, deploying, and transportation, it is more likely that the wanted person(s) will be identified using touch DNA from different components of the IEDs (Rampant 2017).

Many terrorism cases have shown the presence of RDX-C4 in samples such as real IEDs, bombs, pipes, and some packed in bags or wrapped in adhesive film in warehouses.

The estimated number of RDX-C4 cases worldwide is very high, speaking of the studied country as Bahrain, the estimated amount of RDX-C4 between the years of 2015 and 2018 (May) with a total quantity of **370.72** KG, and a total number of 38 cases (Al-Snan 2020).

This dangerously large quantity of RDX-C4 can cause major disaster to the infrastructure and human casualties which we emphasize on the crucial need to

continuously re-evaluate standard operating protocols with empirical studies for such type of cases.

In this chapter, we will display the potential recovery of touch DNA from evidence contaminated with small and large traces of RDX-C4 with no effect of inhibition or degradation of DNA.

Technical Updates

For the first part of the study, a total number of four IEDs cases were identified for detailed analysis from various terrorist incidents (i.e., DFC, DFFC, and EFP) in the kingdom of Bahrain in year 2015.

IEDs components were separated and unsealed to reveal the internal parts for processing along with the external parts of the IEDs. Separation of samples is essential to obtain reliable and beneficial results and to obtain an accurate number of profiles.

In the first case, we have received homemade IED with claymore and phones which was deployed near to a sport club.

Second case, it was a series of IEDs containing black devices linked with wires.

Third case, we received an explosion case in Muharraq area involving a real IED with claymore and phones. Last case, it was a hoax IED containing bundle of tapes and batteries.

Single or double cotton swabs (moistened with DNA grade purified water) were used to collect possible human cells from handles of cylinders, aluminum foils, internal parts of devices, and modified phones (Phetpeng et al. 2015).

Tape lifts and single or double nylon swabs (moistened with DNA grade purified water) were used to collect from adhesive samples, such as long sticky tape.

Moistened wipes (locally sterilized from Kimwipes^{TM} by UV sterilization) were used to collect touch DNA from long wires and large exhibits. Direct cutting of samples was done to sample small pieces of tape endings, small fuses, left over tissues, and wire twists.

For the second part of the study, a total number of five cases were selected and brought to the Forensic Science Laboratory, General directorate of Criminal Investigation and Forensic Science, and Kingdom of Bahrain during the years 2017 and 2018 (Fig. 3). These samples were seized and found within hidden warehouses and some roads, ready to be deployed (Al-Snan 2020).

All samples were made safe by the EOD team by separating the parts of the IEDs and signing the consent that the samples are safe to send to the Forensic Lab. These cases consisted of adhesive films with tapes wrapped around RDX-C4 blocks, black battery, pipes loaded with RDX-C4, black bag which contained RDX-C4, and finally magnetic IEDs, which were used to target specific vehicles. Samples were previously tested positive for bulk analysis of RDX-C4 using DXR Raman Spectrometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA), as shown in Fig. 4 (Al-Snan 2020).

Fig. 3 Some of the processed samples of different terrorism cases, (Al-Snan 2020)

Different collection methods were applied, such as tape lifts, nylons flocked swabs, and direct cutting of small pieces of sample.

Table 1 summarizes the collection method applied. Tape lifts and single or double nylon swabs (moistened with DNA grade purified water) were used to collect from the samples with RDX-C4 such as from handles and zipper of bag, battery body that was part of IED, pipe's opening and from the internal parts of the magnetic (IED).

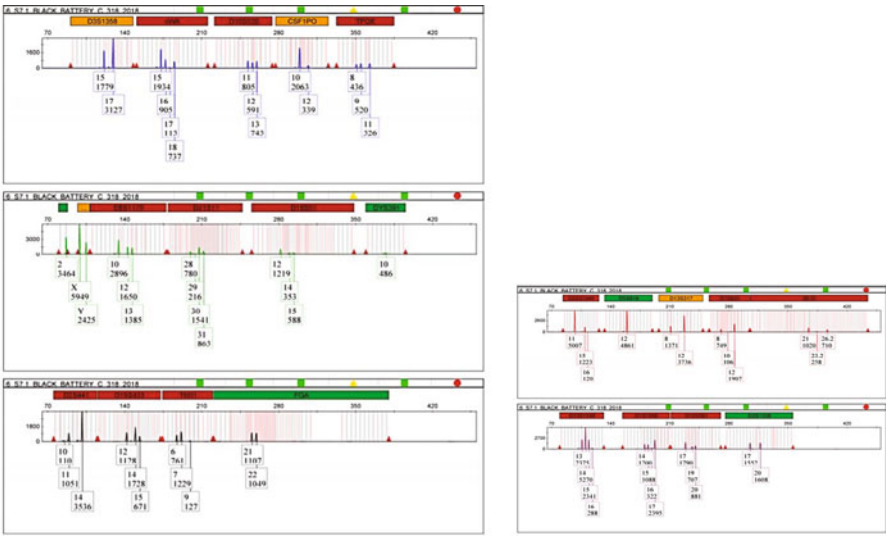


Fig. 4 Chemical analysis of evidence showing 97.94 match with RDX control using DXR Raman Spectrometer (Al-Snan 2020)

Table 1 Summary of the collection method applied to the samples

Collection methods	Samples
Tape lifts	Cloth bag parts inside and outside, battery edges, pipe's opening, internal parts of magnetic IED
Nylons flocked swabs	Cloth bag parts inside and outside, battery edges, pipe's opening, internal parts of magnetic IED, parts of adhesive film around Charge Demolition M112
Direct cutting	Tape endings, wire twists inside the magnetic IED

Direct cutting of samples was done into small pieces of tape endings, wire twists inside of the magnetic (IED) (Phetpeng et al. 2015).

Touch DNA was extracted and purified using the magnetic beads chemistry (i.e., EZ1 Advanced XL, Qiagen, Germany) and AutoMate Express DNA Extraction System, Thermo fisher Scientific, Inc., Waltham, MA, USA) (Davis et al. 2012) with increase time of incubation in EZ1 to 60 min at 56 °C, 850 rpm using 475 μl of undiluted G2 buffer and 25 μl PK. All the samples were added to the Investigator Lyse and Spin tubes (Qiagen, Germany) prior to incubation (Cole 2019).

Quantification was done through Investigator Quantiplex Hyres Kit, Qiagen, Germany, or Quantifiler HP DNA Quantification Kit (Thermo fisher Scientific, Inc., Waltham, MA, USA) using a 7500 Real-Time System (Thermo fisher Scientific, Inc., Waltham, MA, USA) following the manufacturers protocols (Frégeau and Laurin 2015). Most of the samples were subjected to various concentration steps using vacuum dry technique (i.e., using Concentrator Plus – Eppendorf (Müller and

Kache 2005) and using Amicon Ultra-0.5 centrifugal (Garvin and Fritsch 2013) to obtain a reliable quantity for a successful PCR.

About 1.0 ng of the extracted DNA was amplified using GlobalFiler PCR Amplification Kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to manufacturer's recommendation in 29 cycles via MicroAmp Optical 96-Well Reaction Plate (Thermo Fisher Scientific Company, Carlsbad, USA) along with the provided positive control and low TE buffer as a negative control in a 96-Veriti thermal cycler (Thermo Fisher Scientific Company, Carlsbad, USA).

A total of 24 loci were amplified, including 21 autosomal STR loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818, FGA, D12S391, D1S1656, D2S441, D10S1248, D22S1045, and SE33) and three gender determination loci (Amelogenin, Yindel and DYS391).

The PCR products (1 μl) were separated by capillary electrophoresis in an ABI 3500xl Genetic Analyzer (Thermo Fisher Scientific Company, Carlsbad, USA) with reference to LIZ600 size standard v2 (Thermo Fisher Scientific, Inc., Waltham, MA, USA) in total of 9 μl of LIZ600 standard, and Hi-Di formamide (Thermo Fisher Scientific, Inc., Waltham, MA, USA) master mix. GeneMapper ID-X Software v1.4 (Thermo Fisher Scientific, Inc., Waltham, MA, USA) was used for genotype assignment.

DNA typing and assignment of nomenclature were based on the international recommendations. Experiments were performed in the Biology and DNA Forensic Laboratory, Ministry of Interior, Kingdom of Bahrain, which is accredited with CTS testing in regular basis using different quality standards protocols.

Data was captured by 3500 Series Data Collection v3.1. The raw data was then analyzed using GeneMapper ID-X v1.4. RFU values were obtained through in-house validation of the GlobalFiler PCR Amplification Kit with an average of 85 for each dye. Single source samples were checked into the Bahrain DNA database containing ~90,000 DNA STR profiles.

DNA mixtures of 2–3 contributors are supported with Likelihood values using LRmix Studio v2.1 available online (Haned and Gill 2011, Haned and de Jong 2016).

Conclusions

Part One: Forensic DNA Intelligence and IEDs

Based upon our expertise in the forensic terrorism field and from 2011 until 2014, most of the DNA samples collected from the external parts of the IEDs were either negative or generated mixtures containing many numbers of contributors and hence unfit for proper interpretation.

During that period only, a small numbers of academic articles had been published regarding the retrieval of touch DNA in terrorism cases, and some of these academic papers were focusing upon touch DNA retrieval from methods of transferring IEDs in bags or backpacks (Hoffmann et al. 2012). Some of the references applied mtDNA

for the exploded IED parts (Foran et al. 2009). However, due to the loss of individuality power and the long processing time of the mtDNA testing, this test is used mainly for the purpose of research studies.

We have received and processed **2** blasts in 2011, **48** IEDs in 2012, **65** IEDs in 2013, **381** IEDs in 2014, **375** IEDs in 2015, **242** IEDs in 2016, and **76** IEDs in 2017 (up to April 2017), as an average of one IED sample per day (c2014), which indicates the need of establishing advanced protocols to thoroughly investigate these terrorism cases.

In 2015, and with the continuous experience of investigating these IEDs and modified terrorism exhibits, we have followed the above-described pathway using a dedicated and experienced investigative forensics team.

The quantity of extracted touch DNA ranged from 0.025 to 0.09 ng using 7500 Real Time System (Applied Biosystems) prior to DNA concentration and amplification. Any less value of 0.025 ng was considered as negative and unfit for proper interpretation.

The average RFU values obtained using the GeneMapper ID-X (GMID-X) software v1.4 (Life Technologies) were estimated to be around 400–1000.

In the four IED case studies analyzed, two of the cases were negatively reported from the external parts: one DFC case having a claymore, wires, and modified phones and the other case of DFFC with modified transmitter devices and wires.

However, while processing the internal parts of the IEDs, it generated one single profile from the modified phone linked with the DFC (piece of facial tissue) and the other one generated from a piece of tape hidden inside the wire cover linked with the DFFC with modified transmitter devices (Table 2).

Regarding the other two cases, one EFP case with highly number of contributors' mixture in a tape bundle and the other DFC case of two single profiles from the tape on the claymore and the other profile generated from the keyboard of the modified phone. Concerning these two cases, the EFP case with a tape bundle we have found a fuse with bite marks and it generated a single male profile.

Table 2 Description of results obtained from the first and second processed IEDs

IED case	Type	Samples collected	Results	IED Case	Type	Samples collected	Results
1	DFC	Swabs from external claymore and modified phone and wipe from wires.	Neg.	**2**	DFFC	Swabs from external parts of the modified transmitters and wipes from wires	Neg.
		Internal part of phone: small left-over facial tissue	Pos. (fit)			Direct cut of the discovered hidden tape under the wire cover	Pos. (fit)

In the other DFC case, we processed the internal parts of the phone (which gave a profile from the keyboard part) and found out a small tissue paper which generated another single male profile (Table 3).

It was remarkably successful in getting touch DNA with single source profile or mixtures having 2–3 that allowed positive forensic interpretation and led to positive HITS in the DNA Database. Establishing the principle of ***Forensics DNA Intelligence*** leads to two main goals as DNA is not only a human identification tool but also explains exactly why DNA is there (i.e., proves or in fact disproves a suspect's role) through separating the components of IEDs, unsealing electrical devices, and focusing upon modified items:

- Is DNA found in internal or external parts of the evidence?
- Is the terrorist a distributor?
- Is the terrorist an assembler of IED?
- Is the terrorist an electrician?

Once we learn about the role of the suspect (s), it will assist the criminal investigators in their investigation. Where the forensic investigation secures numbers of single unknown profiles, this also assists the investigation in estimating the member size of the terrorist group or organization.

Many IEDs recovered from crime scenes are complex hoax devices; however, these devices cause the same impact of fear among citizens. Nevertheless, same procedure must be followed for any type of IEDs received as many of the hoax IEDs are used in conjunction with live explosive devices and ambush police responders who are dealing with a hoax IED and the live device detonated as a secondary incident, designed to kill the police first responders.

Part Two: The Effect of RDX-C4 on the Recovery of Touch DNA

Full profiles were generated from the different exhibits containing RDX-C4 as shown in (Figs. 5, 6, 7, 8, and 9) (Al-Snan 2020). The RFU of the samples were

Table 3 Description of results obtained from the third and fourth IEDs

IED case	Type	Samples collected	Results	IED case	Type	Samples collected	Results
3	DFC	Direct tape from claymore and swabs from keyboard of modified phones	Pos. (fit)	**4**	EFP	Swabs from tape bundle	Pos. (high mix – unfit)
		Internal part of phone: small left-over facial tissue	Pos. (fit)			Direct cut of a fuse with bite marks	Pos. (fit)

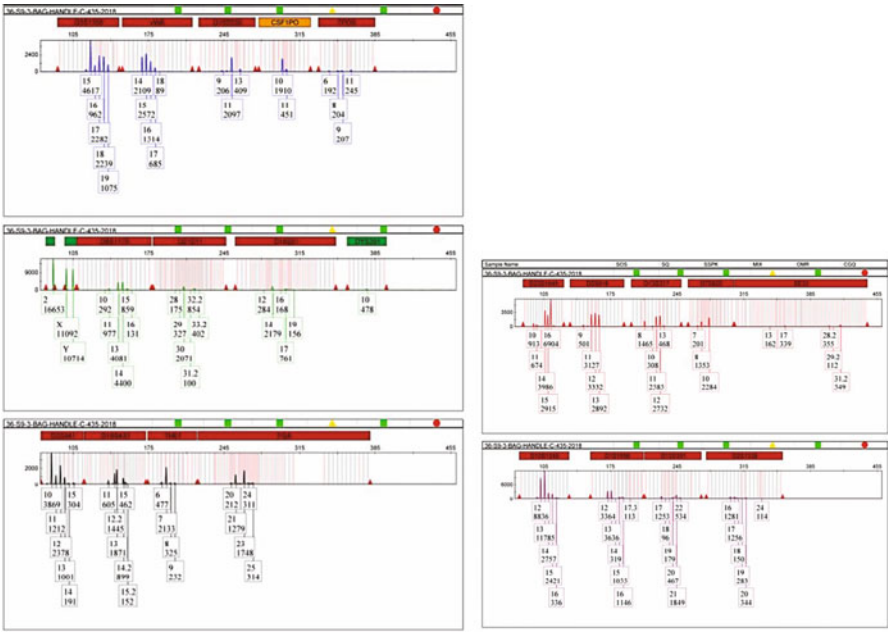


Fig. 5 The results obtained from handles of black bag contaminated with RDX-C4 showed DNA mixtures (0.75 ng/µl) (Al-Snan 2020)

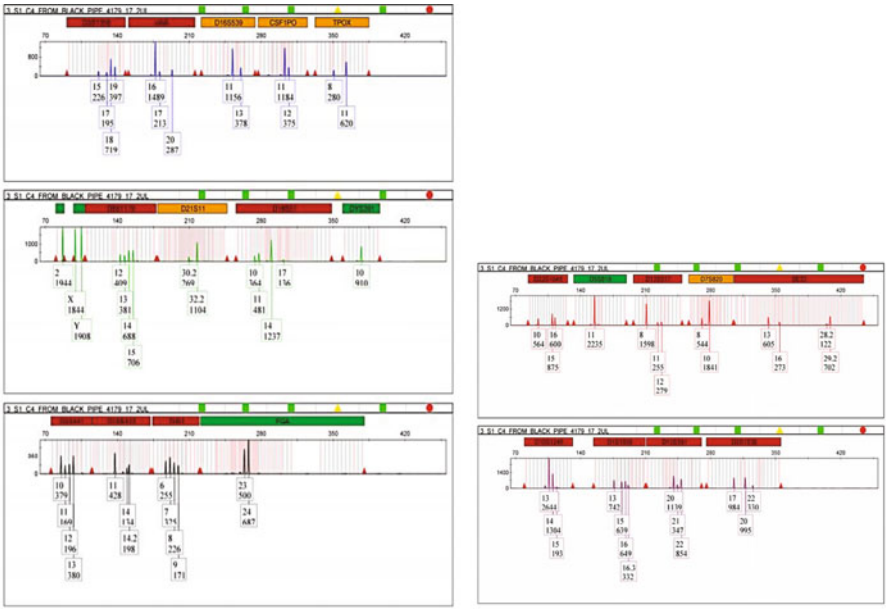


Fig. 6 The results obtained from RDX-C4 inside the black pipe showed DNA mixtures (0.001 ng/µl) (Al-Snan 2020)

Fig. 7 The results obtained from RDX-C4 inside the magnetic IED showed DNA mixtures (00.01 ng/μl) (Al-Snan 2020)

Fig. 8 The results obtained from tape on the Demolition Charge M112 (external surface) contaminated with RDX-C4 showed major contributor of single-source male DNA (0.75 ng/μl) (Al-Snan 2020)

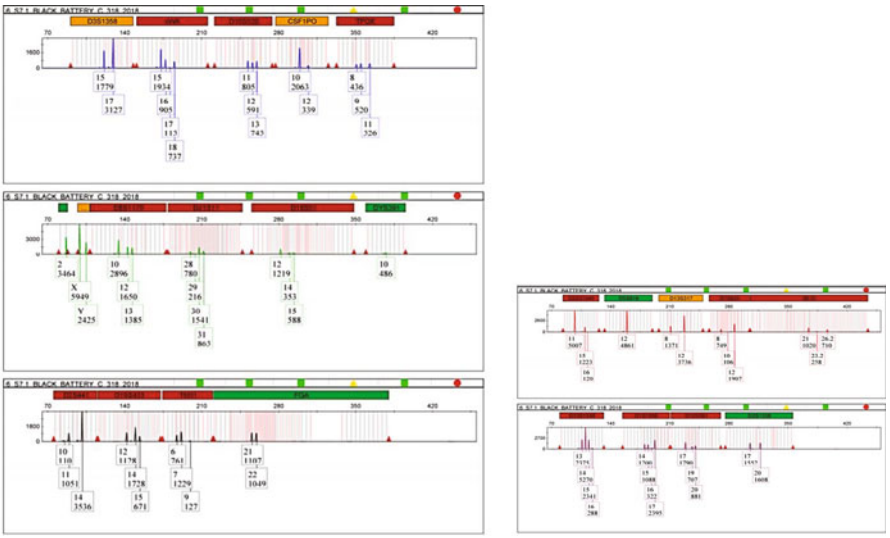


Fig. 9 The results obtained from black battery (external surface) contaminated with RDX-C4 showed DNA mixtures (0.75 ng/µl) (Al-Snan 2020)

Table 4 Quantity of retrieved DNA from RDX-C4 evidence

Samples	Quantity of DNA (ng/ul)
DNA retrieved from handles of black bag contaminated with RDX-C4	0.75
DNA retrieved from RDX-C4 inside the black pipe	0.001
DNA retrieved from RDX-C4 inside the magnetic IED	0.001
DNA retrieved from tape found on the Demolition Charge M112 (external surface) contaminated with RDX-C4	0.75
DNA retrieved from black battery (external surface) contaminated with RDX-C4	0.75

acceptable and fit for interpretation as well as approved by our internal validation for GlobalFiler PCR Amplification Kit (~85 for each loci).

All the samples generated were of DNA mixtures except for the tape wrapped around the adhesive film of RDX-C4 Charge Demolition M112. The quantity of retrieved DNA was displayed in Table 4.

Many terrorism tactics have recently been developed, from Molotov bottles to handmade local weapons and different types of IEDs. These threats need to have optimized way of identifying through specialized forensic investigative team (Al Snan and Ghayyath 2019).

Thus, each forensic specialist has a significant role in identifying the forensic leads which would assist in the identification of how IEDs and components were manufactured or modified and how they were used. RDX-C4 sample is mostly kept

for chemical analysis and most of the labs neglect the crucial need to swabbing or tape lifting the RDX-C4, as RDX-C4 is a sticky solid substance that can retain some of the cells from the shedders or from sweating while assembling the explosive inside compartments of IEDs or bombs. Also, C4 is very stable and insensitive to most physical shocks and can withstand different physical properties such as welding and molding thru metals and electronics (Nagy 2018). Thus, collecting DNA from RDX-C4 sample will give a forensic lead to directly identify the suspect (s) who manufactured the IEDs (Al-Snan 2020).

We have also discovered that C4 cannot bind to the DNA nor to the solutions used in the protocols. Thus, it does not cause inhibition or degradation to the DNA. From this point of view, we were successful in obtaining acceptable and fit results using the above-described methods (Al-Snan 2020).

This chapter is particularly useful and informative to assist the forensic community in terrorism cases applications worldwide. The findings of this study emphasize the need to continuously re-evaluate standard operating protocols with empirical studies for such type of cases.

References

Al H, Colbach E (1969) Composition C-4 induced seizures; a report of five cases. Mil Med 134(13): 1529–1521

Al Snan NR, Ghayyath MA (2019) Potential use of touch DNA in terrorism cases: a report of four cases. Arab J Forensic Sci Forensic Med 1(9):1267

Al-Snan NR (2020) The recovery of touch DNA from RDX-C4 evidences. Int J Legal Med 135:393–397 https://doi.org/10.1007/s00414-020-02407-9

Chong S, Long B, Maddry JK, Bebarta VS, Ng P (2020) Acute C4 ingestion and toxicity: presentation and management. Cureus 12(3):e7294. Published 2020 Mar 16. https://doi.org/10.7759/cureus.7294

Cole KA (2019) Validation of a DNA extraction method using diluted ATL as an extraction buffer with the Qiagen Lyse & Spin Basket kit. Boston University

Daly DJ, Murphy C, McDermott SD (2012) The transfer of touch DNA from hands to glass, fabric and wood. Forensic Sci Int Genet 6(1):41–46

Davis CP, King JL, Budowle B, Eisenberg AJ, Turnbough MA (2012) Extraction platform evaluations: a comparison of AutoMate Express™, EZ1® Advanced XL, and Maxwell® 16 Bench-top DNA extraction systems. Legal Med 14(1):36–39

Field S (2017) Boom!: the chemistry and history of explosives. Chicago Review Press, Chicago

Foran DR, Gehring ME, Stallworth SE (2009) The recovery and analysis of mitochondrial DNA from exploded pipe bombs. J Forensic Sci 54(1):90–94

Frégeau CJ, Laurin N (2015) The Qiagen Investigator® Quantiplex HYres as an alternative kit for DNA quantification. Forensic Sci Int Genet 16:148–162

Garvin AM, Fritsch A (2013) Purifying and concentrating genomic DNA from mock forensic samples using Millipore Amicon filters. J Forensic Sci 58:S173–S175

Gibbons A, Farrier J, Key S (2003) The pipe bomb: a modern terrorist weapon. BMJ Mil Health 149(1):23–26

Haned H, de Jong J (2016) LRmix Studio 2.1 User Manual

Haned H, Gill P (2011) Analysis of complex DNA mixtures using the Forensim package. Forensic Sci Int Genet Suppl Ser 3(1):e79–e80

Hoffmann SG, Stallworth SE, Foran DR (2012) Investigative studies into the recovery of DNA from improvised explosive device containers. J Forensic Sci 57(3):602–609

Magalhães T, Dinis-Oliveira RJ, Silva B, Corte-Real F, Nuno Vieira D (2015) Biological evidence management for DNA analysis in cases of sexual assault. Sci World J 2015
Manual Shore Protection (1984) Department of the Army. US Army Corps of Engineers
Müller H, Kache H (2005) Laboratory centrifuge, comprising refrigeration unit. Google Patents
Nagy B (2018) Grosse point blank microwave C4 mercury switch. Carnegie Mellon University
Petherick W, Turvey BE, Ferguson CE (2009) Forensic criminology. Academic
Phetpeng S, Kitpipit T, Thanakiatkrai P (2015) Systematic study for DNA recovery and profiling from common IED substrates: from laboratory to casework. Forensic Sci Int Genet 17:53–60
Rampant S (2017) The effects of a detonation explosion on the recovery of DNA from fragments of an improvised explosive device. Murdoch University
Ritter N (2008) DNA solves property crimes (but are we ready for that?). National Institute of Justice
Vision of Humanity (2019) Global Peace Index and Positive Peace. Retrieved 06 June 2020, from http://visionofhumanity.org/

DNA Profiling in Forensic Odontology

28

Pooja Puri, Mayank Kumar Dubey, and Naresh Kumar

Contents

Abstract

Right to identity is the fundamental human right, and the extensive efforts have continuously been made to identify unknown human remains after mass disasters like tsunami, cyclones, earthquakes, etc. as well as wars where a large number of unidentified human remains are encountered. Not only this, many times victims of homicide killed brutally with face disfigurement, fire accidents, acid attacks, etc. which cannot be identified from physiognomic characteristics also require identification. In such cases, the forensic analyses are often complicated by sample degradation due to exposure to harsh environmental conditions. DNA

P. Puri (✉)
Amity Institute of Forensic Sciences, Amity University, Noida, India
e-mail: pmalik1@amity.edu

M. K. Dubey
Senior Faculty-Forensic Science Department, Mody University, Laxmangarh, Rajasthan, India

N. Kumar
Home Department, Government of NCT of Delhi, New Delhi, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_63

profiling has become a standard tool for identification of unknown remains in forensic laboratories. The distinctive structure of teeth and their location in the jawbone make them a preferred source of DNA as compared to bones and other tissues. This chapter outlines the role of DNA profiling in forensic odontology.

Keywords

DNA profiling · Human identification · Forensic odontology · Teeth

Introduction

Forensic odontology or forensic odonto-stomatology is a fascinating discipline. It is involves the correct collection, management, interpretation, evaluation, and presentation of dental evidence for criminal or civil legal proceedings: a combination of various aspects of the dental, scientific, and legal professions (Hinchliffe 2011). A forensic odontologist assists the law enforcement agencies by examining teeth and other oral tissues in numerous situations. Though the scope of forensic odontology is well-defined in the dental field, the rational application lies largely in the hands of forensic experts and their interaction with dentist and is pivotal in its ideal application in legal issues. There have been many technical innovations in medical science since the last century that include radiography, novel dental materials, and instrumental and molecular techniques that have revolutionized this field enabling many archived cases to be re-examined (David and Lewis 2018).

The forensic field involves the use of science and technology in discovery and investigation of crime and administration of justice, through the organized efforts of a multidisciplinary team involving specialists in the forensic medicine, forensic odontology, police, and lawyers. The applications of forensic odontology range from age estimation and bite mark comparison to identification of unknown individuals. Dental identification remains one of the most reliable and frequently applied methods for the identification by the comparisons of antemortem and postmortem records. The field of forensic odontology has developed tremendously, and many novel methods for establishing the identity of victims are being assessed, particularly in the cases of disasters, bomb blasts, or air crashes where the human bodies are damaged beyond identification.

Identification Through Teeth

Identification of individuals in the society has both legal and emotional needs, and it becomes imperative when the former aspect is involved (Brown 1984; Buchner 1985). While visual recognition is cited as the most common method of identification, whether it is a case of living or dead, physiognomic characteristics have demonstrated their significance in innumerable instances.

In situations where physiognomic characteristics including facial features can't be used, fingerprints may be helpful, but with the need of existence of standards on the police records. However, in cases involving advanced decomposition of the dead body, incinerated remains, drowning for long period of time, etc., neither of these methods can be satisfactorily used due to obscured features (Table 1). Under most of the severe environmental conditions, the teeth are least destructible part of the body, and they may readily survive all these changes, and hence DNA is preserved. The use of teeth for yielding useful information in such cases has been well documented (Petju et al. 2007).

When the condition of the remains is deemed inappropriate for visual recognition, identification using a scientific technique for determining identity is required. Scientific or primary methods include comparison of fingerprint patterns, medical and dental records, and analysis of DNA. Scientific methods are frequently described as being either comparative, where high-quality pre-death (antemortem) records exist for use in the comparison, or reconstructive, where no antemortem information exists for comparison but unique features in the dentition can be used to generate a postmortem dental dossier. Both these methods have certain limitations and cannot be applied in certain instances (Table 2); then DNA profiling from the dental tissues can be done.

The human tooth consists of two anatomical parts: a root and a crown. The outermost covering of the tooth crown is called enamel. Enamel is the strongest tissue in the body (Nanci and Tencate 2012). Dentine lies beneath the enamel and forms the major bulk of the tooth. The pulp is the soft tissue and is made up of fibroblasts and undifferentiated mesenchymal cells, odontoblasts, endothelial cells, peripheral nerve, and nucleated components of blood which are rich sources of DNA. Cementum is the calcified tissue covering the outermost part of the root. The

Table 1 Potential cases requiring dental DNA profiling

Dental DNA profiling can be applied in
Unidentified dead body with or without decomposition
Incinerated remains
Ancient DNA analysis from archaeological sites
Skeletonized remains
Disaster victim identification
Drowned bodies

Table 2 Limitations of comparative and reconstructive identification

Roadblocks in comparative/reconstructive identification
Unavailability of dental records
Incomplete dental records/records between the treatments not updated
Internationalization of dental treatment
Lack of uniform nomenclature for maintaining records
Lack of awareness/training among the dentists

researches have shown that the pulp is the richest source of DNA. Not only pulp, dentin and cementum too contain enough DNA to be used in the human identification (Malaver and Yunis 2003).

Pötsch et al. (1992) were able to obtain the genomic DNA dental pulp tissue sample that varied between 6 μg and 50 μg in quantity and did not show any major variation when compared to the array obtained from DNA isolated from blood samples.

Applications of DNA in Forensic Odontology

Before the 1980s, the science establishing the identity mainly relied on the serological probes like enzyme and protein typing, blood grouping, and some genetic markers, and many of these modalities has a low degree of discrimination, and, in addition, their use was limited because of low stability in biologically degraded forensic samples (Budowle and van Daal 2008). But with the advent of DNA profiling, establishing the human identity testing using DNA typing methods has been widespread. The past three decades have seen tremendous growth in the use of DNA evidence in crime scene investigations. Today several public and private forensic laboratories in the United States conduct hundreds of thousands of DNA tests every year. In addition, most countries in Europe and Asia and Africa have forensic DNA testing facilities. The number of laboratories around the world conducting DNA testing continues to grow as the technique gains popularity in the law enforcement (Butler 2014), since the DNA profile tests are absolutely consistent as well as admissible in courts throughout the world, such as for investigation of paternity and human identification (Pötsch et al. 1992).

Although the extraction of DNA from the mutilated tissues is quite challenging task, the studies have shown that DNA from hard tissues like bone and teeth (Table 3) are stable even after putrefaction of bodies, and the researchers have successfully extracted DNA from teeth that had been buried for 80 years (Boles et al. 1995). Apart from establishing the identification, another area of study of forensic odontology related to DNA is the analysis of bite mark evidence. In cases of sexual assaults, homicides, and child abuse, bite marks are frequently found evidence on the skin and sometimes even on inanimate objects like fruits, cheese, etc. in the cases of burglary. The perpetrator's saliva is usually deposited on the victim's skin during biting. From the cells present in saliva, it is also possible to isolate DNA for identification (da Silva et al. 2007).

Table 3 The main aim of DNA profiling in forensic odontology (Ziętkiewicz et al. 2011)

DNA profiling in Forensic Odontology is done for:
Identification of victims
Identification of criminals
Co-relating the body parts in case of dismembered bodies or commingled remains

Another potential application of DNA in forensic odontology may be mapping of salivary microbiome as type of bacteria present in saliva can be employed for individual identification. The study of salivary components along with bacteria may give a clue toward health status, lifestyle of individuals, or any disease along with their age and sex. It has been observed that there may be significant variations in the oral microbiome in the certain disease conditions, viz., Behcet's syndrome, aphthous stomatitis, inflammatory bowel disease (IBD), rheumatoid arthritis, hepatic cirrhosis, etc. (Fuentes and Sanchez 2017; Takayasu et al. 2017.)

In the practical scenarios, it has been observed that there are several samples which are encountered at the scene of crime or sent to a Forensic Science Laboratory for DNA Profiling from teeth/oral tissues. Some of them are:

1. In a case where the cadaver/body have decomposed to an extent that tissues can't be used for DNA extraction, one or two teeth can be extracted and sent for DNA profiling.
2. If a skeleton is found which was exposed to harsh environmental conditions for a long time.
3. A mandible/maxilla with teeth can be found and sent to laboratory.
4. One or more avulsed teeth can be found at a crime scene associated with a physical assault or torture.
5. In mass disaster cases like floods, tsunamis, air crashes, and explosions where body is found completely decomposed, charred, or fragmented.
6. In exhumed bodies where there is considerable time gap between the time since death and exhumation, teeth are considered the evidence of choice for DNA extraction and profiling.
7. Mandibular canines are the last teeth to shed with age, and hence in most of the cases where a skull is recovered, it is present as valuable evidence which can be used for DNA extraction.
8. Also, in sexual assault cases, swabs are taken from the bite marks and can be used to generate a full genetic profile of the offender.

Handling and Processing of Tooth Samples

The main purpose of molecular analysis in forensic odontology is to isolate good-quality DNA from the tooth pulp, amplify the DNA, and match the profile prepared from the isolated DNA with the DNA profile generated from the samples of nearest blood relations. When the teeth are encountered as forensic samples, they are usually cleaned with sterilized water without using any abrasive/bleach, soap, etc. As it is regular practice to take a photograph of the forensic evidences to be sampled, a photograph of tooth/teeth with a scale and case number must be taken prior to sampling. The type of tooth sample (incisor/canine/premolar/molar, etc.) that is to be collected for DNA testing is dependent upon the suitability and availability of and the condition of preservation of the human remains. It is the responsibility of the forensic scientist to determine the suitable sample(s) for collection. The molars are

the most preferred teeth. Whole and well-preserved teeth with completely formed root apices should be collected ideally. However, the existence of completely formed root apices is not a prerequisite. The teeth with the carious lesions or dental restorations should be avoided. The sample is preserved in normal saline and may or may not be refrigerated, depending on conditions and availability of resources.

Extraction of Dental DNA

It is the well-known fact that the nuclear or the genomic DNA is found in the nucleus of each cell in the human body and represents a DNA source for most forensic applications. As discussed above, the teeth are an excellent source of genomic DNA because PCR analyses allow matching the postmortem samples to known antemortem samples or parental DNA.

Mitochondrial DNA is another type of DNA that can be used for identification. Its main advantage is the high number of copies per cell. When the extracted DNA samples are too small or degraded, such as those obtained from skeletonized or decomposed tissues, the probability of obtaining a DNA profile from mitochondrial DNA is higher than that with any marker found in genomic DNA as it can survive environments where nuclear DNA does not and can therefore be a powerful tool for human identification.

Although there are several techniques proposed in the literature for sampling of dental DNA (Table 4), cryogenic grinding is one of the frequently used technique.

Cryogenic grinding technique involves cooling the whole tooth to extremely low temperatures using liquid nitrogen and mechanically grinding it to fine powder to get enough DNA (intact, carious, as well as root-filled teeth) (Sweet and Hildebrand 1998).

Although this technique is widely followed in almost all the forensic science laboratories, it has a disadvantage that the entire tooth gets completely crushed and the tooth morphology is destroyed. Also, such sample has calcium in high amount leading to inhibition in DNA amplification. A Dremel (Fig. 1a) is used to clean the sample (Fig. 1b) tooth by removing the stains followed by washing with Tergazyme solution for 1 h; washed tooth is kept for overnight drying in a biosafety cabinet ensuring proper flow of air. Tooth sample is then washed with 70% ethanol twice for 1 h, and again it is kept for overnight drying in a biosafety cabinet. The cleaned tooth

Table 4 Techniques used for dental DNA sampling

Techniques for Dental DNA Sampling
Conventional Endodontic Access
Cryogenic Grinding
Horizontal Sectioning of the tooth
Vertical Splitting
Orthograde Entrance technique
Crushing Complete tooth

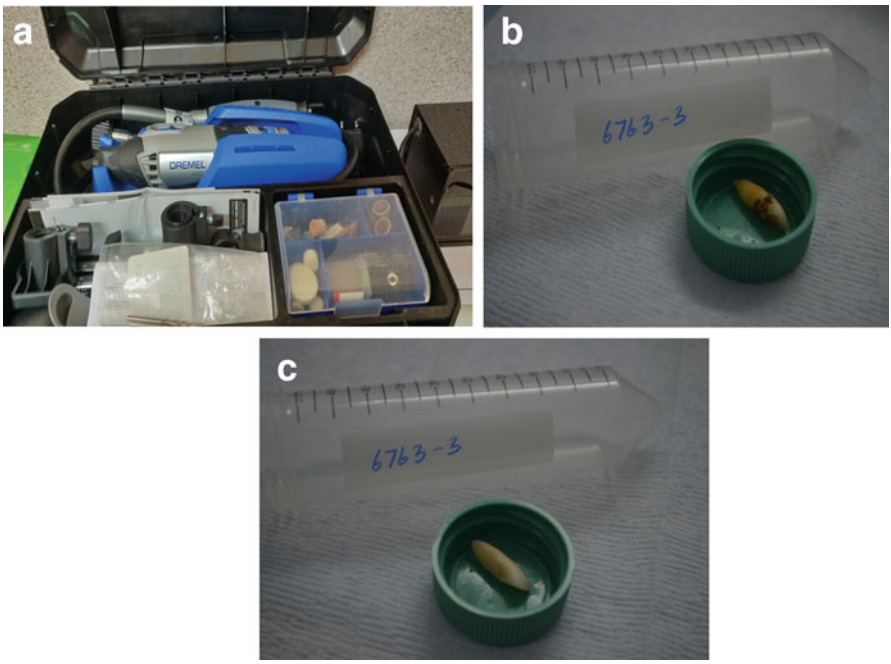


Fig. 1 (**a**) Dremel, (**b**) Sample before cleaning, (**c**) Sample after cleaning

(Fig. 1c) is powdered using freezer mill which works very efficiently and uses liquid nitrogen to maintain extremely low temperature. Care is taken to select the portion of tooth with maximum number of cells so that good quantity of DNA can be extracted. 50 mg of the above powder is taken for the lysis.

Vertical and horizontal sectioning of the tooth is method of choice in many cases. These techniques permit easy removal of the entire pulp tissue with minimal chance of contamination. Splitting is done with carborundum discs, with heavy-duty gloves, preferably under a hood, as aerosol may contact the skin or eyes or can be inhaled. The tooth is mounted in dental stone or may be held tightly. The carborundum discs are used to split the tooth with frequent washing with distilled water. When the pulp cavity is reached, the tooth may be opened by a chisel to avoid mechanical as well as thermal damage to the pulp tissue. The pulp is then excavated and transferred to vials (Muruganandhan and Sivakumar 2011).

Once the pulp is excavated, the next step is the extraction of DNA. The DNA extraction normally consists of three different steps: cell lysis (for rupture of the cell membranes), protein denaturation and inactivation, and finally DNA extraction itself. Lysis is done using PrepFiler™ BTA Lysis buffer where 230 µl of lysis buffer is added to the sample in Bone and Tooth Lysate Tube followed by 7 µl of Proteinase K and 3 µl of DTT (dithiothreitol). Screw the cap on the Bone and Tooth Lysate Tube, vortex it for 5 s, and then centrifuge it briefly. This complete mixture in the tube is kept on a thermal shaker for 12–18 h at 56 °C and 1100 RPM. After the cell

lysis is complete, centrifuge the Bone and Tooth Lysate Tube for 90 s at 10,000 × *g* and transfer the clear (no sediment) lysate to a new PrepFiler™ Sample Tube. With this clear lysate sample, proceed directly to the automated extraction run. Extraction can be done using Automate Express or manually (organic phenol-chloroform extraction or other methods discussed below) depending upon the availability of reagents/equipment).

The techniques of DNA extraction most often employed in forensic odontology are the organic method (phenol-chloroform extraction), Chelex 100 (ion-exchanging resin), FTA Paper, isopropyl alcohol, and silica based. Each of the techniques has its own advantages and disadvantages. The choice of method solely depends upon the forensic examiner and subject to availability of resources. The phenol-chloroform DNA extraction is a gold standard method and widely accepted in genomic research. It yields high molecular weight DNA but involves the use of hazardous chemicals.

Phenol-Chloroform/Organic Extraction Procedure

Step A

1. Collect the sample in a 1.5 ml Eppendorf tube.
2. Add 500 μl of DNA extraction buffer.
3. Add Proteinase K (PK) and DTT according to sample volume and sample type.
4. Vortex, incubate at 56 °C (water bath).

Step B

1. To the lysate add 500 μl of phenol, chloroform, and isoamyl alcohol mixture consisting of equal parts of equilibrated **phenol** and **chloroform/isoamyl alcohol** (24:1).
2. Vortex it for 5–10 s.
3. Keep it for spiral motion of fluid (10 min).
4. Centrifuge it at 1200 RPM for 10 min.
5. Pipette out the upper (aqueous layer) into a new centrifuge tube.
6. Discard the lower (organic layer) and the tube.
7. To the aqueous layer in the new tube, add 500 μl of phenol, chloroform, and isoamyl alcohol mixture.
8. Repeat the steps 1–6 again.
9. Add 100 μl of 2 M sodium acetate and 1000 μl of chilled ethanol.
10. Mix and refrigerate at −20 °C overnight.

Step C

1. Take out the microcentrifuge tube and centrifuge at 14000 RPM for 15 min.
2. Discard the supernatant and keep the pellet portion.
3. Add 500 μl of 70% ethanol. Centrifuge at10000 RPM for 05 min.

4. Discard the supernatant again and add 70% ethanol; centrifuge at 10000 RPM for 05 min (03 ethanol wash is given using 70% ethanol).
5. Discard the supernatant.
6. Air-dry the pellet by keeping the lid of centrifuge tube open in water bath rack/thermal mixer for 30 min.
7. To the pellet add 15–30 μl of TE (Tris-EDTA) buffer or elution buffer.
8. Incubate at 56 °C for 10–15 min.
9. Store the extracted DNA in a freezer for quantification/PCR.

Chelex 100 is a method of choice for extraction of DNA from trace amounts of biological samples as available in forensic cases. Although the method is fast and yields high quantity of DNA, it may not be able to remove some of the inhibitors (Singh et al. 2018); FTA Paper are the specialized papers or cards which can be used to collect the salivary samples from the suspect for molecular analyses. One of the major advantage of the FTA cards is that the samples remain stable for years at room temperature and the integrity is maintained (Beckett et al. 2008), and finally isopropyl alcohol containing ammonium and isopropanol can be used, which is less expensive and also a substitute to the organic method, but produces excellent results for mitochondrial DNA extraction (da Silva et al. 2007).

Alternatively, the extraction can also be performed using Automate Express depending upon the availability of reagents/equipment. Bench-top automated systems ease out the downstream analysis and have the advantages of low-throughput, hands-free operation, and clean processing of evidence samples (Fig. 2). These automated systems are based on magnetic beads and have become more common

Fig. 2 Automate Express Equipment

in the forensic laboratories in the last couple of years. This technique relies on DNA binding to the magnetic beads in an appropriate ionic charge of the environmental surroundings; all unbound contaminants are washed away.

DNA Extraction of Teeth (Using Automate Express)

Reagents and materials required:

- Automate Express kits – PrepFiler BTA
- Heating block
- Centrifuge
- 95% ethanol
- 70% ethanol
- Tergazyme (1% solution)
- Falcon tubes

Procedure

1. The washing/cleaning of the outer surface of the tooth/bone (exhibit) is highly critical.
2. The exhibit is washed for 1 h in a falcon tube with Tergazyme solution (1% solution).
3. After 1 h, the detergent is drained out, and the sample is left to be air-dried overnight inside a biosafety cabinet.
4. There are multiple washes that the same undergoes on the second day.
5. Washing with 70% ethanol for 1 h.
6. Followed by a repeated was with 70% ethanol for 1 h.
7. The next wash is done with 95% ethanol for 1 h.
8. Drain out all the ethanol and let the exhibit dry in overnight inside a biosafety cabinet.
9. On the third day, using Stryker or Dremel, powder the exhibit.
10. Weigh at least 50 mg of each exhibit.
11. After powdering, we go ahead with lysis.
12. Preheat the dry bath or heating block to 56 °C.
13. Place the powdered exhibit in a screw cap tube.
14. Add 250 μL of the PrepFiler BTA Lysis Buffer, 3 μL of 1 M DTT, and 7 μL of Proteinase K to each of the samples.
15. Place on the hot block for 2 h or overnight (18 h) incubation with continuous shaking at 1100 rpm and 56 °C.
16. Once the time is complete, centrifuge the contents using a spin column. The lysate must be collected in the bottom half of the tube.
17. Switch on the Automate Express. Ensure the smart card in in place before you switch on the Automate.
18. Based on the number of exhibits you are extracting at one time, place that many cartridges (pre-made cartridges with magnetic beads) on the cartridge rack.

19. Place the samples with the lysates (at least 300 μL lysate should be there) in the row.
20. Label the elution tubes with the exhibit number/codes, and place them in the row E, with the caps open and facing outward.
21. Place the plunger-tip/holders in the T2 rack.
22. Once everything is in place, start the automated extraction process by following the displayed instructions.
23. Choose the elution volume based on the quantity of DNA you expect (30 μL).
24. Start the protocol. This should take about 30 min (maximum of 13 exhibits at one time).
25. Once extraction is complete, open the Automate Express door, close all the elution tube before removing from the rack.
26. The used tips, holders, cartridges, tubes, etc. must be discarded after every run.

Do's

1. Use only the tips and holders and tubes and cartridges that are provided in the kit.
2. Extract only one type of samples in one run.
3. Follow all the steps for washing the exhibit.
4. Powder the exhibit inside a closed chamber.
5. Wear a facemask while powdering the exhibit.
6. Always handle the Dremel and Stryker with extreme precaution.

Don'ts

1. Do not open the Automate Express while the program is running.
2. Do not use used or expired cartridges for DNA extraction.
3. Do not use 100% ethanol for washing the exhibit.

The techniques for DNA isolation from maxilla or mandible bone (when submitted as forensic evidences) are essentially the same as that for teeth. Crushing is mostly done. Bone marrow extraction may also be performed.

After extracting DNA, an accurate measurement of the amount of DNA and the quality of the DNA extract is desirable. Real-time PCR technique is used for quantification of isolated DNA. Adding the correct amount of DNA to a PCR will produce the best quality results in the shortest time; adding too much or not enough DNA will result in a profile that is difficult or even impossible to interpret. This is especially important when profiling forensic samples, when it is very difficult to know the state of preservation of the biological material and, in many cases, it is difficult to estimate how much cellular material has been collected. It is less important to quantify DNA when using some reference samples – where similar amounts of DNA can be expected to be extracted each time as there are not very many variables. Even so, many laboratories will still quantify the DNA from reference samples as part of their standard analysis. In response to the importance of quantification of samples recovered from the scene of crime, the DNA Advisory Board in the United States adopted rules that made quantification of human DNA mandatory. The PCR-based techniques require as little as 500 picograms; most

PCR-based tests are set up to use between 1 and 10 nanograms of genomic DNA. On the other hand, the RFLP usually needs at least 50 nanograms of intact high molecular-grade DNA.

The quantity of DNA that can be extracted from a sample depends very much on the type of material. Each nucleated cell contains approximately 6 pg of DNA: liquid blood contains 5000–10,000 nucleated blood cells per ml; semen contains on average 66 million spermatozoa per ml (the average ejaculation produces 2.75 ml of semen). Biological samples recovered from the scene of crime are not usually in good condition and can often consist of a very small number of shed epithelial cells; consequently, the amount of DNA that can be recovered can be extremely low and difficult to quantify specially in the cases where we look for DNA from trace materials. In case of the teeth, the total recovery of genomic DNA obtained ranged from 6 to 50ug DNA (Pötsch et al. 1992). The successful recovery of trace or contact DNA is highly variable. It is seemingly dependent on a wide range of factors, from the characteristics of the donor, substrate, and environment to the delay between contact and recovery (Raymond et al. 2009).

Mapping with the molecular markers: A molecular marker is a gene or DNA sequence with a known location on a chromosome that can be used to identify individuals or species. Markers used for human identity testing are found in the non-coding regions either between genes or within genes. The characteristics of different markers currently employed in individual identification are discussed elaborately in literature. In forensic cases, the most applied methodology for individual identification relies on STR (short tandem repeat) profiling. As its name indicates, an STR contains repeating units of a short (typically three- to four-nucleotide) DNA sequence and has been used since 1997 in the FBI Laboratory. STR loci dominate the area of genetic markers in humans (Butler 2014). 13 autosomal STR loci (TPOX, D3S1358, D5S818, FGA, CSF1PO, D7S820, D8S1179, TH01, VWA, D13S317, D16S539, D18S51, and D21S11) and the sex-differentiating amelogenin locus AMEL-X/Y were initially used as the system used for forensic human identification. As of now, 24 marker PCR kit is being used worldwide which consists of 21 autosomal markers (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391, and D2S1338) and 03 sex markers are Amelogenin, Y indel, and DYS391. The likelihood of a person matched to the actual offender having a perfect match for all 13 is typically around 1 in 1 billion (Norrgard 2008). SNP (single-nucleotide polymorphism) markers offer a useful and increasingly important extension to a routine STR-based DNA profiling. SNPs are the most common type of variation in genome involving the single base pair. The advantages of using SNPs in forensic casework and population genetic studies lie in their abundance in the genome – approximately 85% of the human genetic variation can be attributed to SNPs. In contrast to the STRs, SNP markers have comparatively lower mutation rates and are thus more stable, making them more suitable to use in kinship/pedigree studies (Pereira and Gusmão 2013).

Over the last two decades, mtDNA typing has been widely used to solve several human identification-related issues in forensics. In the degraded or compromised

forensic samples, often nuclear DNA is insufficient to generate valuable outcome with current STR typing methods. The teeth are an excellent source for high molecular weight mt DNA that can be crucial for the identification of such degraded or decomposed human remains (Ginther et al. 1992). In contrast to the nuclear DNA, the mtDNA is entirely inherited from mother, due to which not only mutation but the mtDNA sequence of siblings and all maternal relatives is identical (Hutchison et al. 1974). Higgins et al. (2015) studied about the preservation and degradation of nuclear and mt DNA in postmortem teeth and concluded that nuclear and mitochondrial DNA is not distributed evenly throughout teeth and decays at different rates in different dental tissues. Also, mtDNA degrades at a slower rate than nuclear DNA. Further, it is reported that the cementum is a better source of mtDNA than dentine (Adler et al. 2011).

Recent Technologies in Human Identification

The microarray techniques (commonly known as DNA chip or biochip technique) due to the unique ability to screen for large numbers of molecules have been used for use of SNP detection and genotyping. The microarray techniques have not only been used in mutation detection, assessment of gene copy number, comparative genome hybridization, drug discovery, and expression analysis but in individual identification also (Li et al. 2006). A standard microarray experiment involves the hybridization of an mRNA molecule to the DNA template from which it is originated. Many DNA samples are involved in construction of an array. The amount of mRNA bound to each site on the array specifies the expression level of the various genes. Whole data is collected, and a profile is generated for gene expression in the cell. Harada et al. (2009) have studied differences in gene expressions between human dental pulp cells from deciduous and permanent teeth by DNA microarray assays. Not only teeth, the DNA microarray technology has been successfully used to study the various samples from oral cavity, viz., saliva, plaque, etc. to study the microbiome as well as detection of various diseases (Parolin et al. 2017; Kuo et al. 2003).

Next-generation sequencing (NGS), a parallel sequencing technology, has revolutionized the forensic human identification, and these advances in molecular technologies have outpaced the various techniques that were initially used in DNA typing. NGS technology, with its high-throughput capacity, speed, and comparatively low cost, has become an indispensable analytical tool for forensic geneticist. Next-generation genome sequencing technologies available are Illumina Sequencing, Roche 454 genome sequencer, and Proton Sequencing. Not only human identification, NGS technology also has been successfully used in many other areas of forensic studies, viz., database construction, ancestry inference, body fluid and species identification, etc. (Yang et al. 2014). Research has shown that the jaw bone samples and teeth with varying postmortem intervals (PMIs) and degraded DNA samples can yield enough quantity and quality of DNA using DFK technique and reliable genetic profile (Verogen 2020; Inostroza et al. 2020).

Analysis of Y Chromosome

Amelogenin (AMEL) is the major matrix protein secreted by the ameloblasts of the enamel and the AMEL coding gene is located at 11.31 Mb the X chromosome and at 6.74 Mb on the Y chromosomes (short arm). Females (XX) have two identical AMEL genes, but the males (XY) have two nonidentical genes. Therefore, by generating the DNA profile from teeth, the forensic experts obtained 100% success in determining sex of the individual (Nakahori et al. 1991). PCR amplification using the amelogenin gene primers is useful in sex identification of forensic samples. Using other markers on the Y chromosome can help verify male DNA samples (e.g., DYS391, Y-indel). Both the genes are present on Y chromosomes in human males. Any peak if obtained in a DNA profile on these markers confirms the source of the DNA from a male. As Y chromosome genes are inherited from father to son without any alteration, paternal and brotherhood linkage can be established using Y-filer profiling. If the sample contains a mixture of Male and Female DNA as in the case may be in sexual assaults male to female DNA ratio can also be determined. The amelogenin (AMEL) gene, which exists on both X and Y chromosomes, has been used to determine the sex humans (Sullivan et al. 1993). Also, the sex determination based on analysis of Y chromosomal DNA by amplification of specific alphoid centromeric repeat sequence using polymerase chain reaction (PCR) has yielded more accurate results. The X chromosome-specific alphoid repeat sequence can also be detected along with the Y chromosome-specific alphoid repeat sequence (Morikawa et al. 2011).

Hatshepsut, Great Queen of Ancient Egypt

Hatshepsut, one of the great queens of ancient Egypt, who reigned in the fifteenth century B.C. was identified by DNA from her tooth. The mummy, which was discovered way back in 1903, was not identified until 2007 when Discovery Channel supported the project undertaken by the Egyptian Museum, Cairo (Egyptian Mummy Identified as Legendary Hatshepsut, 2020) (Smithsonian Magazine 2020).

Hatshepsut was one of the Egypt's greatest woman pharaoh who reigned for around two decades. When her husband died, she ruled as regent on behalf of his infant son Thutmose III, but effectively took over the throne. Her tomb was discovered by British archaeologist, Howard Carter, while excavating at the Valley of the Kings. The suspicion about the identity began when a body of an overweight woman was recovered lying on the floor in somebody else's tomb, with one arm bent over her chest in the position. CT scan was used to study the physical characteristics of the mummies and the contents of a wooden box found in a different location with Hatshepsut's name on it. Inside the box, some of the queen's internal organs and a single tooth were found (Riaud 2016).

A series of DNA tests were conducted on the body from least destroyed areas of the body including the tooth and compared the DNA samples with those taken from

Hatshepsut's royal relatives – her grandmother Ahmose Nefertari and her father Thutmose I. and the preliminary results suggested a match between the two (Angelique 2007).

Identification of Carbonized Body Through Dental DNA

In another case report by Sweet and Sweet 1995, a burnt body of female was found at the secondary crime scene. The high temperatures produced from the fire induced using the gasoline incinerated the body completely. The body was destroyed beyond the scope of visual recognition, and it precluded the chances of the DNA analysis from conventional sites. The dental pulp proved to be an excellent source of high molecular weight genomic DNA in this case which was well preserved in the unerupted third molar and led to the identification of the victim.

Conclusion

The advent of DNA profiling has transformed the concept of human identification. DNA profiles have been successfully generated from the various dental tissues and saliva. It is reasonable to expect that future advances in DNA technology will assist identification of unidentified bodies by economic, rapid and more sensitive technology. Meanwhile, comparative identification using dental records remains the gold standard for forensic odontology.

References

Adler C, Haak W, Donlon D, Cooper A, Consortium, T (2011) Survival and recovery of DNA from ancient teeth and bones. J Archaeol Sci 38:956–964

Angelique C (2007) University helps with pharaoh's identity. Br Dent J 203(3):123–123

Beckett SM, Laughton SJ, Pozza LD, McCowage GB, Marshall G, Cohn RJ, Milne E, Ashton LJ (2008) Buccal swabs and treated cards: methodological considerations for molecular epidemiologic studies examining pediatric populations. Am J Epidemiol 167(10):1260–1267

Boles TC, Snow CC, Stover E (1995) Forensic DNA testing on skeletal remains from mass graves: a pilot project in Guatemala. J Forensic Sci 40(3):13786J

Brown KA (1984) Dental identification of unknown bodies. Ann Acad Med Singap 13(1):3–7

Buchner A (1985) The identification of human remains. Int Dent J 35(4):307–311

Budowle B, van Daal A (2008) Forensically relevant SNP classes. BioTechniques 44(5):603–610

Butler J (2014) Advanced topics in forensic DNA typing. Academic

da Silva RHA, Sales-Peres A, de Oliveira RN, de Oliveira FT, Sales-Peres SH d C (2007) Use of DNA technology in forensic dentistry. J Appl Oral Sci [Online] 15(3):156–161. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4327460/

David TJ, Lewis JM (2018) Forensic odontology: principles and practice. Academic, London

Fuentes A, Sanchez C (2017) New trends in biomarkers and diseases research: an overview, 1st edn. Bentham Science Publishers, pp 436–468

Ginther C, Issel T, King M (1992) Identifying individuals by sequencing mitochondrial DNA from teeth. Nat Genet 2(2):135–138

Harada R, Watanabe K, Shirasu S, Kato M, Daito M (2009) DNA microarray analysis of dental pulp fibroblasts exfoliated from deciduous teeth. Pediatr Dent J 19(1):38–45
Higgins D, Rohrlach AB, Kaidonis J, Townsend G, Austin JJ (2015) Differential nuclear and mitochondrial DNA preservation in post-mortem teeth with implications for forensic and ancient DNA studies. PLoS One, [Online] 10(5):e0126935
Hinchliffe J (2011) Forensic odontology, Part 1. Dental identification. Br Dent J 210(5):219–224. https://doi.org/10.1038/sj.bdj.2011.146
Hutchison C, Newbold J, Potter S, Edgell M (1974) Maternal inheritance of mammalian mitochondrial DNA. Nature 251(5475):536–538
Inostroza C, Carrasco P, Godoy M, Gatti G, Paulino B (2020) Dental forensic kit (DFK®) and quick extract™ FFPE DNA extraction kit, a new workflow for obtaining dental DNA for human genetic identity. J Forensic Legal Med 73:101992
Kuo WP, Whipple ME, Jenssen T-K, Todd R, Epstein JB, Ohno-Machado L, Sonis ST, Park PJ (2003) Microarrays and clinical dentistry. J Am Dent Assoc 134(4):456–462
Li L, Li C, Li R, Liu Y, Lin Y, Que T, Sun M, Li Y (2006) SNP genotyping by multiplex amplification and microarrays assay for forensic application. Forensic Sci Int 162(1–3):74–79
Malaver P, Yunis J (2003) Different dental tissues as source of DNA for human identification in forensic cases. Croat Med J 44(3):306–309
Morikawa T, Yamamoto Y, Miyaishi S (2011) A new method for sex determination based on detection of SRY, STS and amelogenin gene regions with simultaneous amplification of their homologous sequences by a multiplex PCR. Acta Med Okayama 65(2):113–122
Muruganandhan J, Sivakumar G (2011) Practical aspects of DNA-based forensic studies in dentistry. J Forensic Dent Sci 3(1):38
Nakahori Y, Hamano K, Iwaya M, Nakagome Y (1991) Sex identification by polymerase chain reaction using X-Y homologous primer. Am J Med Genet 39(4):472–473
Nanci A, Tencate A (2012) Ten cate's oral histology. Mosby Elsevier, St. Louis
Norrgard K (2008) Forensics, DNA fingerprinting, and CODIS. Nat Educ 1(1):35
Parolin C, Giordani B, Ñahui Palomino RA, Biagi E, Severgnini M, Consolandi C, Caredda G, Storelli S, Strohmenger L, Vitali B (2017) Design and validation of a DNA-microarray for phylogenetic analysis of bacterial communities in different oral samples and dental implants. Sci Rep 7(1):6280
Pereira V, Gusmão L (2013) Encyclopedia of forensic sciences, 2nd edn. Academic Press
Petju M, Suteerayongprasert A, Thongpud R, Hassiri K (2007) Importance of dental records for victim identification following the Indian Ocean tsunami disaster in Thailand. Public Health [Online] 121(4):251–257
Pötsch L, Meyer U, Rothschild S, Schneider PM, Rittner C (1992) Application of DNA techniques for identification using human dental pulp as a source of DNA. Int J Legal Med 105(3):139–143
Raymond J, Roland A, Oorschot V, Gunn P, Walsh S, Roux C (2009) Trace evidence characteristics of DNA: a preliminary investigation of the persistence of DNA at crime scenes. Forensic Sci Int Genet 4(1):26–33
Riaud X (2016) The dental identification of the Egyptian Queen Hatshepsut. J Dent Oral Care Med 2(1). https://doi.org/10.15744/2454-3276.2.108
Singh UA, Kumari M, Iyengar S (2018) Method for improving the quality of genomic DNA obtained from minute quantities of tissue and blood samples using Chelex 100 resin. Biol Proc Online [Online] 20(1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5984428/
Smithsonian Magazine (2020) Egyptian Mummy identified as legendary Hatshepsut. [Online] https://www.smithsonianmag.com/science-nature/egyptian-mummy-identified-as-legendary-hatshepsut-180940772/. Accessed 8 Dec 2020
Sullivan K, Mannucci A, Kimpton C, Gill, l P. (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. BioTechniques 15(4): 640–641
Sweet D, Hildebrand D (1998) Recovery of DNA from human teeth by cryogenic grinding. J Forensic Sci 43(6):14385J

Sweet D, Sweet C (1995) DNA analysis of dental pulp to link incinerated remains of homicide victim to crime scene. J Forensic Sci 40(2):15365J

Takayasu L, Suda W, Takanashi K, Iioka E, Kurokawa R, Shindo C, Hattori Y, Yamashita N, Nishijima S, Oshima K, Hattori M (2017) Circadian oscillations of microbial and functional composition in the human salivary microbiome. DNA Res 24(3):261–270

Verogen I (2020) Next-generation sequencing workflow for severely degraded DNA. [Online] News-Medical.net. https://www.news-medical.net/whitepaper/20200211/Parallel-Sequencing-Workflow-for-Severely-Degraded-DNA.aspx. Accessed 8 Dec 2020

Yang Y, Xie B, Yan J (2014) Application of next-generation sequencing technology in forensic science. Genomics Proteomics Bioinformatics [Online] 12(5):190–197. https://www.sciencedirect.com/science/article/pii/S1672022914001053

Ziętkiewicz E, Witt M, Daca P, Żebracka-Gala J, Goniewicz M, Jarząb B, Witt M (2011) Current genetic methodologies in the identification of disaster victims and in forensic analysis. J Appl Genet [Online] 53(1):41–60. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3265735/

Part IV

DNA Analysis in Disease Diagnosis

Prenatal Diagnosis of Genetic Disorders by DNA Profiling

29

Inusha Panigrahi and Priyanka Srivastava

Contents

Abstract

Several genetic disorders can have a variable age of onset from newborn period to adulthood. Some of the disorders cause significant morbidity and mortality. Definitive therapy is not available for many disorders of childhood onset. Some of the common genetic disorders of childhood onset include Down syndrome, thalassemia, hemophilia, congenital adrenal hyperplasia, Gaucher disease, and cystic fibrosis. Prenatal diagnosis offers hope in prevention of early-onset disorders associated with significant morbidity and mortality including single or multisystem involvement. Preimplantation diagnosis is also an option especially

I. Panigrahi (✉) · P. Srivastava
Genetic Metabolic Unit, Department of Pediatrics, APC, PGIMER, Chandigarh, India
e-mail: panigrahi.inusha@pgimer.edu.in

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_28

in countries where prenatal diagnosis and selective termination of affected fetuses are not legally allowed. The individual's deoxyribonucleic acid (DNA) carries the genetic code or genetic uniqueness for that person. For the disorders caused by single-nucleotide variants (SNVs), a DNA profiling by Sanger sequencing is a commonly used option. We cover all the techniques used in DNA profiling, the indications, and limitations of the genetic testing using human DNA in the present write-up.

Keywords

Prenatal testing · Chromosomal anomalies · DNA-based techniques · DNA profiling · Fetus · Molecular cytogenetics · Newborn · Prenatal diagnosis

Introduction

The paradigm of examining an infant with the congenital anomalies or birth defects has shifted to evaluate a baby in the womb with the more advanced fetal sampling techniques with testing using better cytogenetics and molecular techniques. Fetal medicine is an upcoming speciality in which embryology, fetal pathophysiology, and several screening and diagnostic techniques are covered. Prenatal screening and diagnosis is an important aspect of clinical genetics and obstetric care that epitomizes the effective coordination between the clinicians and the laboratory. Congenital anomalies account for 2–4% of all newborns but are responsible for around 20% of the perinatal mortality (Murphy et al. 2017). Chromosomal abnormalities and DNA mutations or variants account for most birth defects. Prenatal genetic diagnostic testing is basically intended for determining the status of a specific disorder in the developing fetus with much more certainty and accuracy. It can reduce the burden of genetic disorders and can help the families to take further decision of discontinuation of the pregnancy if the fetus is affected in a stipulated time frame. Advancement and development of various techniques to detect chromosomal aberrations or monogenic defects have optimized the testing and improved reproductive outcomes (Audibert et al. 2011). With availability of molecular cytogenetics and advanced DNA testing methods, chromosomal aberrations can also be detected by DNA profiling. But due to the complex nature and different pathophysiology (epigenetics) of certain diseases, it is not possible to get the certain answers from prenatal testing in several disorders. Prenatal genetic testing is now routinely offered as an essential part of prenatal care for all the pregnant women at an increased risk of getting baby with genetic disorder. In the view of no treatment of most of the genetic disorders, prenatal testing gives promising option for better decision taking. Prenatal diagnosis is justified in most disorders affecting children and causing significant morbidity and/or mortality.

To assess risk of genetic condition in fetus, two options are available: prenatal genetic screening and specific genetic testing if required. Screening is basically an option to check the chances of having certain genetic disorder in a fetus, while

confirmatory genetic testing is intended to determine specific genetic disorder or condition in fetus much more accurately. For confirmation of the particular abnormality after a prenatal screening test, prenatal diagnostic tests are performed. Sampling usually involves invasive procedures like chorionic villus sampling (CVS) and amniocentesis, done at 11–14 weeks' and 15–17 weeks' gestation. In few cases, fetal cord blood sampling may be required to perform any kind of chromosomal or DNA testing to give a definite diagnosis of most genetic disorders.

Imaging techniques to identify or screen structural chromosomal abnormalities include ultrasonography (USG), fetal echocardiography, and fetal magnetic resonance imaging (MRI). Because of certain consequences of an invasive procedure, scientists have tried to find out new ways to reduce the complications. With time, noninvasive prenatal screening (NIPS) technique came into existence which has dramatically revolutionized diagnosis of aneuploidies. These tests are voluntary, and proper pre- and post-test genetic counselling need to be done to inform about the procedure, possible results or outcome of the test, and available options for further decision-making (Cariati et al. 2019).

In this write-up, more advanced current options and indications for prenatal diagnosis, testing by various DNA-based molecular methods, and their advantages and limitations are discussed below.

Indications for Genetic Testing

Mostly genetic testing requires invasive procedures for getting fetal sample. Fetal cells are required for prenatal genetic testing. There are certain factors/indications that put a couple at higher risk of some genetic disorder and necessitate specific DNA testing: The common indications for genetic testing are listed in Table 1. Other indications for prenatal genetic testing are history of infertility, previous child with

Table 1 Familial or prenatal indications for genetic assessment in unborn baby

Situation in which genetic testing may be done
High risk of single-gene disorder
Advanced maternal age (35 or older)
A positive family history increasing susceptibility
A previous baby with genetic disease
Ultrasound-detected fetal malformations
Positive first trimester serum screening
Chromosomal abnormality/translocation in either parent
Multiple miscarriages or prior stillbirth
Oligohydramnios or polyhydramnios
X-linked disorders like Duchenne muscular dystrophy
Inborn errors of metabolism
Advanced paternal age (40 or older)

multiple malformations where cause was not identified, and prenatal exposure to drugs, alcohol, or other environmental factors which can also be the risk factors for some birth defects, developmental delays, and/or illnesses.

Pre- and Post-test Genetic Counselling

Each and every pregnant woman should undergo counselling before any specialized screening or diagnostic test. Genetic testing can be highly specialized and costly and can sometimes result in uncertainty of results, so proper pre- and postnatal counselling is of utmost importance. The following points should be conveyed to couples who are undergoing these tests: (a) available options and their limitations for prenatal testing which should be discussed; (b) explanation of individual risk and background risk of congenital disease and anomaly; (c) possible outcomes of the test; (d) possibility of the diseases which can be detected and feasibility of their management; (e) the risk involved in particular testing; (f) ethical and time-limited issues; and (g) support and additional options available.

Invasive Prenatal Diagnostic Procedures for Chromosomal or DNA Testing

Prenatal cytogenetics with the combination of DNA molecular genetics has opened up new avenues in prenatal diagnosis. The specialized molecular tests are designed on the basis of type of genetic disorder to be detected, and sampling depends on patient preferences and time of pregnancy. Each of these tests provides different kinds of information. After the invasive sampling procedure, the genetic testing is done on the sample in the laboratory. The list of different sampling methods is detailed in Table 2. Guidelines for invasive testing are available from international experts (Ghi et al. 2016). Advantage of molecular methods to detect the most common chromosomal aneuploidies over traditional karyotyping is that rapid results are available in 1–3 days. The lab tests usually performed include (a) fluorescence in situ hybridization (FISH); (b) PCR-based tests, RFLP, ARMS-PCR, and Sanger sequencing; (c) quantitative fluorescence (QF)-PCR; (d) multiplex ligation-dependent probe amplification (MLPA); and (e) cytogenetic microarray (CMA).

Prenatal Cytogenetics and Its Limitations

Majority of structural chromosomal aberrations (deletions, duplications, translocations) and all chromosomal aneuploidies can be detected prenatally by traditional karyotyping. It involves analysis of full chromosome complement – 46 chromosomes. Sufficient metaphases have to be evaluated and reporting can take around

Table 2 Invasive procedures used in prenatal diagnosis for genetic disorders

Technique	Timing of procedure	Application	Benefits	Limitation and risk
Chorionic villus sampling	11–14 weeks	Chromosome analysis (karyotyping) Molecular genetic diagnosis	Early diagnosis and decision-making	0.5–1.0% risk of miscarriage
Amniocentesis	15–17 weeks	Chromosome analysis Molecular genetic diagnosis Biochemical diagnosis like enzyme testing	Multiple types of testing are possible on the sample	0.5–1.0% risk of miscarriage
Placental biopsy	From 15 weeks	Chromosome analysis Molecular genetic diagnosis	Detects abnormalities when patients present late in pregnancy	0.5–1.0% risk of miscarriage
Cordocentesis	16–20 weeks	Chromosome analysis Hematological tests Biochemical diagnosis	Especially used for direct hematological testing involving HPLC, flow cytometry, factor assay	0.5–1.0% risk of miscarriage
Fetal skin biopsy	20 weeks	Diagnosis of specific genetic dermatoses	Selected skin disorder detection	0.5–1.0% risk of miscarriage

2 weeks. It is technically difficult, and the couples at higher risk of having a genetic disorder may have to wait for up to 2 weeks for getting the prenatal karyotyping report.

The primary aims of developing DNA-based tests were to alleviate parental anxiety and give the report more rapidly while retaining the accuracy. With the advancement in the techniques, these became available for many rare disorders also and contributed to their clinical management. Different tests used for several genetic disorders are short-listed in Table 3.

Technical details are mentioned below:

(a) **Fluorescence In Situ Hybridization (FISH)**

FISH is a powerful combination of molecular and cytogenetics (molecular cytogenetics) where a DNA sequence which is fluorescently labelled based on the sequence complementarity is hybridized onto a slide carrying the metaphase spread.

Table 3 Lab methods used for prenatal diagnosis in different genetic disorders

Test	Sample	TAT	Abnormalities detected	Advantages	Disadvantages
Conventional karyotyping	Amniotic fluid	2–3 weeks	Deletions or duplications of size 5–10 Mb	Translocations, inversions, and ring chromosomes can be seen	Deletion/duplication <5 Mb cannot be detected
FISH – direct preparation (interphase)	Cord blood	24–48 h	Chromosomes 13, 18, 21, X, and Y	Rapid assessment of major aneuploidies and specific microdeletion syndromes	Direct testing on CVS on amniotic fluid is usually not reliable and may not be accurate
Metaphase FISH	Amniotic fluid culture	1–2 weeks	Known microdeletion or microduplication syndromes	Used to detect known specific abnormalities	Specific probes are needed for particular disorder
QF-PCR	DNA from CVS/amniotic fluid	5 days	Common aneuploidies	Can be used as standalone test	All chromosome abnormalities cannot be detected
Chromosomal microarray	DNA from blood/ amniotic fluid/ CVS	1–2 weeks	CNVs of size over 200 kb can be easily identified	Major chromosomal abnormalities can be detected including aneuploidies. The resolution can vary with different microarray platforms. For prenatal, usually low-resolution microarray is used	Balanced rearrangements and some triploidies cannot be detected. Mosaicism may also be missed
Multiplex ligation-dependent probe amplification (MLPA)	DNA from blood/ amniotic fluid/ CVS	2–3 days	Copy number variants in common microdeletion/ duplication, SMA, DMD, etc.	Highly specific, multiplexed, and fast	Whole genome/novel variants cannot be Detected; Sensitivity to contaminants
MS-MLPA	DNA from blood/ amniotic fluid/ CVS	2–3 days	Methylation status	It can detect Angelman and Prader-Willi syndrome	Only for selected disorders

PCR-based mutation detection (RFLP, ARMS)	DNA from blood/ amniotic fluid/ CVS	Around 1 week	Used to test a known mutation for a specific genetic disorder as per previous affected child in family	Used for some commoner disorders and to detect locally prevalent mutations	Previous identification in index case or carrier parents required
Targeted Sanger sequencing	DNA from CVS/amniotic fluid	7 days	Known mutation detection previously identified in proband	Usually commonly found mutations detected in some genetic disorders	Not done for detection of larger exonic deletions or duplications
Preimplantation genetic diagnosis (PGD)	Cell taken from embryo	Couple of days	Mutation known in family or disorder	Selected few mutations can be tested for which couple is at high risk	Error rate can be high but acceptable in experienced centers

This method is useful for the identification of a particular gene, chromosomal abnormalities (numerical and structural), possible origin of marker chromosome, etc. Complete work flow (method) is depicted in Fig. 1.

For the detection of known aneuploidies of chromosomes 21, 13, 18, X, and Y, pericentromeric satellite probes or alpha satellites are used for hybridization and subsequent detection.

Whole chromosome painting or WCP refers to use of several probes for the chromosomes targeting the repetitive sequences and then labelling with specific fluorochrome. Apart from this, these probes can also be combined with locus-specific probes to identify a particular missing part of a chromosome in case of contiguous gene deletion syndromes. The microdeletion and microduplication syndromes which can be detected include DiGeorge syndrome, Prader-Willi syndrome, Angelman syndrome, Miller-Dieker syndrome, Williams syndrome, and Potocki-Shaffer syndrome. DiGeorge syndrome is common cause of cardiac defect, and Prader-Willi syndrome is common cause of genetic obesity.

(b) **Screening for Common Mutations**

Mutation screening techniques are the methods used for screening or scanning a gene or gene fragment for the presence of mutations or variants. They are used in identifying mutations in genes which are very large and genetically heterogenous. They help in sparing the effort and expense of meticulous and complete sequencing

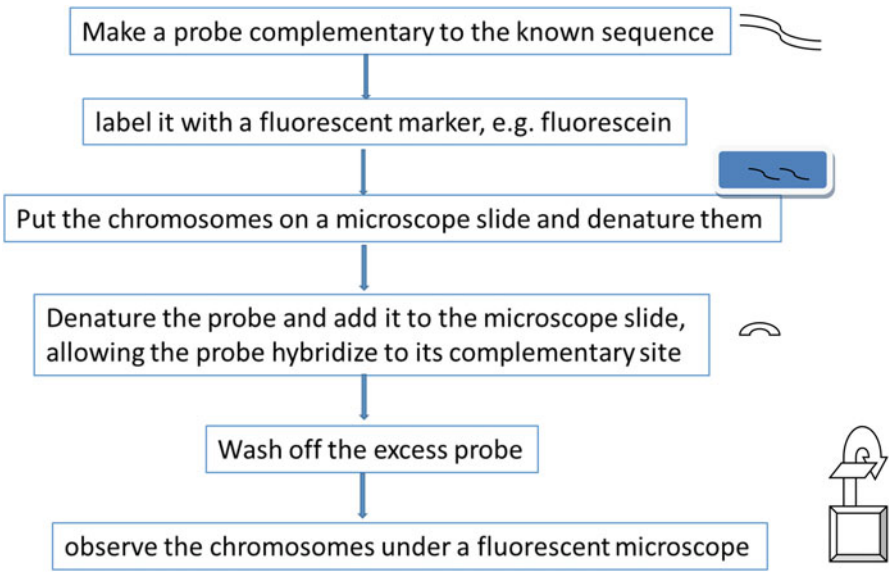


Fig. 1 Steps involved in FISH testing for specific disorders

of the genes for identifying the mutations. Mutation screening by PCR-based method requires confirmation by DNA sequencing. These techniques are used to detect point mutations, small insertions, and deletions. If there is deletion of one or more exons, then it can be detected by multiplex PCR. Various methods of mutation screening are discussed hereunder:

Polymerase Chain Reaction (PCR)

This technique is used to make multiple copies (microgram) of a specific DNA segment which is present in very small amount. Primers against known sequences are developed and the specific target PCR can be designed. PCR is the basic technique for all further subsequent advanced testing. It consists of a series of cycles of three major steps conducted at different temperatures. The advantages and applications are listed in Table 4. The test is usually done in most genetic disorders on genomic DNA extracted from CVS or amniotic fluid for prenatal diagnosis. The successive steps in a PCR reaction are listed below. These steps are redone again and again around 30 times in a thermal cycler to get the desired PCR product.

1. **Denaturation**: The starting DNA material is heated up to 93–95 °C to separate the two complementary strands of DNA. This much high temperature is enough to break the hydrogen bonds between two strands. As a result of denaturation, both the strands get separated and form single-stranded DNA.
2. **Annealing**: After denaturation, the temperature is cooled down as per melting temperature of the synthetic oligonucleotide primers, so that primers are allowed to sit on the complementary single-stranded DNA sequence and bind by base pairing.
3. **Extension**: At a temperature of 70–75 °C, DNA polymerase (Taq polymerase) synthesizes the sequence from dNTPs which include dATP, dCTP, dGTP, and dTTP. The new strands are complementary to the DNA template or target DNA sequence of specific gene.

Table 4 The advantages, disadvantages, and utility of PCR testing

Advantages of PCR	Disadvantages of PCR	Applications of PCR
• PCR is very simple and rapid process • It is extremely sensitive and robust • It allows the amplification of minute amounts of target DNA	• Size range of the amplification products in a standard PCR reaction is rarely more than 1 kb • It is sensitive to impurities and temperature	• PCR is usually used for identification and analysis of mutations in specific genetic disease • DNA deletions and insertions can be easily detected by a change in the size of the PCR products

PCR, Polymerase chain reaction

Amplification-Refractory Mutation System (ARMS)-PCR

This test uses modification of the simple PCR method. Separate reverse primers are used for normal and mutant alleles of a target sequence in the tested gene. This method is usually used to detect common thalassemia mutations in parents of an affected child.

Two separate reaction tubes are used: one for the wild-type allele and another for mutant allele. The first tube uses the forward primer and reverse primer for wild or normal allele. In the second tube, the forward primer and reverse primer for mutant allele detection are used. The DNA polymerase is unable to repair single-base mismatch between template and primer at 3′ end of the primers because polymerase lacks a 3′ exonuclease activity. PCR product is then electrophoresed. If the band is seen only in the reaction with the wild-type primer, the person is homozygous for the mutant primer only. In heterozygous individuals, amplification will be seen in both sets of primers. In autosomal recessive disorder, the patient will be homozygous for one mutation or compound heterozygous for two mutations.

Applications of ARMS-PCR

1. Detecting point mutations, e.g., in beta-thalassemia. The common Indian mutations detected include IVS 1-5(G→C), Co41-42(-TCTT), IVS1-1(G→T), Co8-9 (+G), and -619 bp deletion. The other less common mutations include Co16(-C), Co30(G→C), Cap+1(A→C), Co15(G→A), and -88(C→T).
2. Analysis of polymorphisms related to genetic susceptibility.
3. Detection of hepatitis B virus (HBV) variants to assess response to a therapy.

Advantages

- Ease of developing, validation, and implementation.
- It is simple and does not require the use of sophisticated and expensive detection systems.
- Nonradioactive.

Limitations

- Only the known variants or mutations can be detected.
- It should be combined with other molecular diagnostics strategies like sequencing for providing comprehensive mutation detection.

Restriction Fragment Length Polymorphism (RFLP)

Restriction enzymes are endonucleases that identify and cleave a specific sequence on the double strand on the DNA sequence. These restriction sites are present throughout the genome. RFLP refers to the variability in the restriction pattern in different individuals or in relation to disease. Thus, it is used to identify mutations, deletions, or polymorphisms. Resultant fragments are called restriction fragment and

size is determined by comparing the size with a DNA ladder. Most of thalassemia mutations can be detected by RFLP method. This method is used to identify specific DNA mutation or variability causing genetic disease like hemophilia and cystic fibrosis.

(c) **Quantitative Fluorescence Polymerase Chain Reaction (QF-PCR)**

This technology is a very robust, rapid, cost-effective, and more accurate prenatal genetic test for major aneuploidies. Different from other techniques, this has benefit of detecting mosaicism, triploidy, and maternal cell contamination (MCC). Copy numbers can also be detected by amplification of repeat sequences at polymorphic loci. Gel electrophoresis is used to separate these amplified repeat sequences and labelled products. Peak height ratio is indicative of abnormalities. Presence of two peaks of equal height within the same chromosomal region is a diagnostic of two copies of the allele. Three peeks or two peaks with a ratio of 2:1 or 1:2 are indication of trisomy. QF-PCR is much more robust, rapid, less tedious and more accurate, and suitable for large number of samples as compared to FISH. This test is economical and less expensive diagnostic test for major trisomies (13, 18.21, and X, Y chromosomes) and triploidy.

The advantage of this rapid testing is very obvious for a woman who is at increased risk of common aneuploidy in the pregnancy. This can end up the dilemma of the couple and alleviate their anxiety of having a normal outcome. Abnormal result can be helpful for planning of subsequent management and early decisions.

Applications of QF-PCR

- Major chromosomal abnormalities like trisomy of 13, 18, 21, and sex chromosomes and triploidy can be detected.

QF-PCR for Maternal Cell Contamination (MCC): It is the presence of maternal cells in prenatal test sample. It is more common in CVS cultures because of intimate association between maternal tissue and chorionic villi. To minimize this, maternal deciduas must be carefully dissected. MCC can be tested with the help of QF-PCR.

Indications

- To provide a correct result interpretation, one should be confident that the tissue is purely fetal in origin.
- MCC analysis can be done during zygosity testing in multiple pregnancy.
- MCC analysis also provides an internal control that the biological mother is matched with concordant fetus, thus avoiding the risk of sample mix-up in setting of prenatal testing.

(d) **Triplet-Primed PCR (TP-PCR)**

This is a specialized PCR technique developed to test for triplet repeat disorders. One of the disorders causing mental retardation or intellectual disability in males is

fragile X syndrome (Rajan-Babu and Chong 2019). Increased number of CGG repeats >200 causes the disease. In carrier females, the CGG repeats can be 50–200. Earlier X-chromosome breaks were seen to identify the disorder, but it is now considered a crude and insensitive method. The TP-PCR enables testing and hence diagnosis in fragile X syndrome-suspected patients and also carrier detection with increased sensitivity and reliability.

Sanger Sequencing

Prenatal Sanger sequencing can be offered to the family where specific mutation for a particular disease is already known in the family (Goodeve 2015; Tanacan et al. 2019). It involves following consumables in the reaction: single-stranded DNA template, DNA polymerase, oligonucleotide primers, dNTPs, and dideoxynucleotides (ddNTPs).

The main difference between ddNTPs and dNTPs is that ddNTP lacks a hydroxyl group at the 3′-carbon position which stops the phosphodiester bond formation, resulting in chain termination. In the reaction, DNA fragments of different lengths or sizes are generated ending at the different ddNTPs. These are characterized by separating on capillary electrophoresis method using specific analysis software which shows the sequence of the target DNA in the sample. Further analysis can be done by using various online tools (NCBI BLAST) to find out the mutation or variant in the sequence.

(e) **Multiplex Ligation-Dependent Probe Amplification (MLPA)**

A single MLPA reaction can detect copy number variants (CNVs), maximum of 50 sequences. It is a modified multiplex PCR method which is fast, technically less demanding, cost-effective, yet sensitive requiring DNA (Gouas et al. 2015). There is a probe set of two for each target region with a sequence (attached to one of the probes) and common sequences as primers for PCR reactions. PCR products are separated by capillary electrophoresis, and peak height or area for each probe reflects the relative abundance of the target sequence in the DNA compared with that of normal control. Using specific probes, rapid prenatal diagnosis of aneuploidy, sub-telomeric CNVs, and common microdeletion syndromes can be done (Fig. 2). Probe mixes are commercially available for chromosomes 21, 18, 13, X, and Y. A peak ratio of over 1.3 would indicate duplication or gain in the sequence in the DNA sample tested, whereas a value below 0.7 would indicate deletion or specific sequence loss in the sample DNA.

Applications of MLPA

- For quantitative detection of genomic deletions (common and sub-telomeric deletions), duplications, and point mutations, e.g., BRCA1, BRCA2, MSH2, MLH1, DMD, and SMA. Separate kits are available for DiGeorge syndrome, Williams syndrome, and Prader-Willi syndrome (PWS).

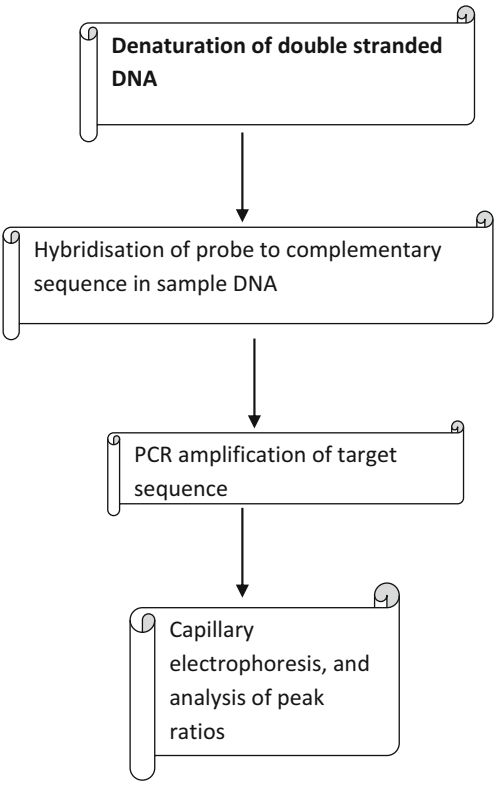


Fig. 2 Steps in a MLPA reaction for deletion or duplication analysis

- Detecting aneuploidy in prenatal samples.
- Centromeric probe sets are used to detect marker chromosome.

Methylation-Specific MLPA (MS-MLPA)

In this method, the probe has a restriction site, and methylated and unmethylated sequences can be identified. In case of methylated DNA, restriction enzyme HhaI digestion will not occur, and PCR amplification will occur and will be detected by the peaks generated.

Uses of MS-MLPA

1. Identification of epigenetic alterations.
2. Methylation-sensitive MLPA is used for the detection of imprinting disorders like Silver-Russell syndrome, PWS, Angelman syndrome, and Beckwith-Wiedemann syndrome (BWS).

Limitations

- MLPA cannot detect balanced translocations.
- Unable to detect unknown point mutations.
- Can't detect any case of a triploidy with a 69,XXX karyotype due to normalization algorithms.
- It can't detect low-level mosaicism (<20%).
- It may miss maternal cell contamination. So it can be combined with QF-PCR for prenatal diagnosis.

(f) **Chromosomal Microarray (CMA)**

CMA testing involves an advanced type of molecular karyotyping which identifies aneuploidies as well as submicroscopic deletions and duplication difficult to be identified by traditional karyotyping (Panigrahi et al. 2018). However, balanced translocations and triploidy/polyploidy cannot be identified. It has been found that abnormal CMA variants can be identified in around 1.7% prenatal samples with normal ultrasound and karyotype result (Wapner et al. 2012). Now CMA has been made available and recommended to be offered in every case undergoing invasive diagnostic procedure. Because of higher sensitivity and higher resolution, CMA can be done if multiple abnormalities are found on prenatal ultrasound (Committee on Practice Bulletins 2016).

Procedure

The test is performed on DNA sample from uncultured cells and also on cultured cells from amniotic fluid. In this technique, small fluorescently labelled single-stranded oligonucleotides are attached to a glass microscope slide known as probes. It is used to detect hybridization of target DNA. It allows simultaneous analysis of several million targets. The data is analyzed using a scanner. SNP array can also detect copy-neutral genetic anomalies such as uniparental disomy (UPD).

Advantages of Microarray

- Can detect known or recurrent as well as novel microdeletion or microduplication syndromes.
- Automated steps.
- Rapid as compared to karyotyping.
- Low level of mosaicism can be detected.

Limitations

- Balanced chromosomal rearrangements cannot be identified by genomic arrays.

(g) **Next-Generation Sequencing (NGS)**

The advancement of technology has given the opportunity to explore more set of genes in one go through next-generation sequencing (NGS) options. NGS has brought a revolution in prenatal diagnosis. Not only targeted sequencing selected

countable genes with known phenotype but also prenatal NGS testing is now an available option (Vora and Hui 2018). Because of complexities of data interpretation and lots of variants of unknown significance, it has not been used in clinical settings but adapted for research purposes (Vora et al. 2017; Mone et al. 2018). Likelihood of detection is higher if multiple abnormalities are found in prenatal ultrasonography, and in presence of cardiac defects, or skeletal anomalies (Lord et al. 2019). Yield is lowest if isolated increased nuchal translucency is found.

Noninvasive Prenatal Testing (NIPT) for Prenatal Cytogenetic Abnormalities

Cell-free fetal DNA (cffDNA) is a new hope for women who don't want to undergo invasive procedures or any procedure associated with high risk. NGS has been successfully integrated into NIPT and used for detection of chromosomal aneuploidies using shotgun approach or targeted panel testing for aneuploidies (chromosomes 13, 18, X, and Y chromosomes). Only a maternal blood sample is needed for the test in a special tube (D'ambrosio et al. 2019).

There are three main NGS-based approaches for NIPT:

1. Aneuploidy testing on cffDNA
2. Targeted region sequencing
3. Single-nucleotide polymorphism (SNP) analysis

For aneuploidies, the fragments generated from different chromosomes are analyzed, and this testing is most sensitive for detection of trisomy 21 or Down syndrome (DS). However this is considered a screening test and can be reliably done from 8 weeks' gestation onward. It requires a confirmatory testing if positive for aneuploidy.

Second strategy is to amplify specific genomic regions and then sequence them. In comparison this technique is cheaper, as it is targeted sequencing, but has the limitation of missing the whole genome which can be important in some cases.

The third strategy refers to amplification and analysis of selected SNP regions in special situations. Thus, it has good specificity and acceptable sensitivity in most situations especially in case of prenatal DS detection.

This is a relatively expensive test. There are resource limitations in developing countries. Also, problems also occur in finances if the prenatal tests are not done with governmental support. Unexpected results can create undue anxiety.

Illustrative Cases of Prenatal Diagnosis by Various Methods

Case 1: A 26-year-old female, resident of Rupnagar, Punjab, presented at 12 weeks' gestation in pregnancy with history of previous child affected with *spinal muscular atrophy*. The DNA testing done in the child had shown homozygous deletion in the exon 7 of the *SMN1* gene.

The couple was counselled regarding 25% risk of recurrence in every pregnancy and feasibility of prenatal testing. A CVS sample was obtained under ultrasonographic guidance, and mutation testing was done by MLPA on the DNA extracted from the CVS sample. The fetus was found heterozygous for the exon 7 deletion and hence was unlikely to be affected with SMA. Additionally a maternal EDTA sample was also collected for testing of maternal cell contamination (MCC) testing, and this test showed no maternal contamination on the CVS DNA. Post-test counselling was done in the couple and they continued the pregnancy. SMA is a common neuromuscular disorder, in which in severe SMA type I, babies present with neonatal respiratory problems and floppiness/hypotonia. Nusinersen and gene therapy are treatment options in only selected patients (Bharucha-Goebel and Kaufmann 2017). Prenatal diagnosis is an option for preventing recurrence in family. Carrier frequency can be tested by MLPA or droplet digital PCR or quantitative PCR (Dejsuphong et al. 2019).

Case 2: A young couple from Sangrur, Punjab, presented in pregnancy with previous child affected and expired with GM1 gangliosidosis at 1.5-year age. The enzyme testing done in the child had shown low activity of beta-galactosidase consistent with ***GM1 gangliosidosis***. Since mutation testing was not done in the child; the couple was counselled, and carrier testing was done in both husband and wife by next-generation sequencing (NGS). NGS is a test in which massive parallel sequencing is done for likely mutations in targeted gene or genes which can explain the phenotype or genetic disorder in the family. The wife showed a variant in the *GLB1* gene: c.65-75del:p.Arg22fs; and the husband showed the variant *GLB1*: c.276G>A:p.Trp92*. A prenatal amniotic fluid sample was taken at 17 weeks' gestation and tested for the variants identified in the parents by Sanger sequencing. The fetus was found unlikely to be affected. Genetic counselling was performed. GM1 gangliosidosis is a lysosomal storage disorder in which no definitive treatment is available. The infants present with hepatosplenomegaly, coarse facies, macular cherry-red spots, and dysostosis, and diagnosis is by enzyme and/or DNA testing (Verma et al. 2015). Late-onset cases can present with neurologic symptoms like dystonia and dysphagia (Arash-Kaps et al. 2019).

Case 3: A 28-year-old female, resident of Chandigarh, came with history of previous child 6 years old with Duchenne muscular dystrophy (DMD). The MLPA done in the child had shown deletions of exons 48–50 in the dystrophin gene. The female was carrying pregnancy of 9 weeks. Apart from routine prenatal workup, genetic counselling was done for the couple. A CVS sample was collected at 12 weeks' gestation, and MLPA testing was done along with MCC testing for prenatal diagnosis. Post-test counselling was done. DMD affects males who present with frequent falls, calf hypertrophy, increase in creatine kinase, and positive Gower sign. Becker muscular dystrophy (DMD) is milder form of the disorder. Different trials for therapy are in progress but most patients cannot get any definitive treatment (Tyagi et al. 2019). Carrier detection can be done also in females by MLPA (Deepha et al. 2017).

Case 4: A 29-year-old female came with a child affected with suspected muscular dystrophy. The child had increased creatine kinase levels and positive Gower sign.

MLPA testing in the child for DMD was negative for deletions or duplications in the dystrophin gene. A NGS testing was done in the child to look for likely point mutations causing the phenotype in the child or other muscular dystrophy gene mutations. The child was found homozygous for a *SGCA* gene variant in the exon 3 of the gene, a variant of unknown significance (VUS). The variant was ENST00000262018_ *SGCA*: c.293C>A:p.Arg98His. The parents were tested by Sanger sequencing and found to be carriers for the variant. A prenatal testing was done on amniotic fluid sample, and the DNA was tested for the same variant by Sanger sequencing. The fetus was found heterozygous for the variant, and possibly unlikely to be affected (Fig. 3). Homozygous or compound heterozygous pathogenic mutations in the *SGCA* gene on chromosome 17q21.33 lead to autosomal recessive limb-girdle muscular dystrophy (LGMD) type 3 or LGMD2D (Passos Bueno et al. 1999). Onset occurs in first decade and the children may lose independent ambulation in teenage years. Scapular winging, contractures, and scoliosis can be seen with increased age. Cardiomyopathy may also be seen. However severity can be variable in different families. *CAPN3* gene-related LGMD2A patient can also present with calf hypertrophy and also have severe phenotype.

Case 5: A 39-year-old female presented with previous 9-year-old child having global developmental delay and dyskinesias. The child started having symptoms at

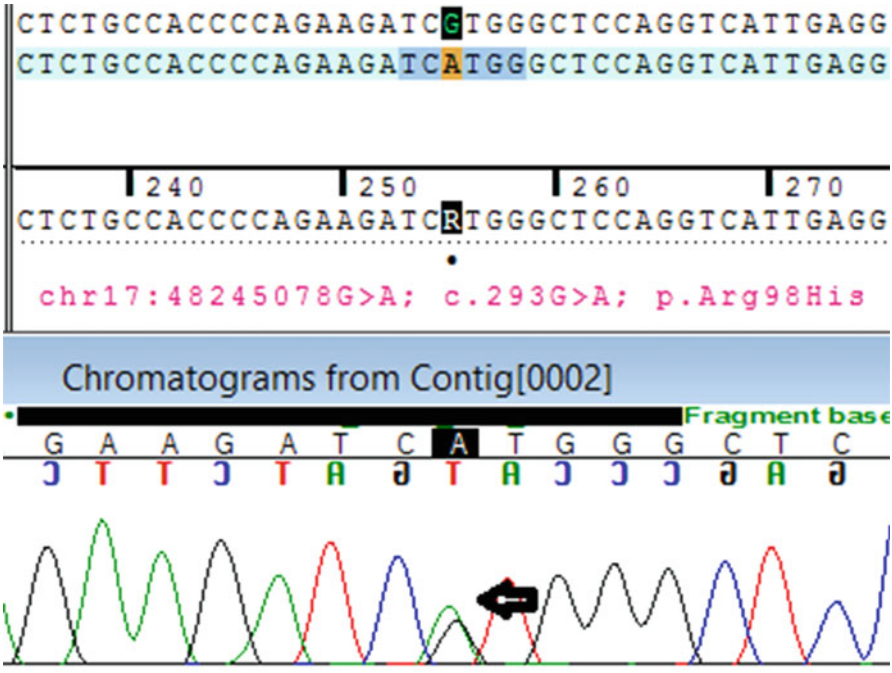


Fig. 3 Sanger sequencing output showing heterozygosity for the SGCA variant in the tested fetus suggesting fetus is unlikely to be affected

7 years of age with decreasing scholastic performance and behavioral problems. She was investigated for possible autoimmune encephalitis and metabolic workup was done. The urine gas chromatography-mass spectrometry (GC-MS) showed excretion of methylmalonic acid and increased homocysteine levels. The child's brain magnetic resonance imaging (MRI) showed basal ganglia hyperintensities with cortical atrophy. A targeted NGS was done and she was found to have a homozygous variant in exon 3 of the *MMACHC* gene. The nonsense variant detected was ENST00000401061_*MMACHC:*c.394 C>T:p. Arg132*. Both parents were heterozygous for the variant on Sanger testing. A prenatal testing was done on CVS sample with MCC testing. The fetus was found to be heterozygous for the same variant and hence unlikely to be affected. Since there was advanced maternal age, an aneuploidy testing was also performed on the amniotic fluid by karyotyping. No autosomal abnormalities were identified on karyotyping in the fetus. The sex chromosomal status cannot be revealed on prenatal sample as per Indian Pre-Conception and Pre-Natal Diagnostic Techniques (PC-PNDT) Act, 1994 and 2003. Biallelic mutations in *MMACHC* gene lead to inherited cobalamin defect (Carrillo-Carrasco et al. 2012). Infantile presentation includes hypotonia, failure to thrive, hemolytic uremic syndrome, and hydrocephalus. In later-onset forms, neuroregression, behavioral problems, encephalopathy, and thromboembolic complications are noted. Increased plasma homocysteine and increased excretion of methylmalonic acid (MMA) in urine give clues to the diagnosis. Management includes appropriate nutrition and hydrocobalamin (OHCbl) supplementation.

Case 6: A 26-year-old female resident of Jind, Haryana State, came with previous history of two infantile losses, and the third baby was evaluated for possible Leigh syndrome. The girl was symptomatic from infancy with predominant motor delay and hypotonia. She also had small head, nystagmus, and dystonia noted at 2 years of age. Additional features in the child included poor weight gain and constipation. Her hemoglobin was 10 g/dl and serum lactate level was 110 mmol/l. Her EEG was normal. The previous siblings were similarly affected. Mutation testing was done by targeted NGS keeping a possibility of Leigh syndrome (LS). This revealed two variants in *PNPT1* gene. The variants were *PNPT1*: c.G1583A:p.Arg528His and *PNPT1*:c.1822+1G>A; the latter was a splicing variant. The couple was counselled accordingly and options for prenatal testing by DNA analysis were discussed. Subsequently DNA analysis was performed on CVS sample for prenatal diagnosis. In Leigh syndrome, mutations in around 75 genes are known to lead to the phenotype in humans (Lee et al. 2020). These can be genes in nuclear genome or the mitochondrial genome. Some cases of LS may be biotin and/or thiamine responsive and thus this treatment can be tried in selected cases. Basal ganglia involvement leads to dystonias in LS. It is important to confirm the mutation findings, and correlate with phenotype, and do parental testing for the two variants in case of autosomal recessive disorder, before doing a prenatal diagnosis.

Conclusions

Several genetic testing methods are available for prenatal diagnosis and many of these are DNA-based tests. Newer tests more frequently being used nowadays are MLPA, QF-PCR, and chromosomal microarray. The choice of the test would depend on the disorder suspected or disorder diagnosed in the previous affected child. For most single-gene disorders, a CVS is done followed by PCR and Sanger sequencing. MLPA is done mainly in previous DMD and SMA cases in the target family. The testing should be done with appropriate pre-test and post-test genetic counselling.

References

Arash-Kaps L, Komlosi K, Seegräber M, Diederich S, Paschke E, Amraoui Y et al (2019) The clinical and molecular spectrum of GM1 gangliosidosis. J Pediatr 215:152–157

Audibert F, Gagnon A, Genetics Committee of Society of Obstetricians and Gynecologists of Canada, Prenatal Diagnosis Committee of the Canadian College of Medical Geneticists (2011) Prenatal screening for and diagnosis of aneuploidy in twin pregnancies. J Obstet Gynaecol Can 33(7):754–767

Bharucha-Goebel D, Kaufmann P (2017) Treatment advances in spinal muscular atrophy. Curr Neurol Neurosci Rep 17(11):91. https://doi.org/10.1007/s11910-017-0798-y

Cariati F, D'Argenio V, Tomaiuolo R (2019) The evolving role of genetic tests in reproductive medicine. J Transl Med 17:267

Carrillo-Carrasco N, Chandler RJ, Venditti CP (2012) Combined methylmalonic acidemia and homocystinuria, cblC type I. Clinical presentations, diagnosis and management. J Inherit Metab Dis 35(1):91–102

Committee on Practice Bulletins – Obstetrics, Committee on Genetics, and the Society for Maternal-Fetal Medicine (2016) Practice bulletin no. 163: screening for fetal aneuploidy. Obstet Gynecol 127:e123–e137

D'ambrosio V, Squarcella A, Vena F, Di Mascio D, Corno S, Pajno C, Piccioni MG et al (2019) Update on non-invasive prenatal testing. Minerva Ginecol 71(1):44–53

Deepha S, Vengalil S, Preethish-Kumar V, Polavarapu K, Nalini A, Gayathri N, Purushotta M (2017) MLPA identification of dystrophin mutations and in silico evaluation of predicted protein in dystrophinopathy cases from India. BMC Med Genet 18(1):67. https://doi.org/10.1186/s12881-017-0431-6

Dejsuphong D, Taweewongsouton A, Khemthong P, Chitphuk S, Stitchantrakul W, Sritara P et al (2019) Carrier frequency of spinal muscular atrophy in Thailand. Neurol Sci 40(8):1729–1732

Ghi T, Sotiriadis A, Calda P, Da Silva CF, Raine-Fenning M, Alfirevic Z et al (2016) ISUOG practice guidelines: invasive procedures for prenatal diagnosis. Ultrasound Obstet Gynecol 48 (2):256–268

Goodeve AC (2015) Hemophilia B: molecular pathogenesis and mutation analysis. J Thromb Haemost 13(7):1184–1195

Gouas L, Kémény S, Beaufrère A, Eymard-Pierre E, Pebrel-Richard C, Tchirkov A et al (2015) Prenatal screening of 21 microdeletion/microduplication syndromes and subtelomeric imbalances by MLPA in fetuses with increased nuchal translucency and normal karyotype. Cytogenet Genome Res 146(1):28–32

Lee JS, Yoo T, Lee M, Lee Y, Jeon E, Kim SY et al (2020) Genetic heterogeneity in Leigh syndrome: highlighting treatable and novel genetic causes. Clin Genet 97(4):586–594

Lord J, McMullan DJ, Eberhardt RY, Rinck G, Hamilton SJ, Quinlan Jones E et al (2019) Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): a cohort study. Lancet 393(10173):747–757

Mone F, Quinlan Jones E, Kilby MD (2018) Clinical utility of exome sequencing in the prenatal diagnosis of congenital anomalies: a review. Eur J Obstet Gynecol Reprod Biol 231:19–24

Murphy SL, Mathews TJ, Martin JA et al (2017) Annual summary of vital statistics: 2013–2014. Pediatrics. https://doi.org/10.1542/peds.2016-3239

Panigrahi I, Jain P, Didel S, Agarwal S, Muthuswamy S, Saha A, Kulkarni V (2018) Identification of microdeletion and microduplication syndromes by chromosomal microarray in patients with intellectual disability with dysmorphism. Neurol India 66(5):1370–1376

Passos Bueno MR, Vainzof M, Moriera ES, Zatz M (1999) Seven autosomal recessive limb girdle muscular dystrophies in the Brazilian population: from LGMD2A to LGMD2G. Am J Med Genet 82:392–398

Rajan-Babu I-S, Chong SS (2019) Triplet-Primed PCR and apillary electrophoresis for characterizing the Fragile X mental retardation 1 CGG repat hyperexpansions. Methods Mol Biol 1972: 199–210

Tanacan A, Gurbuz BB, Aydin E, Erden M, Coskun T, Beksac MS (2019) Prenatal diagnosis of organic acidemias at a tertiary center. Balkan J Med Genet 22(1):29–34

Tyagi R, Kumar S, Dalal A, Mohammed F, Mohanty M, Kaur P, Anand A (2019) Repurposing pathogenic varants of *DMD* gene and its isoforms for DMD exon skipping intervention. Curr Genomics 20(7):519–530

Verma J, Thomas DC, Sharma S, Jhingan G, Saxena R, Kohli S et al (2015) Inherited metabolic disorders: prenatal diagnosis of lysosomal storage disorders. Prenatal Diagn 35(11):1137–1147

Vora NL, Hui L (2018) Next-generation sequencing and prenatal omics: advanced diagnostics and new insights into human development. Genet Med 20:791–799

Vora NL, Powell B, Brandt A et al (2017) Prenatal exome sequencing in anomalous fetuses: new opportunities and challenges. Genet Med 19:1207–1216

Wapner RJ, Martin CL, Levy B, Ballif BC, Eng CM, Zachary JM et al (2012) Chromosomal microarray versus karyotyping for prenatal diagnosis. N Engl J Med 367:2175–2184. (Level II-3)

CBU Posttransplant Chimerism Analysis Using ChimerMarker™

30

Donato Madalese, Roberta Penta de Vera d'Aragona, Federica Schiano di Tunnariello, and Giovanna Maisto

Contents

D. Madalese (✉) · R. Penta de Vera d'Aragona
Histocompatibility and Immunogenetics Laboratory, UOSD Cryopreservation and Ba.S.C.O, Oncohaematology Department, Pausilipon Hospital, A.O.R.N. Santobono-Pausilipon, Naples, Italy
e-mail: r.penta@santobonopausilipon.it

F. Schiano di Tunnariello
Manipulation and Cryopreservation Laboratory, UOSD Cryopreservation and Ba.S.C.O, Oncohaematology Department, Pausilipon Hospital, A.O.R.N. Santobono-Pausilipon, Naples, Italy

G. Maisto
Flow Cytometry Laboratory, UOSD Cryopreservation and Ba.S.C.O, Oncohaematology Department, Pausilipon Hospital, A.O.R.N. Santobono-Pausilipon, Naples, Italy

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_31

Abstract

In the last decade, short tandem repeat (STR) analysis of the posttransplant umbilical cord blood unit (CBU) chimerism has become a remarkable and versatile tool in pediatric treatment. Its use enables to effectively monitor the CBU engraftment every 30 days. The employment of both a 24-loci kit multiplex polymerase chain reaction (PCR) and software developed for this purpose has made possible the transplant progress evaluation and the risk of a graft versus host disease (GVHD). CBU was chosen because, compared to matched unrelated donor (MUD), it is less alloreactive even with one mismatch in human leukocyte antigen (HLA). In fact, several studies have demonstrated that the presence of a third cell population of maternal origin generates an added value on this type of tissue. Herein we present a general overview of the CBU posttransplant chimerism analysis, exploring clarifying cases recently reported and, finally, a befitting and practical ChimerMarker™ software description.

Keywords

Chimerism · Cord blood unit posttransplant · Multiplex PCR · ChimerMarker™ · STR analysis

Introduction

Nowadays in the transplant field, the possibility to monitor over time the progression in the bone marrow (BM) transplant is of crucial importance. In particular, when it involves the use of umbilical cord blood (UCB) for pediatric subjects, this kind of patient is often very young (sometimes they are infants of a few months) and affected by previous pathologies such as leukemia, thalassemias, and anemias.

Why exactly use cord blood? Because the UCB is composed of a high number of stem cells, useful to repopulate marrow tissue, depleted of its initial population; In Haematopoietic Stem Cells transplantation UCB are always avabile, while sibiling and unrelated donor may not avabile. Indeed, for these cases, the unit also has greater variability of HLA (human leukocyte antigen) and, in case of mismatch, gives low immunogenicity.

In this context, different methods have been developed to monitor the post-transplant. Based on our experience, of significant impact is the use of the semi-quantitative method of the STR analysis, exploiting the capacity of a peculiar software: ChimerMarker™.

Why did we decide to talk about the chimerism in this context? DNA profiling is usually associated with forensics applications, involved in cases of “death.” However, in contrast, it is interesting to note how these same methods can be also helpful in a “life” field. This is because they are able to accurately identify a specific allelic profile, allowing a better monitoring of the patient during the time and giving the possibility to intervene in advance in case GVHD appears.

Definitions

Complete chimerism (CC) indicates a total replacement of the patient's cells of BM, with cells of the chosen donor, after the transplant of staminal cells – hematopoietic stem cell transplantation (HSCT). Instead, mixed chimerism (MC) indicates a simultaneous presence of donor's and host's cellular population (Andreani et al. 2014).

MC has three different outcomes:

- Complete chimerism (CC), in which there is a total replacement of the host's hematopoietic compartment.
- Persistent mixed chimerism (PCM), in which populations continue to coexist for the rest of the subject's life.
- A third is a condition that can completely end. In this case, let it be a resumption of pathology.

Instead, with the term microchimerism, it is defined as the presence of a minimum number of cells in a posttransplant patient that have a different origin both from the individual and the donor.

Chimerism and Transplantation

The word chimerism recalls to *chimera*, creature of the mythological Greek. The ancient authors described this animal as a monster with a head and body of a lion, a second head of goat, and a poisonous snake as a tail.

It's interesting to see how the concept of chimera has been applied in the biology field. It is defined as *chimera*, a subject who has two completely different genetic heritages, who came from two zygotes not related to each other. Instead, in case two cellular populations are from the same zygote, mosaicism, the subject is defined as "mosaic."

Chimeras are formed by four parental cells (two fertilized eggs or early embryos combined together) or by three parental cells (one fertilized egg is combined with a non-fertilized egg or the fertilized egg is melted with further sperm). Each cellular population preserves its characteristics and the resultant animal/subject is a mixture of mismatched regions.

This phenomenon, that may seem rare, is in the contrary a normal condition when a patient is submitted on an organ transplant, in our case, on BM transplantation. In fact, when the patient is subject to this procedure, his immune system is quite completely depleted, through a cycle of chemoradiotherapy pretransplant (said "conditioning"). The substitution of the staminal cell compartment often involves not only for the hematopoietic tissue but also other districts: macrophages of pulmonary alveoli, cells of Kupffer in the liver, osteoclasts, the cells of Langerhans of the skin, the microglia cells of the brain, and also the striated muscle cells.

The evaluation the state of engraftment, how much is eradicated the new group on staminal inside medullary niches, is carried out a posttransplant evaluation by cytogenetics, assessing karyotype banding, through blood group, with the study of the HLA system, or also through methods of molecular biology as analysis of STR or Variable Number Tandem Repeat (VNTR) and relative study of DNA's.

Three chimeric statuses will ensure:

- Complete chimerism (lack of patient's hematopoietic cell residue)
- Mixed chimerism (coexistence of donor's and receiving cells)
- Absence of chimerism (autologous hematopoietic reconstitution)

During monitoring of the hemopoietic taking root posttransplant, the study of chimerism provides important information on kinetics of engraftment and represents a unique method of molecular monitoring for those patients without disease-specific gene alterations identified at diagnosis. The study of posttransplant chimerism may be performed through two different methods: the study of STRs by a multiplex PCR and the study of biallelic polymorphic insertion/deletion markers (Ins/Del) by real-time quantitative PCR (Q-PCR). For these methods, there are different commercial kits, many of these, as STR, have been standardized for application in the forensics medicine, where the amount of DNA available often is really minimal (1–2 ng di DNA). In case of the application of these kits on clinical studies, as chimerism study, the quantity of material involved is considerable. A large number of studies demonstrated how under 10 ng of DNA correspond approximately to 1500 starting cells. At the moment, the science community considers acceptable a value of >25 ng of DNA by reaction. These ratings are quite generic and can change in consideration of the methodology used (STR or Q-PCR). However, to ensure a complete reproducibility of the experiment in terms of specificity and accuracy, it is necessary to establish in each laboratory the optimal range of quality of the DNA (260/280 nm $>$ 1,8) and of quantity (10–250 ng) and arrange an experiment of dilution of DNA to scalar concentrations, thus determining the more adequate amount to use depending on the method used.

In the case of young patients, even a few months old, the methods of forensics molecular biology allow the use of a very little quantity of material for the necessary analysis. In fact, it isn't always possible to withdraw more of 1 ml of blood, by patients which are just transplanted, and often, if they had a course of drugs that brought them to aplasia (condition which white blood cells of the patient are broken down through chemotherapy), nucleated cells to extract DNA became very low.

HLA System

The major histocompatibility complex (MHC) is a chromosome gene pool which has the relevant rule to mediate the gene complex and proteins of the humoral and cell-mediated reaction. Different HLA setting could cause an immunologic reaction.

The HLA system mismatching between donor and recipient is the cause of immunological response and reject. Today it's well known that different families of HLA genes are grouped together in different regions of the complex and which are divided in classes.

- Class I: Regions A, B, and C are highly polymorphic membrane glycoproteins, constituted by a heavy chain inserted in the cellular membrane and a light chain represented by microglobulin β2 coded by a gene located on chromosome 15. They have variable expression in different tissues, with high representations in lymphocytes T CD8, and are placed on all nucleated cells and on platelets.
- Class II: Regions DR, DQ, and DP, composed of two glycosylated polypeptide chains, have a smaller distribution and are present only on B lymphocytes, macrophages, some myeloid progenitors, some T lymphocytes, and cells of Langerhans. In complex with an HLA antigen on antigen-presenting cell (APC, typically a macrophage), they are often involved in the regulation of the immune response with T lymphocytes that can recognize a non-self-antigen. T Lymphocyte are divide into two classes: T4 and T8. The first are know as Helper while the second as Cytotoxic.
- Class III includes genes which encode for some fractions of the complement.

HLA complex, except the occurrence of crossing over that, intervenes in about 1.08% (Donors Candidate) of cases and is transmitted as a single block with the first law of Mendel. The combination of genes on the same chromosome is called "haplotype," and the whole of the two parental haplotypes (paternal and maternal) is called "genotype," determined only with a family survey and haplotypic segregation. The products of HLA genes were characterized with serological methods. Nowadays, instead, molecular biological methods are more used, such as the sequence-specific primers (SSP), sequence-specific oligonucleotides (SSO), Sanger sequencing, and next-generation sequencing (NGS). These products, commonly known as "HLA antigens," are very highly polymorphic. This entails a higher number of haplotypic combinations and an unlimited number of genotypes. However, polymorphism is restricted in the family, and there is a high probability of HLA compatibility in about 25% of siblings.

Therefore, to proceed with a transplant, proband and his family have to be previously typed for the search of a compatible subject. In case of negative results, it becomes necessary to research a compatible donor through international Bone Marrow Donor Registry (BMDR) and UCB stored in cord blood banks scattered around the world, managing possible donors (MUD or matched unrelated donor).

The research of a compatible donor has to match loci A and B at antigenic level, the DRB1 at allelic level. This gives also the possibility to have a rank that will be extremely precious and useful for the transplantologist, at the time to select the cord unit.

In clinical practice, CBU is often use as boost for an existing transplant, and UCB, MUD, or another familiar related will be used. This occurs due the rejection of the first transplant.

Umbilical Cord Blood (UCB)

In 1974, for the first time, it was possible to proof that into the cord blood there was a strong presence of HSC, used for transplantation purposes. The presence of HSC was afterward clearly demonstrated with a successful transplant in 1989 on a patient suffering from "Anaemia di Fanconi."

These stem cells are originated in the yolk sac, to migrate in the fetal liver and then in the BM. This last is the principal source of stem cells after birth. The general tissue immaturity at the birth, and the functional anatomical peculiarity of the maternal-fetal circle, contributes to confer specific characteristics to the cellular component of the placental blood both under the hematopoietic profile and under the immunological profile.

According to several studies, HSCs are more immature compared to the adult BM and with a higher proliferative potential. Moreover, it is characterized by a higher lymphocyte number compared to a peripheral blood type, with T lymphocytes much more immature and allogenically less reactive. All of these UCB characteristics give a lower risk of disappearance of GvHD in UCB transplant.

This tissue sample doesn't involve any kind of risk for the unborn and his mother; it can be collected once the childbirth is complete, as it's Cesarean section as it's natural childbirth. This occurs with the clamping of the funiculus and consequent cannulation of the umbilical vein to a closed system with sterile pouch.

When the collection is completed, the entire system is transported to the reference banking center (cord blood bank) defined "bank," where it is valued, processed, typed, and cryopreserved. The entire process can be carried out entirely on site or rely on bank reference efferent services.

The choice of cord unit for transplant is performed first of all verifying the compatibility of HLA for loci A, B, DRB1 and so assigning a score; then the quality of cryopreserved cells and their number are evaluated and finally they are validated post-freezing. For pediatric patients, namely, with an age between 1 and 16, it is required an infusion of two cord blood units with a pro kg cell dose of:

- Pre-cryopreservation total number cells (TNC) 3×10^7/kg in recipients of a single unit
- Defrosting 2×10^7/kg for single unit

Another fundamental value is the $CD34^+$ number, available inside the unit. Wagner et al. showed that UCB must have a minimum number of 1.7×10^5 CD34 cells/kg to permit an optimal engraftment. They demonstrated the existence of a correlation between the concentration of CD34 and TNC. The engraftment of the UCB unit was also reported by Van Heeckeren with a value of $\geq 1.2 \times 10^5$ CD34 cells/kg. To date, with the aim to carry out the evaluation of CD34, it is necessary to perform a flow cytometric survey by the use of a standardized protocol called International Society of Hematotherapy and Graft Engineering (ISHAGE). This is to enumerate in short times: stem cells of the peripheral blood, cord blood, or apheretic collection. The method provides not only marking of $CD34^+$ cells but

also the use of 7-aminoactinomycin D (7-AAD), a DNA marker, and use of fluorescent beads with known value. The gating strategy provides a physical dot plot using SSC vs. CD45 for establishing which are mononuclear cells. These cells are identified as $CD34^+$ and $7AAD^-$ in a new plot. Then they are all related to the number of beads of the kit. In this way, it is possible to establish the number of CSE present in the sample analyzed, the TNC, and viable cells.

Graft Versus Host Disease (GvHD)

The graft versus host disease (GvHD) is a kind of pathology that could arise after transplant, the donor's cells may not be recognized if the pharmacological protocol has not depleted all recipient immunological system cells.

This effect was noticed, for the first time, in 1956 when two mice died after some days that grafting was performed.

Nowadays, the incidence of this pathology is still high. It affects the 30–50% of the subjects who submit themselves to this treatment; in the case in which it's used as a donor, a familiar, or a MUD, its casuistry increases up to 80%, with a long-term survival rate inferior to 30%.

The white blood cells (WBC) involved in this response are lymphocytes $CD4^+$ and $CD8^+$, as well as natural killer; it has been suspected also of the involvement of inflammatory cytokines and of TNF (tumoral necrosis factor).

We can distinguish two levels of disease severity:

- Acute GvHD: It appears within the first 100 days of graft, on average at 15° day of marrow infusion. It's clinically characterized by skin rash, diarrhea, and liver dysfunction (cholestatic jaundice). All three organs are not necessarily interested; however, acute GVHD is given an overall degree of severity ranging from I to IV, in relation to organ compromise and to clinical manifestations.
- Chronic GvHD: It appears after 100 days of graft in comparison with the acute version. It can be a consequence of acute form or appear. It affects more areas, compared to the acute form, such as the skin, eyes, oral or esophageal mucosa, liver, lung, neuromuscular system, and intestine; it may hire characteristics of scleroderma, biliary cirrhosis, or obliterative bronchiolitis. Chronic GVHD can be limited or extensive based on organ impairment.

In order to treat this serious pathology, a strong prophylaxis is carried out providing the use of different classes of drugs such as methotrexate (MTX), cyclosporine (CSA), and corticosteroids.

IMTX is a cytotoxic drug that interferes with nucleotide synthesis of thymidine and purine. Initially used on canine transplant models, subsequently it was applied on human treatment. In 1977 the group of Seattle (Bleyer 1978) used the MTX in low doses following this scheme: 15 mg/m^2 giorno +1 post-TMO, so 10 mg/m2 in +3, +6, +11 days and after each week till day 100. In this way the group of Seattle

managed to reduce the incidence of GVHD by about 50%. The main complications, related to the use of MTX, are mucositis and medullary hypoplasia (Bleyer 1978).

CSA is a hydrophobic cyclic peptide of mycotic origin, is a immunosuppressant, reduces the interleukin 2 (IL-2) production by T CD4 lymphocytes, and interferes with cell receptor, thereby blocking the amplification of the alloimmune response. The use of this drug may be associated with renal toxicity, especially concomitant use of other nephrotoxic (antifungal, antibiotics, etc.) or in the presence of other complications such as sepsis or VOD. Moreover, CSA can determinate hirsutism, retinal lesions, hypertension, neurological disorders, liver dysfunction, hypomagnesemia, microangiopathic hemolysis, and myalgia.

Pre-analytic

A EFI-accredited (European Federation for Immunogenetics) or an ASHI-accredited (American Society for Histocompatibility and Immunogenetics) laboratory is required to perform chimerism analysis.

The collection of all the information about the blood samples as donor (UCB, MUD, Sibling) and receiving registry, date of transplant, and number of days since the first infusion (fundamental to evaluate the chimerism outcome) is of paramount importance; indeed, there are time points to assess the trend of engraftment:

- 10/20 days: engraftment's state.
- 60 days.
- 90/100 days: Confirm the engraftment.
- At the request of the clinician.

Q-PCR

Real-time quantitative PCR (Q-PCR), which is the quantity of PCR products, is measured in each cycle of reaction using fluorescent molecules. With this technique, it is possible to monitor the entire reaction from the beginning to exponential phase and define with extreme precision the DNA concentration. The application of Q-PCR to chimerism analysis arises from experience made on the gene expression studies. Real-time PCR involves the biallelic polymorphic markers of insertion/deletion (Ins/Del) distributed on several chromosomes, presenting a fair variability in the subjects.

The DNA amplification for the genome areas with insertion/deletion is done with a number of singleplex PCR, equal to the number of markers available for analysis: positive and negative control. Each reaction contains a TaqMan probe detectable by fluorescence, and each PCR has been studied to have a comparable efficiency, as indeed is comparable to the length of PCR products. For real-time PCR, in addition to the definition of the thermal program, it's necessary to well setup: threshold, baseline, reporter, quencher, passive reference, and chemistry of reagents. In order to

ensure the data reproducibility, in terms of specificity, sensibility, accuracy, and linearity, the DNA quality range (260/280 nm >1,8) and DNA quantity material (10–250 ng) must be established internally in the laboratory. This is *via* a process that allows to evaluate, define, and classify the quantity of DNA more adequate to use. Moreover, a no template control (NCT) test can be added to the result in order to avoid contamination of the sample.

Real Time quantitative PCR requires a specific setting defined below.

Reference DNA: For quantification, it should be used as a DNA containing the 100% of the receiving (quantification to evaluate the percentage of the recipient) or a DNA containing the 100% of the donor (quantification to evaluate the percentage of the donor).

Reference marker: For the quantification, it should be used always a reference marker, both for recipient's percentage calculation and for donor's percentage calculation.

- Informational/s marker/s: For the quantification, it should be used at least two informational markers for the recipient, if they are available, or an informational marker for each donor, if it's available.
- Define the real-time PCR analysis software; results are viewable in the form of amplification plot: this is a description about PCR analysis software and Graphical User Interface. For the pretransplant analysis through real-time PCR, the informativeness of a marker is given by the presence/absence of the amplification in the counterpart. The threshold cycle, Ct, of each informative marker gives a sensibility indication of that marker for that subject. For the posttransplant analysis, being a relative quantification, the calculation method is that of comparative Ct or Ct method; define the normalization method, using a validated algorithm given by markers (that can be more sensitive or more accurate) and knowing that markers can be subjected to chromosomal rearrangements/mutation, constitutional and/or acquired.

Short Tandem Repeats (STR)

Short tandem repeats (STRs) are a subclass of variable number of tandem repeat (VNTR); they are constituted by repeated sequences not higher than two to five of a base pair (bp). They are also called microsatellites. The STR differs from the other VNTR. The minisatellites are constituted by tandem repeat sequences of 20–100 bp; instead, the microsatellites are 2–5 pb. STRs are particularly informative and useful in forensic genetics and in clinical studies. Their wide distribution inside the human genome allows the chimerism analysis between patient and donor and also in the case of chromosomal rearrangements or loss of genetic information (monosomies or deletions); these generally are frequently in leukemia or in hematological disorders. These sequences are transmitted to lineage according to Mendelian laws, and they are located in DNA noncoding regions and don't get eliminated by natural selection. Moreover, STRs have a great variability, showing a clear heterozygosity in subjects.

The multiplex PCR is performed using a master mix (mixture of reaction with specific primer for each locus). More STR loci simultaneously are amplified in one reaction by this technique. Every specific primer has very similar annealing temperatures, and they are associated to fluorochromes for allelic discrimination. The data acquisition process is performed via the capillary electrophoresis. The electrophoresis peak needs to be converted in genetic profile (genotype), and one allele (in case of homozygosity) or two alleles (in case of heterozygosity) need to be assigned to each locus. Every peak is defined with a number, indicating the number of tandem repetitions of the "core" sequence. The conversion of the electropherogram into a genotype is carried out using the ChimerMarker™ software.

Standardization

The standardization is a crucial point to evaluate the quantity and quality of DNA extracted from blood samples or from cell suspensions (peripheral blood, BM, lymphocyte subpopulations). It is strongly recommended to perform a dilution serial of a high-quality DNA sample, starting from 10 ng to 0.25 ng, and on these dilutions set up a study of STRs. Now it is possible to establish the optimal concentration of DNA so that the fluorescence peaks are between 500 and 4000 relative fluorescence unit (RFU) range. It is appropriate to carry out this rating on a certain number of samples (10–20) and establish the optimal concentration of DNA to use in each experiment. If the DNA will be extracted by suspensions of lymphocyte subpopulations, the minimum quantity of cells to consider is about 1500/μL. The quality and quantity of DNA depends on the type of extraction performed. Therefore, it is more important to set up an intra-assay and inter-assay quality control (QC) with the aim of evaluating the data reproducibility. CQ intra-assay reproducibility is controlled by analyzing some samples and evaluating the 250 bp fragment of size standard, used as internal control.

The time of electrophoretic separation of this fragment must be similar in all of the analyzed samples, with a maximum variation of 4–5%. The CQ inter-assay reproducibility can be checked analyzing the same samples in different analytical sessions. The reanalysis of samples should report a variation of the area of the lower peaks at 3%. In order to constantly check the inter-assay variability, the sample already tested should be fully reproducible even in the next assay. It's important to define the methodic sensibility and simulate a chimeric state between blood samples of two subjects, one female sex (R) and one male sex (D).

In two samples, a blood count is performed and concentration of leukocytes is defined in each of them. Moreover, five or more mixes are set up at theoretical note percentage following the list below:

- Cycle 1
  - Mix 20%
  - Mix 40%
  - Mix 60%

  - Mix 80%
  - Mix 100%
- Cycles 2–3
  - Mix 25%
  - Mix 50%
  - Mix 75%
  - Mix 100%
- Cycles 4–5
  - Mix 90%
  - Mix 95%
  - Mix 97%
  - Mix 100%

The DNA is extracted by each mix and it is amplified performing the multiplex PCR. The test is repeated for four to five times, and the quantitative analysis of the area of the allele peaks of every markers is expressed as percentage of donor's cells (D) present in the sample analyzed. The sensibility of the dosage of STR should be declared by the manufacturer.

ChimerMarker™

ChimerMarker™ is a software analysis, commonly used in this field. It is characterized by a simple and intuitive interface, with the possibility to create different panels of analysis, to compare two different donors optimizing the quality of the analysis. The software is typically used to approach the follow-up analysis of the posttransplant chimerism. Indeed, after appropriate settings, it is able to identify the best allelic ladder to use for the analysis, the positive and negative control, comparing them independently with the reference standard, identifying donor and recipient. After the first stage, the software allows to follow the reference patient in time and make a dedicated longitudinal graph.

During the first stages, the user can easily choose to analyze a previously saved project or to start a new project, importing electrophoretic runs (fcs files). When the desired operation is set up, a screen appears with two lateral columns, in which there are the samples and loci of the kit in use and a central screen, where the raw data of the allelic ladder is reported. Selecting by the left column various samples, by double-clicking, it is possible to assess the data of each sample, activating and deactivating them. The toolbar on top side has a button to change channel (color) and assess them individually and simultaneously in the activated samples. There are four color channels (blue, green, yellow, and red) for sample markers and another one channel (orange) for the internal lane standard (ILS).

When the setup of the raw data is completed, the analysis can proceed and associates the allele peaks in the respective locus. In the toolbar on the top side, there is a green button "run"; clicking it a new screen will appear where it is possible to select the panel kit (*Powerplex_fusion_panel_v1.0*). At this stage, to perform the

chimerism studying, it is possible to simply recall the allelic ladder, and automatically all the available peaks will be associated to it. Moreover, ChimerPanel will be created and it will be specific for every project.

ChimerPanel® is a unique and useful tool; indeed it is built on alleles of the donor/s and recipient, showing and naming only peaks of interest for the analysis. Once the specific ChimerPanel is set up for the case study, the nonspecific peaks of stutter will be removed in each sample. When this clean operation is completed, the chimerism analysis can start on the mixes.

The real analysis is made with *single-donor chimerism analysis*, and the software using a specific formula will give a percentage of chimerism for each marker and an average percentage of all analyzed markers.

Now it is possible to ignore some markers, for some reasons we can exclude some loci from analysis and these values can help in this.

The chimerism calculation occurs in an automatic way and it is established with the following equation:

$$\%D = [(D1 + D2)/(D1 + D2 + R1 + R2)] \times 100$$

This formula is referred to an ideal case, in which the markers present both alleles of the donor and the recipient, their resulting heterozygotes. Often the markers are completely homozygous in both samples or also they are mixed. These cases are classified as follows:

- Loci Type I: Recipient and donor don't have common alleles. This locus type provides a more accurate and reproducible result. It is recommended to use only this type of loci, defined "fully informative," to calculate the chimerism percentage. In this case there are some variation:
  - Donor is homozygous and the recipient is heterozygous; the formula will be $\%D = [(D)/(D + R1 + R2)] \times 100$.
  - Donor is heterozygous and recipient is homozygous; the formula will be $\%D = [(D1 + D2)/(D1 + D2 + R)] \times 100$.
  - Both recipient and donor are homozygous; the formula will be $\%D = [(D)/(D + R)] \times 100$.
- Loci Type II: In this case, one marker allele is shared between recipient and donor while the other is in heterozygous status. If in the sample fully informative loci are available, as those described before, it is preferable to use those ones for the percentage calculation of chimerism.
- Loci Type III: In this case, recipient and donor are homozygous for a locus and the analysis isn't possible using it. This locus type has to be discarded for chimerism calculation and analysis.

Loci that are included in types II and III must be considered not very reliable, and so they have never to be considered in the analysis to have a truthful result.

In addition, the software allows to make a correction of this allele type through a useful tool for activation or not the deconvolution.

Deconvolution is a mathematics process based on an algorithm that improves signal and which finds application in many fields. In chimerism and STR study, the deconvolution applies a special algorithm to permit a correction of these shared alleles studied between samples. These alleles can give a truthful result on the percentage of chimerism for each available locus. After the "single-donor chimerism analysis" selection, on central frame will appear two options:

- "With deconvolution"
- "Without deconvolution"

The default function is set on "without deconvolution."

Indeed, in familiar posttransplant analysis, it is recommended to activate the "with deconvolution" function because donor and recipient are related (siblings/consanguineous). The first analysis is always set through the "without deconvolution" option. Upon activation of this option, we'll have the possibility to choose a few more settings to be able to make the finer and more real analysis.

As it is possible to see in Fig. 1, this software gives the possibility to exclude shared alleles, by selecting some options listed below (SoftGenetics LLC 2015):

- Shared allele imbalance (SAI): If selected, the marker will be ignored if a non-shared peak in the marker has higher RFU than a shared peak.
- With informative peak in stutter position (N-1): The marker will be ignored if in a donor or recipient samples, a shared peak is in a stutter position.
- Heterozygous imbalance: Ignores a locus if the smaller of two donor or recipient peaks is shorter than the specified percent times the height of the larger of the two peaks.

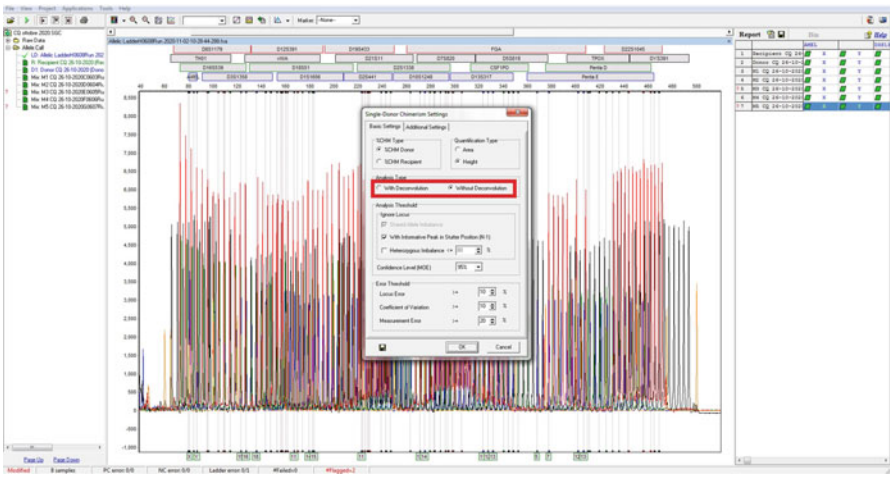


Fig. 1 "With deconvolution" is selected and other options are now available

The software can make an analysis on two different donors; in this case it will be enough to import the genotype of the second donor, do a new ChimerPanel, and remake the analysis using the "double donor Chimerism analysis" option. The software, activating this function, deactivates deconvolution option and all the alleles present will be considered.

Two pediatric clinical cases are reported below, in which a transplant was performed by MUD or by a familiar. Cases of CBU transplantation will be analyzed later.

Case Report 256

Pediatric subject suffering from β-thalassemia (Mediterranean anemia).

This pathology is caused by a mutation of the chain β of the hemoglobin. This mutation induces a form of anemia that can generate serious conditions. The gene involved is the HBB, positioned on chromosome 11p (short arm); it is a hereditary pathology with an autosomal recessive inheritance. β-Thalassemia is a widespread pathology in the Mediterranean basin with an incidence of porters of 15% in the coastal areas of Greece and Turkey and a 9–15% of the Italian coastal areas with a high frequency in regions of Sicily and Sardinia (Berg et al. 2013).

The β-thalassemia severity depending on the mutation is classified in two groups:

- Thalassemia minor: Only one allele is changed and the subjects suffer from microcytic anemia.
- Thalassemia major: All alleles are changed; the affected subjects have usually less than 2 years and are affected by microcytic anemia, hepatosplenomegaly, and splenomegaly. It has an incidence of 1/10,000 in Europe.

The therapy for affected patients provides continuous transfusions and finally the BM transplant.

The case study is a pediatric patient affected by β-thalassemia major, resulting from haploidentical to paternal grandfather and from whom the patient revived the marrow (familiar transplant). Therefore in this case the sample taken into consideration is to +1175 days from transplant.

Once the DNA extraction has been made and the electrophoretic STR performed, ChimerMarker™ software can start the fragment analysis.

First of all, the new project is opened and named with a sample ID of the case; the user launches a quick checkup to evaluate the quality of the run process. It's important to check if the negative control gives ILS peaks and if the positive control is compatible with known peaks, after activation of the allelic ladder panel, in order to name the peaks for the samples.

In Fig. 2 many peaks of markers are turned out to be shared between donor and recipient because they are related. This is a helpful data because it allows the identification and elimination of any unspecific peaks: It could also be turned into an issue, with decreases to the number of informative loci to use, loci of Type I.

When this "cleaning" phase is completed, the user can launch a single-donor chimerism analysis "without deconvolution" to make a rawer valuation. The number of informative markers is only 11 on 24 loci with a medium chimerism percentage to 100%: This because more alleles are to be shared and defined non-informative (Fig. 3).

Later in this first analysis, the analyst runs a new process called "with deconvolution": In this case, the number of informative alleles switches from 11 to 15, with a chimerism percentage that is lowered to an average value of 99.74%. This occurs due to the fact that two of the alleles were not considered before; indeed they had different height and area compared to the sample, and therefore they are excluded from analysis. Consequently, the percentage is lowered. Specifically, CSF1PO and D5S818 alleles are analyzed with an average chimerism value of 99.73% and 96.35%.

In detail, both of these loci have at least an allele shared between donor and recipient; therefore, they are classified as locus Type II. The software allows to make an analysis even more accurate, and it gives a general vision of the numeric values of height and areas of each peak. The analysis by default is made considering the area values and shown below (Fig. 4).

The peaks related to shared alleles are marked in green with the label "D1R"; their areas are wider than donor peaks labeled with "D" letter and blue color. It shows that repetitions of the same DNA transplant sequence (donor) aren't higher than the patient ones (recipient), and for this reason, there is a discordant result compared to other loci that have a value of 100%.

After getting the previous result, it's necessary to proceed to a second analysis which examines the heights of peaks and not the areas; in this case the result of total chimerism will be completed of 99.66%. Considering that informative loci will be 14, the CSF1PO is directly excluded by the software, while D5S818 is attributed a

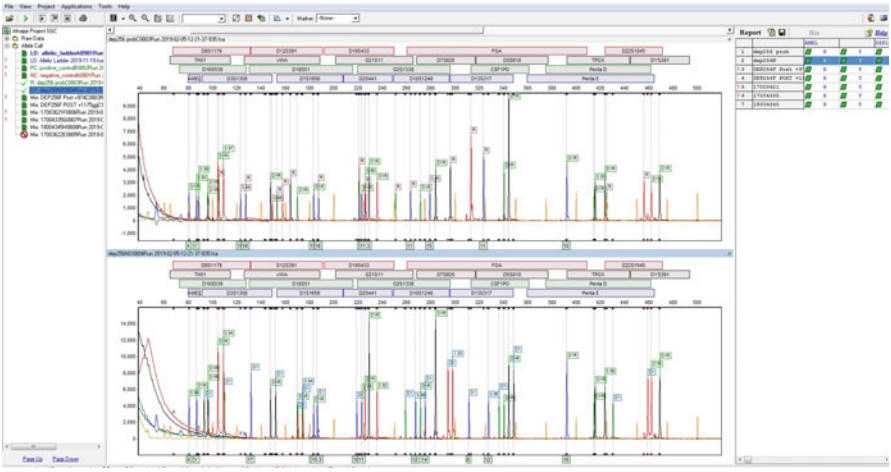


Fig. 2 Donor and recipient compared

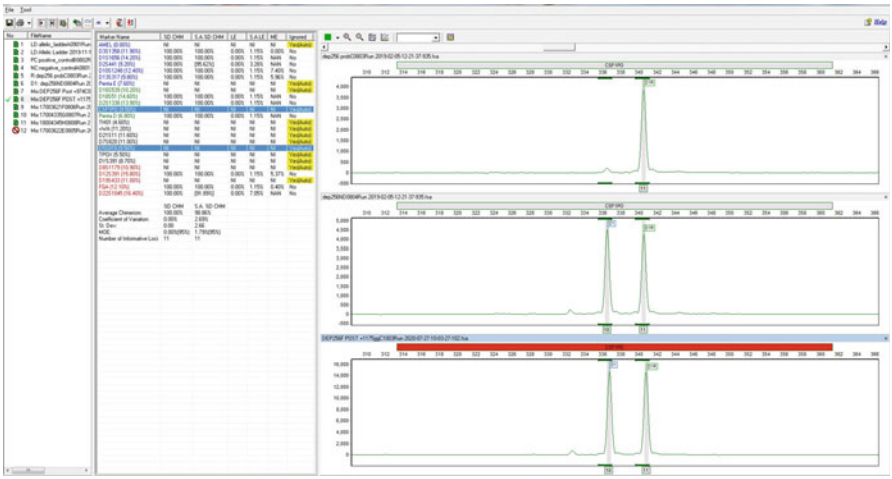


Fig. 3 "Without deconvolution" analysis. CSF1PO and D5S818 are selected

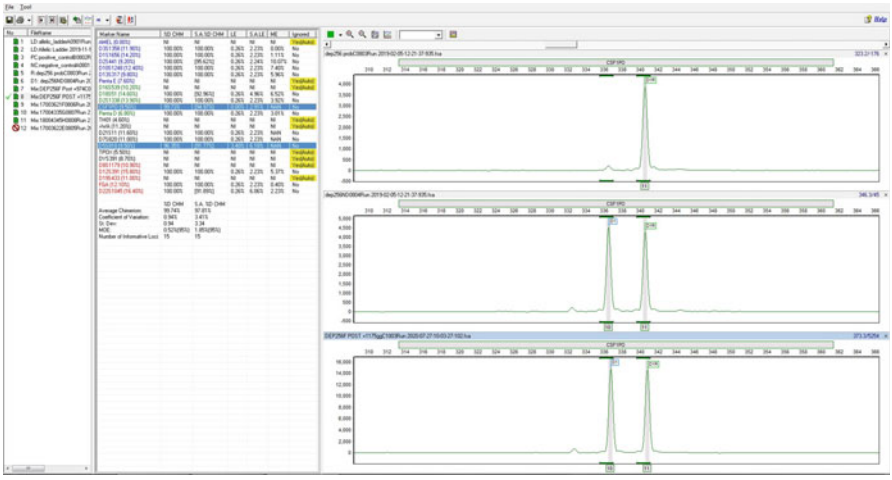


Fig. 4 "With deconvolution" analysis. CSF1PO and D5S818 are selected

value of chimerism equal to 95.30%. Also, in this case, analyzing the numeric values and heights, D1R allele is higher than D1; it proves that a portion of the recipient hasn't been eliminated by the transplant treatment and it's still present.

Continuing using the software, the peak area is set as the reference unit to obtain a truthful result. Two new options are now available for a deep research (while "deconvolution" mode is still running): First option allows to eliminate shared imbalance alleles (SAI) and the second one to eliminate stutter peaks in N-1 position. Now the informative alleles (locus Type I) are reduced to 9; those of Type II that

appeared previously are eliminated because SAI or it appears to be in stutter position. The total chimerism came back to 100%.

It is now possible to define that the case study of chimerism 256 has a chimerism of 100% (full donor) to 1175 days by transplant. The patient continues the checks on an annual basis to monitor chimerism.

Case Report 449

In this case, the 9-month-old pediatric subject suffers from a lymphohistiocytosis disease.

It's a rare disease that causes an immune dysfunction in babies and infants in an age minor than 18 months. Many patients have an underlying immune disorder, although in some patients the base disease is undiagnosed. Manifestations can include other pathologies as lymphadenopathy, hepatosplenomegaly, fever, and neurological anomalies. The diagnosis is obtained by specific clinical criteria and tests (genetics). There is a defect in cell elimination and inhibitory controls of natural killer cells and cytotoxic T lymphocytes, resulting in an excessive production of cytokines and accumulation of T cells and of macrophages in different organs. Cells in the BM and/or the spleen can attack red blood cells, leukocyte, and/or platelets.

The disease may be hereditary or acquired, and it is diagnosed when or the known gene mutation is identified or when five of the eight criteria indicated below are matched.

The criteria are as follows (Stadt et al. 2006):

- Fever (peak temperature of >38.5 ° C for >7 days)
- Splenomegaly (palpable spleen >3 cm under costal margin)
- Cytopenia that involves >2 cellular lines (hemoglobin <9 g/dL [90 g/L], absolute count of neutrophils <100/mcL [0.10 × 109/L], platelets <100,000/mcL [100 × 109/L)
- Hypertriglyceridemia (fasting triglycerides >177 mg/dL [2.0 mmol/L] or >3 standard deviations more than the normal value for age) or hypofibrinogenemia (fibrinogen <150 mg/dL [1.5 g/L] or >3 SD less to normal value for age)
- Hemophagocytosis (in biotypic samples of BM, spleen, or lymph nodes)
- Reduced or absent activity of natural killer cells
- Serum ferritin >500 ng/mL (> 1123.5 pmol/Lng/mL)
- High levels of soluble interleukin-2 (CD25) (>2400 U/mL or a very high value for age)

Instead, the more frequent mutation associated to pathology subtypes are as follows:

- FHL1: HPLH1
- FHL2: PRF1
- FHL3: UNC13D

- FHL4: STX11
- FHL5: STXBP2

The treatment is usually characterized by both chemotherapy and stem cell transplant. It entails, in the first instance, using high doses of cyclosporine, methotrexate, and immunoglobulin; in case of most severe cases, it is treated with hematopoietic stem cell transplant (HSCT).

The 9-month-old subject received a HSCT from a French umbilical cord blood unit. After 30 days by the process, the peripheral blood sample is analyzed to estimate the chimerism percentage. Using software, the user is able to estimate the electrophoretic run, checking up if the negative control gives ILS peaks and if the positive control matches with its previously added peaks. After that, the process proceeds with the activation of the allelic ladder panel, in order to name the peaks for the samples.

In this case, the "single-donor chimerism analysis" evidenced how there is a chimerism average in all loci of 100% with a higher number of informative loci, 19 on 24 loci. There are loci of Type I and Type II, which may be completely / partially informative, due to the fact that there isn't a kinship degree of common territorial origin, so probability of finding alleles in common is very low.

In Fig. 5, five non-informative loci are shown, and they are excluded from the analysis because they have identical alleles from donor and recipient samples.

Additionally, a deeper and accurate analysis can be started "with deconvolution" function by flagging the option "ignore shared imbalance alleles" in order to eliminate loci that have an allele shared between them (Fig. 6).

In this case, the number of alleles decreases from 19 to 12 fully informative (Type I). The average of chimerism results does not change and it is shown as 100%.

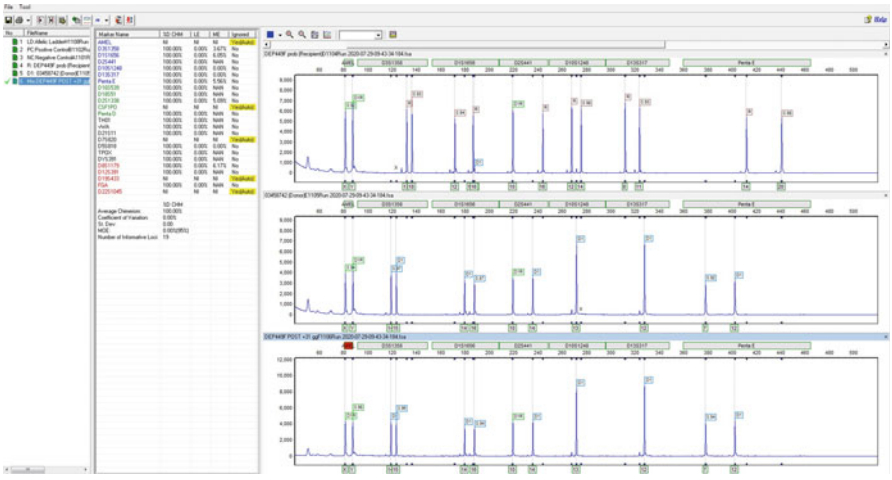


Fig. 5 "Without deconvolution" analysis. Nineteen informative loci

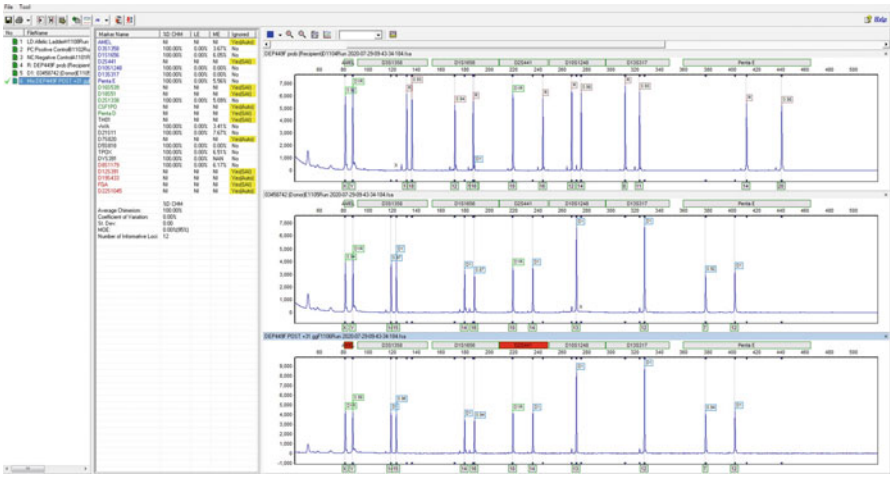


Fig. 6 "With deconvolution" analysis. Twelve informative loci

The patient died 10 days after the analysis despite having an optimal chimerism percentage, due to congenital skills already present at the birthing, and without correlation to the transplant procedure.

Cell Sorting

The magnetic-activated cell sorting (MACS) is a system used to perform a positive magnetic selection on the cells that are meant to have specific membrane antigen. With the aim to isolate each "population," 2 mL of whole blood or marrow is used. This is first evaluated using blood count to enumerate the nucleate cells and then to define the lymphocyte/granulocyte portion. All the systems follow a basic immunomagnetic principle: The cells have to be marked with a specific antibody conjugated to a bead, and the sample after incubation is eluted in a magnetic column that will allow the separation of the respective "population."

In the specific case, transplants of leukemic subjects are treated with two categories of markers: the $CD3^+$ for the lymphocyte series and the $CD15^+$ for the granulocyte lineage. Starting from 4 mL of whole blood, they are set up two vials, each one containing 2 mL of sample added with 100μL of specific antibody. After that, the incubation is carried out for 15 min to 4 °C, while the magnetic column for separation is set up for the next step.

A vertical block is used, by hooking a magnetic support with four housing; in each of them, a column for separation is positioned. The column is previously hydrated with a PBS albumin buffer solution and then left to rest until the end of incubation of samples is completed. After 15 min, the sample is dispensed into the column and left eluted; then 2 mL of buffer are slowly added in order to improve the fluidity and make the elution easier. At the end of this phase, beads with cells of

interest are linked to the magnet of the column, so, at this point, it is necessary to disconnect the column by the support. Now in the column, 4 mL of manufacturer elution buffer are added. The column is then placed on a collection tube, and applying a light pressure on the column, all beads will be released with the cells into the solution. Finally, the eluate efficiency is carried out before via blood count, to have an overview, using flow cytometry.

The cytometry evaluation provides a marking of two samples ($CD3^+$ e $CD15^+$) with a pattern recommended by the manufacturer.

The resulting cells have to be concentrated WBC (white blood cells) to form a pellet that can be used to extract the DNA having the genetic material of origin exclusively lymphocyte/granulocyte. The sample is centrifuged to 3500 rpm X 15 min at 4 °C; now it is ready for extraction using an automatic extractor eluting the DNA in 50μL.

CBU Case Report

The use of hematopoietic stem cells (HSC) to treat blood pathologies such as leukemia is the gold standard, but today it is characterized by several problems related to GvHD with a very high risk of relapse and also a high mortality ratio. The choice of donor, familiar or MUD (matched unrelated donor, by international register), can be of critical importance for the patient health as well as a crucial decision that totally depends on the transplantologist.

It's well known that UCB, compared to other origin tissues, has been less responsive and gave fewer adverse reactions with a reduced mortality of pediatric subjects. The cord blood is under study for the peculiarity to give lower allo-reactivity, and, for this reason, HLA compatibility between the donor and recipient is not only a condition for a successful transplant using this tissue (Keating et al. 2019). An essential requirement is also the presence of a very limited number of maternal origin cells of the cordon (Kanaan et al. 2017). During the fetal development stage, there is a continuous exchange of cells between maternal and fetus, through the placenta, and it is not uncommon finding maternal cells in the cord blood. For solidarity purpose, the cord blood is collected and cryopreserved. This cellular exchange induces the mother immune system to develop more resistant lymphocytes T and B and a minor number of antigens compared to HLA. In this way, the blood cord units can be used, but they must comply specific requirements: I) HLA compatibility with pediatric patients and II) presence of an adequate number of nucleate cells, as well as a high percentage of hematopoietic stem cells. Usually, a CBU to be used in a transplant should have 3×10^7 nucleate cells pro/kg, compared to the weight of the patient, and assumes that the 1% are $CD34^+$ HSC and the 33% are T cells. Therefore, 5×10^4 maternal cells T/ kg are transferred with every CB treatment.

These cells of maternal origin, which are present in very small quantities in the cord blood, are highly sensitized toward fetal antigens of father origin and, for this

reason, are defined inherited paternal antigens (IPA) (Ichinohe 2010). The proof of the presence of these cells is given by CBU study, picked up by some fetuses stillborn: The presence of cells of maternal origin in a concentration that is approximately 20% is registered. Another important evidence has occurred when these IPA cells started a graft versus leukemia (GvL) effect: It means that IPA started to attack "sick" cells, giving a positive effect and permitting the transplant to take over the illness.

Case Report 269

Pediatric subject of 7 years old affected by acute lymphatic leukemia (ALL).

ALL affects patient of whatever age but is the most frequent tumor in children and constitutes 75% of all leukemia affecting children under 15 years. The ALL generally affects children 2–5 years old. In adults, it is most frequent in subjects of 45 years old.

In ALL, leukemic blast cells accumulate in the BM, destroying and replacing normal cells. They reach through the bloodstream, liver, spleen, lymph nodes, brain, and testicles. ALL cells can accumulate everywhere into the body, wherever the cells can continue to proliferate. They can spread in layers of tissue that cover the brain and the spinal cord (leukemic meningitis) and can cause anemia, hepatic and renal insufficiency, as well as damage to other organs.

The disease manifests its symptomatology in evident and acute or in a sneaky way, with symptoms that could last for months.

Typical detectable symptoms are those due to neutropenia, thrombocytopenia, and anemia. For this reason, the appearance of recurrent infections, bleeding episodes or asthenia, tachycardia, and paleness should target the person toward a hematological control.

The ALL diagnosis requires a series of analysis to totally characterize the clinical-biological situation. This characterization is fundamental for choosing the best therapeutic strategy, and this is possible after performing morphologic analysis, cytochemistry, immunophenotypic analysis, and molecular genetics.

The morphological analysis of cells, that is the observation under the microscope of the streak of blood (periferic and/or medullary), is the first step in the diagnostic procedure. The target is to detect the presence of immature lymph node cells. If they are present in a higher percentage of 20% compared to the total, an ALL case is diagnosed.

The cytochemistry analysis based on the cell's relation to specific coloration might not work, because there is no specific test for the ALL.

A fundamental immunophenotypic analysis consists of highlighting some peculiar characteristics of leukemic cells. With the typing, we get to define the B or T phenotype of the cells, present with a frequency of 86% and 14%, respectively. It could be possible to define the phenotype of these cells as "hybrid" or "biphenotypic" leukemia.

Typically, the antigen $CD34^{+}$ is the most commonly used to identify the leukemic cell.

Molecular biology is a crucial method for diagnosis with cytogenetics analysis. These techniques, if used independently, could miss the identification of all the chromosomal abnormalities that are typically structural, largely translocation.

A typical translocation is known as Philadelphia chromosome, whereby in the exchange of genetic material between chromosomes 9 and 22, a new functional protein is born (BCR-ABL), the main cause of the disease. When this mutation is present, the disease is defined positive Philadelphia (Ph^+), negative when it isn't detected (Ph^-).

The human Philadelphia chromosome 22 is modified by the insertion of a terminal fragment coming by the chromosome 9. Following a translocation, the Abelson gene (ABL) passes by the chromosome 9 to the grouping region of the breaking points (breakpoint cluster region, BCR) of the 22 chromosome, with a formation of a chimeric gene Bcr-Abl. The mutant chromosome will be found in all the cellular lines and therefore can determine proliferative changes.

In 2017, the patient was submitted to UCB transplant and then was periodically checked every 30 days for the first 5 months and then each 3 months.

Even in this case, a chimerism analysis has been done, first of all by doing a ChimerPanel with donor and recipient alleles and then monitoring it during the time. This specific case was followed by two different centers to better monitor the posttransplant follow-up.

After the first 30 days in which the sample had a chimerism of 100%, it then changed over time until it stabilized 29–35%, having fluctuating values in the last 2 years. In the specific, analyzing two random samples of those available in posttransplant, it was clear that 16 loci on 24 are informative, with an average chimerism of 29% and a standard deviation of 2.55.

To date, the subject has not shown worrying symptoms to supposing a reappearance of the pathology. For this reason, we decided to better analyze his chimeric state, using a specific lineage cellular separation for lymphocytes (CD3) and granulocytes (CD15) exploiting the immunogenetic system.

We have used peripheral blood sample for perform MACS separation. The sampling timing is decided from Transplant Unit.

- +733
- +566
- +596
- +918
- +932
- +998

The graphs in Fig. 7 show that the total chimerism has a fluctuation from 42% to 29%; instead focusing on the specific lineage, it is possible to notice for the populations $CD3^+$ that there is an average value of 65% while for $CD15^+$ a value between 15% and 20%. These results and the absence of other symptoms for the case study lead to the conclusion that hematopoiesis of donor and recipient is healthy able

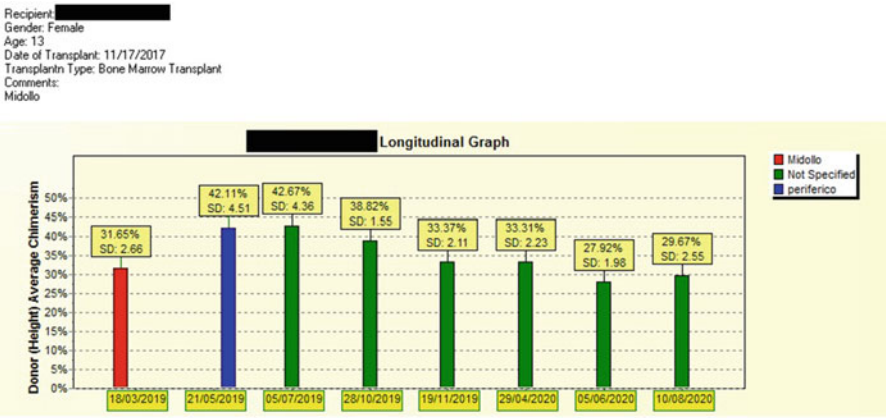


Fig. 7 Chimerism on whole blood samples

to resist in the same person and that the granulocytic line of the recipient has taken over on that of donor. The CD15+ percentage is 15%, while the lymphocyte line (CD3+), which is the lymphoid origin of the pathology, is approximately 65%; this means that the lymphopoiesis is mostly of the donor instead of the receiving. This state of constant mixed balance (mix of donor and recipient with different hematopoietic lines) has a positive prognostic index, compared to a full donor transplant, and often it is a clinician desired state (Figs. 8 and 9).

Case Report 176

The pediatric subject of 1 year old (born in 2014) suffering from acute myeloid leukemia (AML), transplanted in 2015.

AML is an acute leukemia most common in adults and its incidence increases with age. Despite it being a relatively rare disease, representing approximately 1–2% of deaths due tumors in the USA, its incidence is designated to increase with the progressive aging of the population (Deschler and Lübbert 2006).

The replacement of normal BM by leukemic cells causes a drop of red blood cells, platelets, and normal white blood cells. AML symptoms include fatigue, shortness of breath, bruising, bleeding, and increased risk of getting infections. In AML, there are different factors of risks and some chromosomal abnormalities are identified. AML is an acute leukemia rapidly progressing and has fatal outcome in a few weeks or months. This leukemia presents more subtypes with different treatment and prognosis between them. The acute myeloid leukemia is cured in 35–40% of under 60 population and 5–15% in those of older age. Elder people that cannot be treated with chemotherapy are characterized by an average lifetime of 5–10 months. The mortality index varies with the age (2.7 to 18 per 100.000 men and women).

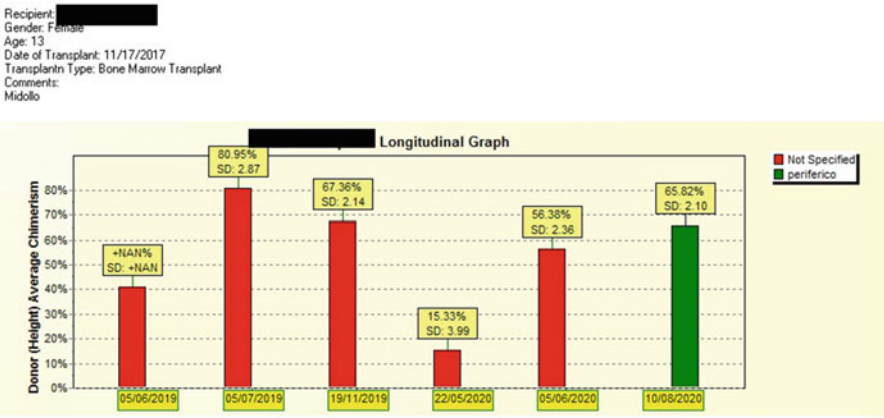


Fig. 8 Chimerism on CD3+ MACS samples

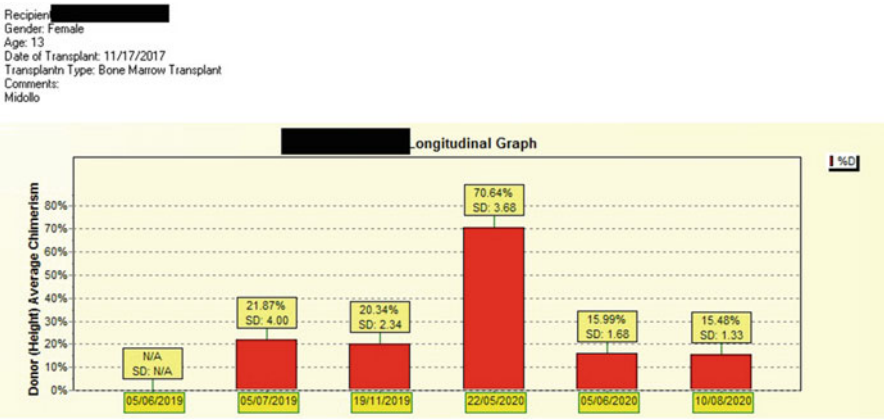


Fig. 9 Chimerism on CD15+ MACS samples

Myeloblasts are the tumoral cells in acute myeloid leukemia. In normal hematopoiesis, myeloblast is an immature precursor of myeloid white blood cell and gradually matures into white blood cell. However, in myeloid leukemia, a single myeloblast accumulates genetic changes that "block" the cell in its immature state and prevent differentiation. This mutation alone does not cause leukemia. However, when this stopping of differentiation is combined with other mutations that disrupt the genes that control proliferation, the result is an uncontrolled growth of immature cells which leads to the establishment of the clinical entity of acute myeloid leukemia.

Specific cytogenetic abnormalities can be found in many people with acute myeloid leukemia; abnormalities of chromosomes often have prognostic significance (Abeloff 2004). Chromosomal translocations usually encode abnormal protein transcription factors whose altered properties can cause "differentiation arrest." For

example, in acute promyelocytic leukemia, the chromosomal translocation 15;17 produces a PML-RARa fusion protein that binds to the retinoic acid receptor element in the promoters of several specific myeloid genes and inhibits myeloid differentiation (Melnick and Licht 1999).

Clinical signs and symptoms of the condition arise from the growth of leukemic cells that tend to shift or interfere with the development of the normal cells in the BM. This leads to neutropenia, anemia, and thrombocytopenia. Symptoms of acute myeloid leukemia are, often in turn, caused by the lower number of these elements in the blood. In rare cases, people with acute myeloid leukemia may develop a chloroma or a solid tumor in extramedullary leukemic cells that can cause different symptoms depending on its position (Ronald et al. 2013).

The first clue for an acute myeloid leukemia diagnosis is, in general, an abnormal result on a complete blood count exam, which is an excess number of abnormal white blood cells (leukocytosis), after a morphological exam of the peripheral blood colored with May-Grunwald solution and of Giemsa stain. The BM or the blood is examined under an optical microscope and with flow cytometry, to diagnose the presence of leukemia and to differentiate the acute myeloid leukemia from other types of leukemia (e.g., from the acute lymphoblastic leukemia) and to classify the disease subtype. A sample of blood or BM is tested also for chromosomal abnormalities of routine cytogenetics or fluorescence in situ hybridization (FISH).

The treatment subsists with chemotherapy, radiotherapy, and BM transplant.

All of this can be schematized in three phases: induction, consolidation, and maintenance. The induction phase has the purpose to eliminate as many neoplastic cells as possible and induce complete remission. The maintenance phase uses similar drugs but not cross-resistant to those used for the period of induction; this phase aims to eliminate the residual leukemic cells (in particular those that were in G0 phase and that have returned to replicate). The consolidation phase can be followed by the BM transplant. After the consolidation phase and/or the BM transplant, there is the maintenance phase; in this phase, we use cytostatic drugs to prevent the onset and/or the growth of new leukemia cells.

This pathology can be classified in seven different forms using a classification called Frank-American-British (FAB) depending on the level of undifferentiated of the blast and the cellular line of origin.

The patient received a transplant by UCB in 2015, the chimerism analysis wasn't carried out for the first 2 years, because it is not planned yet in the protocol after it was cadenced annually starting to the end of 2017. Have been identified 17 informative loci and so useful to analyze over time for each patient we perform a chimer panel. in this case there was not peak related to recipient (relapse). there was some peak unrelated to donor or recipient and we analyzed mother CBU. The patient never gave a sign of possible recovery of the disease during this period, so the research evidence of hematopoietic maternal cells coming from the cord unit blood was carried out. After recovering the DNA of the maternal cord, it was added to the analysis of its genotypic profile. The method provides the analysis of semiquantitative STR that points out a minimal presence of DNA of maternal origin at irregular intervals in different samples examined. Samples used for the follow-up are all

coming by peripheral blood, the marrow sample has never been analyzed, and it's interesting to notice how hematopoiesis of maternal cells is swinging in the analysis carried out. Below is a summary of samples analyzed with relative timing:

- MIX 1–2018
- MIX 2–2019
- MIX 3–2020

This time exploiting the function "double-donor chimerism analysis" is possible to use the genotype of two different donors to find them in the mix of the analysis.

For convenience, the recipient will be called R, CBU will be called D1, and the maternal donor of the CBU will be called D2.

In Fig. 10, starting from MIX 1 it is possible to appreciate a relative peak to allele 17 of locus D3S1358, albeit with a low score and a chimerism percentage of 11.26%. In locus D2S441 the allele 14 (D1D2) is much higher than allele 11 of D1; this observation, as explained above, makes us to suspect that is not previous only the DNA of D1 but also of D2, and that gives a peak far higher. The average percentage of chimerism results is 100% with 98.64% of origin CBU and 1.36% of maternal origin. In the other canals, it wasn't possible to identify the presence of more D2 peaks.

In MIX 2 the score of the D2S441 locus, allele 14, is slightly lower; its value is, in fact, identical to the allele 11 of the D1 because the amount of maternal DNA is lower, while of the D3S1358 locus, allele 17, chimerism continues to persist with an average of 7.34%. The total mean chimerism analysis is 100% with 99.13% of CBU origin and 0.87% of maternal origin.

Finally, in the MIX 3 no allele of maternal origin can be identified. Therefore, to be sure that hematopoiesis of the mother is exhausted, immunogenetics selection has

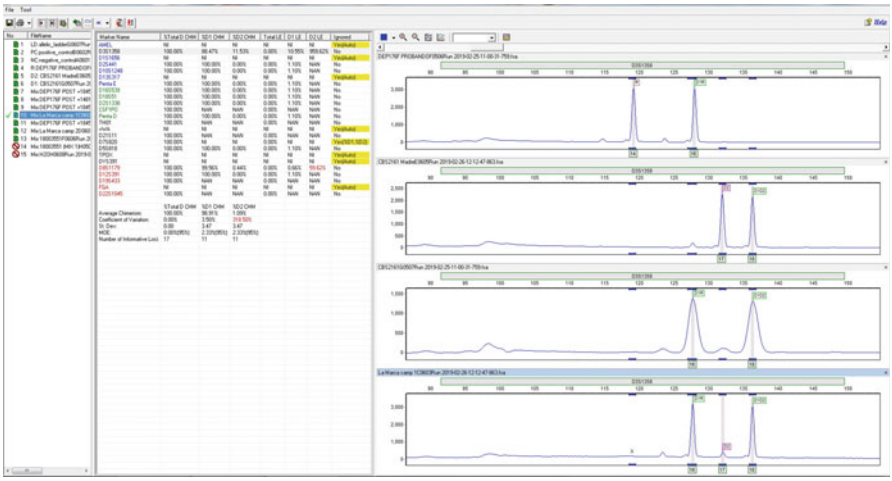


Fig. 10 Locus with double donors is selected

been carried out using MACS of $CD3^+$ and $CD15^+$ on a peripheral blood sample of the patient and extracting the DNA and the analysis of chimerism. In both cases, it wasn't possible to determine the presence of DNA of maternal origin because the maternal hematopoiesis disappeared or the concentration of maternal leukocytes is so low that it cannot afford an STR analysis. It could occur if the DNA concentration is lower from the threshold of identification method. After 1 year a new test will be performed on new peripheral blood tube.

Conclusion

There are many diseases (leukemia, thalassemia, anemia, etc.) that can be treated with HSCT of UCB origin, and the use of UCB doesn't require a high rigidity in the selection of HLA mismatch if it is compared with a MUD. Recent analysis of National Marrow Donor Program (American Donor Registry) has evidenced the possibility to find an HLA donor matched 8/8 unrelated adult donor and the results are reported below (Gragert et al. 2018):

- 75% for White European
- 41% for Chinese
- 40% for Korean
- 37% for Japanese
- 34% for Hispanics
- 19% for African Americans
- 18% for Africans

The UCB increases the probability for the patient of ethnic minorities to bring the HSCT and to find a compatible donor. The data provided by Center for International Blood and Marrow Transplant Research (CIBMTR) indicate that UCB is currently the 40% of unrelated donor transplantation in pediatric patients and the 10% in adult patients. Indeed, the lower alloreactivity of CBU allows search and select units with some mismatch HLA especially in pediatric transplant. Below is the list about HSCT by CBU in pediatric patient with HLA mismatch 4/6 (Gragert et al. 2018):

- 99% for White European
- 98% for Chinese
- 98% for Korean
- 97% for Japanese
- 98% for Hispanics
- 96% for African Americans
- 96% for African

The posttransplant chimerism study plays a paramount role in the improvement of a new therapeutic strategy, in particular in pharmacological therapy or when it is needed a second transplant to reinforce the previous one. Nowadays, the CBU is

widely used, in particular in pediatric subjects. Furthermore, the possibility of having this value added given by microchimerism has also a positive feedback on activities of umbilical cord blood banks.

A specific software, as the one reported in this study, allows a faster analysis and standardization of chimerism, which consents to improve the analysis quality, exploiting different tools and specific options according to the requirements. Moreover, the software provides a detailed and linear report during the time of the graft's trend. A real benefit of the software is the possibility to create a personalized panel for each patient and analyze solely the alleles and markers of interest. In addition, this function allows to identify low concentrations of DNA of a possible second donor, as it can occur in cases of maternal-fetal microchimerism or of double transplant.

The powerful tools of science, nowadays, provide us the possibility to divide the single cell lines with immunomagnetic methods, increasing the quality of the chimerism analysis. This state-of-the-art technology allows the detection of 200 cell/μL, which would not otherwise be identified impeding, in the past, the microchimerism identification. The current separation technique limitations are correlated uniquely to the antibody variety provided by the manufacturers.

The modern molecular biology techniques give us the opportunity to go deeper in life science, to extract DNA even in lower cell concentration, and software like ChimerMarker simplifies the chimerism analysis.

Acknowledgments D.M. was supported by the "Associazione Genitori Insieme ONLUS." Thanks go to Francesco Paolo Tambaro, M.D., PhD for having believed in this project and Antonio Di Maio, PhD for helping draft editing and critically revising the paper in keeping with intellectual content.

References

Abeloff M et al (2004). Clinical Oncology 3rd Edition (2004), pp 2831–2832

Andreani M, Testi M, Lucarelli G (2014) Mixed chimerism in haemoglobinopathies: from risk of graft rejection to immune tolerance. https://doi.org/10.1111/tan.12313

Bleyer A (1978) The clinical pharmacology of methotrexate. New applications of an old drug. https://doi.org/10.1002/1097-0142(197801)41:1<36::AID-CNCR2820410108>3.0.CO;2-I

Deschler B, Lübbert M (2006) Acute myeloid leukemia: epidemiology and etiology. https://doi.org/10.1002/cncr.22233

Gragert L, Eapen M, Williams E, Freeman J, Spellman S, Baitty R, Hartzman R, Douglas Rizzo J, Horowitz M, Confer D, Maiers M (2018) HLA match likelihoods for hematopoietic stem-cell grafts in the U.S. Registry. https://doi.org/10.1056/NEJMsa1311707

Kanaan SB, Gammill HS, Harrington WE, De Rosa SC, Stevenson PA, Forsyth AM, Allen J, Cousin E, van Besien K, Delaney CS, Lee Nelsona J (2017) Maternal microchimerism is prevalent in cord blood in memory T cells and other cell subsets, and persists post-transplant. https://doi.org/10.1080/2162402X.2017.1311436

Keating AK, Langenhorst J, Wagner JE, Page KM, Veys P, Wynn RF, Stefanski H, Elfeky R, Giller R, Mitchell R, Milano F, O'Brien TA, Dahlberg A, Delaney C, Kurtzberg J, Verneris MR, Boelens JJ (2019) The influence of stem cell source on transplant outcomes for pediatric patients with acute myeloid leukemia. https://doi.org/10.1182/bloodadvances.2018025908

Melnick A, Licht JD (1999) Deconstructing a disease: RAR alpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia. Blood 93:3167–3175

SoftGenetics LLC (2015) ChimerMarker user manual. SoftGenetics LLC

Stadt UZ, Beutel K, Kolberg S, Schneppenheim R Kabisch H, Janka G, Hennies HC (2006) Mutation spectrum in children with primary hemophagocytic lymphohistiocytosis: molecular and functional analyses of PRF1, UNC13D, STX11, and RAB27A. https://doi.org/10.1002/humu.20274

Diagnosis of Genetic Disorders by DNA Analysis

31

Parag M. Tamhankar, Vasundhara P. Tamhankar, and Lakshmi Vasudevan

Contents

P. M. Tamhankar (✉) · V. P. Tamhankar · L. Vasudevan
Centre for Medical Genetics, Mumbai, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_30

Abstract

DNA diagnostic tools have revolutionized modern medicine. Rapid and cost-effective technologies have changed the way genetic diseases are being recognized and diagnosed. Doctors are adopting these tests increasingly as a genotype – first approach to comprehensively screen patients instead of multiple costly, invasive tests or tests requiring anesthesia in children such as MRI, etc. In cases without a diagnosis such as fever of unknown origin or chronic illness without a diagnosis, doctors are increasingly adopting these technologies to find answers. The challenges lay in betterment of software technologies for genotype and phenotype correlation so as to minimize the uncertainty in diagnosis and availability of functional analysis of novel variants for betterment of understanding of molecular pathophysiology of these diseases.

Keywords

Genetic disorders · Diagnosis · PCR · Sequencing · DNA testing · Chromosomal disorders

Introduction

Human genetic disorders are the diseases caused by changes in the sequence and/or composition of its DNA. Genetic disorders could be broadly divided into chromosomal diseases and single-gene diseases. Chromosomal diseases are caused by error in the number or structure of chromosomes, whereas single-gene diseases are caused by errors in the sequence of genes. Genetic disorders are diagnosed by clinical presentation and can be confirmed by performing genetic testing. Genetic testing can be of various types: cytogenetics which includes study of chromosomes, biochemical genetics which includes testing various metabolites or enzymes to detect inborn errors of metabolism, and molecular genetics which includes DNA diagnostic techniques.

The concept of hereditary traits was published by Gregor Mendel in 1865. In his honor, the single-gene disorders are called as Mendelian diseases. In 1902, Sir Archibald Garrod published a book on alkaptonuria an inborn error of metabolism leading to accumulation of homogentisic acid in cartilage leading to premature osteoarthritis. He proposed it as a recessive hereditary disorder occurring primarily in offspring of cousin marriages. This disease was simply diagnosed by demonstrating darkening of the urine on exposure to air. This was the first genetic disease diagnosed by a simple genetic test. Later on, in 1909, the concept of gene was proposed by Wilhelm Johannsen. Joe Hin Tjio (1956) demonstrated that there are 46 chromosomes in human cells and revolutionized modern cytogenetics. In 1959, the French physician Jerome Lejeune identified Down syndrome as a chromosomal condition caused by 47 chromosomes in each cell. The DNA was first extracted by Swiss physician Friedrich Miescher in 1869. In 1949, Linus Pauling identified

defective hemoglobin molecule as the basis of sickle cell disease (Eaton 2020). In 1977, Frederick Sanger and colleagues pioneered the method of DNA sequencing using chain terminating inhibitors. The first DNA-based diagnosis was done by Orkin et al. (1982) for sickle cell disease by restriction fragment length polymorphism using enzyme M*st*II. In 1985, Kary Mullis invented the process named as polymerase chain reaction wherein a small amount of DNA can be copied into large quantities using primers, polymerase enzyme, nucleotides, and heat cycling.

In February 2021, the Human Genome Project celebrated the 20th anniversary of its completion wherein the first draft of the human genomes nearly complete sequence was published in Nature and Science, two most prestigious journals. The project identified and sequenced the approximately 20,500 genes in human DNA and determined the sequence of the 3 billion letters that make up human DNA. The DNA can be compared to a book, the chromosomes being the chapters, the genes being the pages, and the actual written material being the code. The entire genome contains six billion letters, the protein coding genes or exomes contain around 35 million letters (35 MB data), and each gene contains few letters to thousands of letters. These letters or nucleotides are of four types, adenine (A), guanine (G), cytosine (C) and thymine (T).

Thereafter Mark Schena and Patrick Brown at Stanford University in 1995 invented the chromosomal microarray, a collection of microscopic DNA spots on semiconductor chip to measure the copy number of chromosomal loci or the gene expression. Later, around 2006 Shankar Balasubramanian and David Klenerman co-invented the new DNA sequencing technology Solexa sequencing now known as next-generation sequencing technology that used highly sensitive lasers to capture base by base addition of the nucleotides by the polymerase enzyme to generate millions of sequences of shredded DNA molecules that could be aligned by bioinformatics to produce a continuous DNA sequence. Sarah Ng and Jay Shendure at the University of Washington pioneered exome sequencing using next-generation sequencing technologies to identify the genetic basis of Freeman-Sheldon syndrome. Genetic testing technologies can be compared to counting chapters (karyotyping), counting pages (microarray), or reading the book (sequencing page by page: Sanger sequencing or massively parallel sequencing, next-generation sequencing technologies).

DNA Testing for Classical Chromosomal Disorders

Chromosomes are the units of heredity. There are 23 chromosome pairs in each cell making up the diploid DNA complement. Disorders of chromosomes include either numerical abnormalities or structural abnormalities. Aneuploidy occurs when there is an extra or less copy of one or more chromosomes. For example, trisomy occurs when there are three copies of a chromosome instead of normal two, whereas monosomy occurs when there is single copy of the chromosome instead. Polyploidy occurs when there is an extra haploid set of chromosomes, for example, triploidy (69 chromosomes). Traditionally chromosome analysis or conventional cytogenetics

involves culturing cells such as peripheral blood lymphocytes in suitable medium, arresting the cells in metaphase using colchicine, harvesting the cells, preparing the cells for slide using hypotonic saline followed by fixation, later dropping the cells on the slide to prepare metaphase spreads, and staining the slides using sequential Giemsa stain-trypsin-Giemsa stain (GTG banding) to give the chromosome threads an alternating light and dark band appearance. Using a light microscope, camera capture, and software-assisted chromosome identification, the chromosomes are arranged in a photograph in systematic fashion to give us the karyotype. Chromosomal structural disorders such as translocation, inversion, ring chromosome, deletions, or duplications can be also diagnosed on a karyotype.

Typically chromosome disorders are suspected when a child has developmental delay or facial features suggestive of Down syndrome (47,XY,+21), or proportionate short stature/delayed puberty/amenorrhea/infertility in a girl child/woman when Turner syndrome (45,X) is likely, or male hypogonadism/infertility when Klinefelter syndrome (47,XXY) is likely. Recurrent abortions or miscarriages can be attributed to either husband or wife being carrier of balanced chromosomal translocation. This involves a mutual exchange of chromosomal segments between two different chromosomes. In this case, two acrocentric chromosomes are involved, a Robertsonian translocation results. The commonest chromosomal translocation is t(13;14). For an anomaly to be detected on chromosomal karyotype, at least 3–5 megabases of DNA letters have to be missing or extra. This is the resolution of a karyotype. Fluorescence in situ hybridization is another conventional cytogenetic technique, wherein fluorescently labeled DNA probes are used to count chromosomes/parts of chromosomes on a slide. The cells are usually colored fluorescent blue by 4′,6-diamidino-2-phenylindole (DAPI), and the chromosome probes are usually Spectrum Red (Texas red), Spectrum Orange, Spectrum Aqua, Spectrum Green, Spectrum Gold, or Spectrum Blue in color.

Rapid and efficient DNA diagnostic or screening techniques for chromosomal disorders include chromosomal microarrays, quantitative fluorescent polymerase chain reaction (QFPCR), BACS-on-Beads (BoBs) technology, noninvasive prenatal testing, and multiplex ligation dependent probe amplification (MLPA). Chromosomal microarray or array comparative genomic hybridization technology is used to count copy number of thousands of genes and chromosomal loci on a semiconductor microchip. The patient DNA is labeled with green fluorescent signal, and the reference DNA is labeled with red color. The patient DNA and reference DNA are co-hybridized onto the chip and scanned for the fluorescent signal using a high-resolution camera. If the patient DNA and reference DNA have equal number of loci/genes, a yellow color will be seen; if there is deletion of a particular chromosomal locus in the patient, a net gain of red color is seen; and if there is duplication of a particular chromosomal locus in the patient, a net gain of green signal is seen. Thus a log ratio is calculated by software, and the statistical significance of the deletion/duplication can be estimated. The microarray chips are of two types primarily depending on the types of probes used. If bacterial artificial chromosomal clones (BAC clones) are used as probes, it's called an array CGH. In case single-nucleotide

polymorphisms are used as probes, it's called as SNP array. The SNP arrays not only help find the copy number variations in genes but also helps in loss of heterozygosity (LOH) determination, detection of uniparental isodisomy (UPD), and mapping recessive genes using regions identical by descent. Affymetrix chips are primary SNP arrays, and the diagnostic chips are of following types CytoScan Optima (or 315 K chip), CytoScan 750 K, and CytoHD Scan and CytoScan XON array. CytoScan Optima is an array featuring 315, 608 probes (including control probes, 18,018 copy number markers and 148,450 SNP markers) giving whole genome coverage with optimal focus (25 markers/100 kb) on 396 relevant regions for applications related to prenatal diagnosis or for products of conception to identify chromosomal disorders. It offers a minimum resolution of 1 MB for losses, 2 MB for gains, and 5 MB for LOH/AOH. The CytoScan 750 K is a high-resolution chip which contains 750,000 markers including 5,50,000 copy number markers for constitutional and cancer genes and around 2,00,000 genotype able SNP probes. The CytoScan HD is an array which contains 2.6 million markers (1.9 million oligonucleotide probes and 7,50,000 SNPs). The probe spacing is one oligonucleotide every 2 kb and one oligonucleotide every 400 base pair in focused regions. There are 200 SNP probes per MB. The CytoScan XON is an exon-level copy number solution and is designed to cover the entire genome with a focus on 7000 clinically relevant genes. All the probes are 25 base pair long; each SNP is targeted by six probes (three probes per allele). The Chromosome Analysis Suite (ChAS) software is used for analysis.

The derived copy number variants are compared to established databases such as DECIPHER, Database of Genomic Variants, PubMed, Online Mendelian (2021) Inheritance in Man, and PubMed to know their clinical significance. Known syndromes or microdeletion/duplication syndromes are reported as such. CNVs with loss of exonic region of a gene correlating with the phenotype was reported as likely pathogenic and if within a gene not correlating with the phenotype was reported as variant of uncertain significance. CNVs reported in control population were considered as benign. Loss of heterozygosity more than 10 megabases was considered significant, and segmental uniparental disomy was considered possible. Confirmation of segmental UPD leading to imprinting disorders would require methylation study of the imprinted genes. Example for segmental UPD covering 15q11.2 region, methylation study of SNRPN gene is advised to know whether it is paternal UPD (leading to Angelman syndrome) or maternal UPD (leading to Prader-Willi syndrome). LOH can be classified as pathologic when the entire chromosome was in LOH (except X chromosome in males). Where possible, CNVs were advised to be confirmed by an alternative technology such as MLPA or QFPCR or FISH.

Quantitative fluorescent PCR includes amplification, detection, and analysis of short tandem repeats which are polymorphic repeats. After the PCR is done using fluorescent primers for the STRs, the product is run on a capillary sequencer. The number of chromosomes is derived from the number of peaks seen in each STR or the relative heights of the peaks in STR. For example, in case of normal copy number of chromosomes 21 (disomy), two peaks of STR alleles should be seen with equal

heights; in case of trisomy, three peaks of STR alleles (three different sized alleles) can be seen, or two peaks can be seen with height ratio of 2:1 (if the size of the STR is common to two chromosomes and different in one chromosome). Common kits for rapid prenatal diagnosis of chromosomal numerical errors of common chromosomes 13, 18, 21, and sex chromosomes include Devyser commercial QFPCR kits (Devyser website 2021). These kits can also be used for rapid detection of maternal cell contamination in DNA samples prior to microarray analysis.

BACs on Beads technology is a new technology for rapid prenatal diagnosis of chromosomal disorders. It uses bacterial artificial chromosome clones attached to dyed microspheres. It contains probes for 75 chromosomal regions involving the 13, 18, and 21, sex chromosomes, as well as common microdeletion syndromes such as DiGeorge, Williams-Beuren, Prader-Willi, Angelman, Smith-Magenis, Wolf-Hirschhorn, Cri du Chat, Langer-Giedion, and Miller-Dieker syndromes.

MLPA is a single-tube multiplex polymerase chain reaction that uses a single primer pair (one of them is fluorescent) to amplify up to 60 probes, each corresponding to a unique genomic target and length. For example, detection of exonic deletions or duplications will require probes for various exons, and detection of subtelomeric deletions or duplications will require probes for various gene/s in the subtelomeric region of the chromosome. The PCR amplicons are fluorescent (since one of the primers is fluorescent), and hence they can be run and detected in a capillary sequencer (Fig. 1). By comparing the resulting pattern of peaks in a sample to the pattern of a set of reference samples, the copy number analysis is derived using a standard software named as Coffalyser.

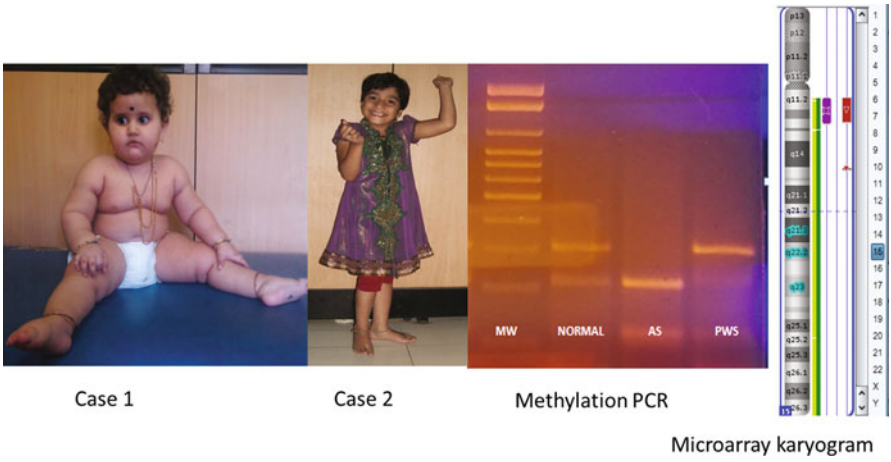


Fig. 1 Methylation PCR and chromosomal microarray test: Angelman and Prader-Willi syndrome. Case 1 shows child with Prader-Willi syndrome, and case 2 shows child with Angelman syndrome (AS). Methylation PCR for SNRPN gene shows two bands, upper being maternal and lower band being paternal; maternal band is absent in AS and paternal band is absent in PWS. Microarray karyogram shows the chromosome 15q11.2 deletion depicted by red band

Common Microdeletion Syndromes

DiGeorge Syndrome/Velocardiofacial Syndrome (22q11.2 Deletion)

Children with 22q11.2 deletion syndrome have congenital heart disease such as conotruncal malformations, ventricular septal defect, palatal abnormalities, immune deficiency, facial features such as elongated face, small low set ears, small palpebral fissures, bulbous nasal tip, short or flattened groove in upper lip, and learning difficulties. The chromosomal region is divided in to following segments A-B-C-D-E-F-G-H, majority (85%) are typical large 3 megabase deletions A to D, and common nested deletions are A to C (2 MB), A to B (1.5 MB), and B to D. Most (90%) of large deletions are sporadic, whereas 60% of nested deletions are inherited from either parent. The genes included in this region are *TBX1* (congenital heart defects), *DGCR8* (learning difficulties), *CRKL* (NK cell immune deficiency), *SNAP29* (learning defects), and *PRODH* (learning defects). The diagnosis is done by commercially available FISH probes N25 and TUPLE, along with *TBX1* located between locations A to B. Hence atypical deletions such as B to C or B to D are not identified by FISH. DNA diagnosis can be done by MLPA probe set SALSA Probemix P250 which contains the following probes region A to B, *CLTCL1*, *HIRA*, *CDC45*, *GP1BB*, *TBX1*, *TXNRD2*, and *DGCR8*; region B to C, *ZNF74*, *KLHL22*, and *MED15*; region C to D, *SNAP29* and *LZTR1*; region D to E, *HIC2*, *PPIL2*, and *TOP3B*; region E to F, *RSPH14*, *GNAZ*, and *RAB36*; region F to G, *SMARCB1*; and region G to H, *SNRPD3*. These can diagnose typical as well as atypical deletions.

Prader-Willi Syndrome/Angelman Syndrome (15q11.2-q13 Deletion)

These are two distinct imprinting disorders linked to deletions at the chromosomal locus 15q11.2-q13. PWS is characterized by severe hypotonia/floppy child, feeding difficulties in infancy to excessive eating and obesity in adolescence, delayed milestones, mild to moderate cognitive impairment, and behavioral abnormalities. AS is characterized by severe developmental delay, speech impairment, ataxia or tremulousness, “happy puppet”-like behavior (frequent smiling/ laughing), microcephaly, and seizures. Deletion of the paternal allele leads to PWS, whereas deletion of maternal allele leads to AS. The chromosomal region is divided into following segments BP1-BP2-BP3 deletions. Typical deletions range from BP1 to BP3 region around 5–7 MB, whereas atypical are from BP2 to BP3. Other mechanisms that lead to these disorders are uniparental disomy (UPD: maternal UPD leading to PWS and paternal UPD leading to AS), i.e., both chromosomes inherited from same parent, imprinting center defect (with or without deletion), and unbalanced chromosomal translocations. In the case of PWS, deletion accounts for 65–75% cases, maternal UPD 20–30%, imprinting defect (ID) without imprinting center (IC) deletion 2%, ID with IC deletion 0.5%, and unbalanced chromosomal rearrangement 2%. In AS, deletion accounts for 65–75% cases, paternal UPD 3–7%, ID without IC deletion

2.5%, ID with IC deletion 0.5%, unbalanced translocation 2%, *UBE3A* gene sequence variant 11%, and other unidentified mechanisms in 10–15% cases. The recurrence risk of PWS/AS in future sibs in cases of interstitial deletion is 1%, maternal/paternal UPD is 1%, unbalanced translocation up to 50%, t(15;15) also known as Robertsonian translocation (mother carrier in the case of PWS and father carrier in the case of AS) 100% risk, ID with IC deletion 50% risk (mother may be carrier in AS, father may be carrier in PWS), ID without deletion 1% risk, and *UBE3A* gene sequence variant 50% risk if mother is carrier. Detection of interstitial deletions is possible by FISH/chromosomal microarray/MLPA. FISH or SNP chromosomal microarray has 65–75% detection rate in AS and up to 90% in PWS. Detection of UPD is possible by chromosomal microarray using SNPs which will show LOH region at that location. Detection of UPD is also possible by short tandem repeat (STR) analysis which uses polymorphic markers and compares the markers between the child and parents. Detection of ID with/without IC deletion is possible by methylation analysis (PCR or MLPA). MRC Holland MLPA kit ME028 is used for detection of PWS/AS and has 99% sensitivity to diagnose PWS and 80% sensitivity to diagnose AS. Sequencing of *UBE3A* gene is used to detect point mutations in the gene.

Subtelomeric Deletion/Duplication Syndromes

Telomeres are the ends of the chromosomes, and rearrangements at these regions are quoted to be responsible for 5–7% of all cases of mental retardation. Deletion of some ends of chromosomes causes a specific phenotype. For example, Wolf-Hirschhorn syndrome due to deletion of chromosome 4p16.3 is clinically characterized by pre- and postnatal growth deficiency, developmental delay of variable degree, characteristic craniofacial features (Greek helmet appearance of the nose, high forehead, prominent glabella, hypertelorism, high-arched eyebrows, prominent eyes, epicanthic folds, short philtrum, downturned corners of the mouth and micrognathia), cleft lip/palate, eye coloboma, and cardiac defects. The critical genes within these regions are *WHSC1* and *LETM1*. Another example is alpha-thalassemia – mental retardation ATR-16 syndrome which is caused by deletion of chromosome band 16p13.3 which removes the alpha globin genes *HBA1* and *HBA2*. Patients present with alpha-thalassemia trait or mild hemoglobin H disease, facial dysmorphism, genital anomalies, short stature, and intellectual disability. The *SOX8* gene deleted in this region contributes to mental deficiency. Using a mapping MLPA kit, the critical region has been narrowed to an 800 kilobase region localized between 0.9 and 1.7 megabases from the telomere (Harteveld et al. 2007). The various MLPA probe sets available to detect all chromosome subtelomeric deletions or duplications are P036 and P070, and probe sets for common microdeletion syndrome is P245 and P371–P374.

DNA Testing for Classical/Important Single-Gene Disorders

Beta-thalassemia

Beta-thalassemia is one of the commonest single-gene disorders in India and several parts of the world. Mutation in HBB gene on chromosome locus 11p15.4 that produces the beta globin chain of hemoglobin molecule leads to ineffective erythropoiesis and chronic hemolytic anemia. Children with the thalassemia major form of the disease with two deleterious mutations on both the maternal and paternal allele present usually within the first year of life with growth retardation, chronic anemia, hepatosplenomegaly, and skeletal manifestations such as frontal bossing and decreased bone density. The clinical diagnosis is achieved by hematological tests showing anemia, microcytosis, target cells in peripheral smear, and increased hemoglobin F (fetal) on high-performance liquid chromatography for hemoglobin variant analysis. The HBB gene consists of three exons. Around 400 unique mutations have been described in the HBB gene. In the Indian cohort of patients, five common mutations account for 85% of mutation spectrum IVS1 + 5G > C, IVS1 + 1G > T, codon 8/9 (+G), codon 41/42(-CTTT), and 619 base pair deletion. The first four mutations are initially screened by amplification refractory mutation screening (ARMS) a technique, wherein their PCR is done using specific primers for each mutation separately. The primers include a common primer and a mutant/wild-type primer, which are matched to the mutation/ wild-type nucleotide at the 3′ end. For example, for detecting the IVS1 + 5G > C mutation, there will be two sets of PCR reactions: (1) to detect the wild-type G nucleotide, the PCR reaction will employ common primer and a wild-type primer with C at the 3′ end complementary to the G nucleotide. (2) To detect the mutant nucleotide C, the PCR reaction will employ the common primer and a mutant type primer with G at the 3′ end complimentary to the C nucleotide. The 619 base pair deletion can be detected on PCR using agarose gel electrophoresis by determining the size of the PCR products (Tamhankar et al. 2009). Other techniques used for detecting common mutations include reverse dot blot hybridization technique (El-Fadaly et al. 2016). An example of commercially available is beta globin StripAssay (Vienna Diagnostics). PCR is performed with biotin-labeled beta globin mutation and wild-type specific primers for detection of common 22 mutations (Mediterranean area) in a single tube followed by hybridization to a test strip containing wild-type and mutant oligonucleotide probes immobilized as parallel lines. Bound biotinylated sequences are detected using streptavidin alkaline phosphatase and color substrates. In case patient does not have common mutations, complete HBB gene sequencing can be done for all three exons. In case sequencing does not detect a mutation, deletion-duplication assay (MLPA), for example, MRC Holland P102 probe mix, can be used.

Alpha-thalassemia

The adult hemoglobin HbA1 is composed of two alpha strands and two beta strands. The alpha chains are produced by the alpha genes, four in number, i.e., two each (HbA1 and HbA2) (αα/αα) on locus 16p13.3. Mutations or deletions in the alpha chain produce four distinct syndromes. Large deletions that knock out both genes on one strand (- -/αα) can be of following types: Southeast Asian deletion (--SEA), Filipino deletion (--FIL) and Mediterranean deletion (--MED). Homozygosity for large deletions ie (--/--) with all four alpha genes deleted or no alpha gene copy present on either chromosome, produces severe anemia in utero leading to hydrops fetalis, and the hemoglobin produced is hemoglobin Bart which is composed of four globin chains (ϒ4) instead of hemoglobin F (fetal hemoglobin, which is α2 ϒ2, composed of two alpha strands and two ϒ strands). Smaller deletions that remove one alpha gene are usually of the following types α 3.7 kilobase deletion (α3.7/αα) or α 4.2 kilobase deletion (α4.2/αα). The other clinical possibilities are three alpha genes deleted (--/-α) (large deletion on one chromosome and small deletion on other chromosome) which produces hemoglobin H disease resulting in microcytosis, anemia and splenomegaly. Hemoglobin H is an adult hemoglobin made up of four beta chains. In case, two alpha gene copies are present in cis (--/α α) or in trans (-α/- α), heterozygous alpha-thalassemia or alpha-thalassemia trait is produced. In case only one copy of α gene is deleted (α α/α-), the individual has silent alpha-thalassemia carrier state. There could also be non-deletional mutations or point mutations that could inactivate the alpha gene. Detection of large as well as small deletions is done by gap-polymerase chain reaction. The principle is that primers will form a PCR product if the large deletion is present since the complimentary regions come close due to the deletion, and if the deletion is absent, no PCR product is produced. Non-deletional mutations can be detected by Sanger sequencing of the alpha genes. Multiplex ligation-dependent probe amplification technology can also be used for detecting the large and small deletions as well as copy number analysis of *HbA1* and *HbA2* genes.

Cystic Fibrosis

Cystic fibrosis is an autosomal recessive genetic disease caused by mutation in *CFTR* gene on chromosome locus 7q31.2 which encodes a chloride transporter, which leads to viscous mucus in the epithelial tracts of the lungs, pancreas, and gut and absence of vas deferens leading to elevated sweat chloride, frequent respiratory tract infections, food malabsorption, failure to gain weight, and male infertility. Cystic fibrosis represents the first genetic disorder elucidated by process of reverse genetics or positional cloning. Rommens et al. (1989) used technique such as chromosome walking and jumping. Kerem et al. (1989) found that delta F508 (phenylalanine at 508) position accounted for 70% of Caucasian mutant alleles. Newborn screening is done by measuring immunoreactive trypsinogen on dried blood spots around 48 h of life using time-resolved fluoro-immunoassay (PerkinElmer AutoDELFIA). Patients with more than IRT 70 ng/ml are considered

positive on screening, and repeat IRT was measured around 20th to 30th day of life. In those with positive second screening test, sweat chloride (minimum sweat amount is 75 mg) is performed by iontophoresis of pilocarpine nitrate (Gibson and Cooke method) and measured by coulometric analysis. Sweat chloride tests were considered positive if more than 60 mEq/l (diagnosed CF), borderline range (30–60 meq/l), and negative (CF unlikely) (if less than 30 meq/l). In patients with positive sweat chloride test, or borderline results, CF diagnosis can be confirmed by DNA testing for *CFTR* gene mutations (Castellani et al. 2016; Farrell et al. 2017; Paracchini et al. 2012). More than 2000 mutations have been described till date (http://www.genet.sickkids.on.ca/cftr/app). Welsh and Smith (1993) classified mutations in four types: (I) defective protein production such as nonsense mutations p.G542X, frameshift mutations 3905insT, and splice site mutation c.489 + 1G > T (II) defective protein processing or failure to traffic to the correct cellular location such as deltaF508, p.S549R, (III) defective protein regulation such as mutations in the nucleotide binding domain such as p.Gly551Asp which alter the way ATP regulates the way channels open, and (IV) defective protein conductance such as mutations in the first membrane spanning domain p.R117H which lessens the chloride current generated. The traditional sweat chloride testing various strategies for screening for common mutations are employed. Inge Buyse et al. (2004) developed an extended cystic fibrosis panel for 51 known pathogenic *CFTR* gene mutations that incorporate 25 recommended *CFTR* mutations and 26 additional mutations including 3199del6 which was associated with I148T (Buyse et al. 2004). The I148T is now considered a benign allele after functional in vitro studies showed that this does not lead to alteration of CFTR protein function. Genomic DNA was isolated from blood. PCR amplification was performed in a 384 well format, extension of primers adjacent to the mutation of interest using defined combinations of deoxy and dideoxynucleotides, and the resolution of the corresponding wild-type and mutant primer extension products by their molecular weights. The platform used is matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry platform (Sequenom). The system can also detect novel alleles as aberrant peaks on the chromatogram that can be later confirmed on sequencing. System advantages include rapid turnaround time (24–48 h), high-throughput testing, and flexibility in assay design. Patients who turn negative for the common mutations may need complete sequencing covering all 24 exons and exon-intron boundaries. This can be done by either Sanger sequencing or may be included in clinical exome/whole exome panel.

Congenital Adrenal Hyperplasia (CAH)

This is a group of diseases including failure of adrenal gland to produce mineralocorticoids and glucocorticoids due to deficiency of enzyme producing them. The commonest type is caused by mutation in *CYP21A2* gene (21OHB) which encodes 21 hydroxylase enzymes. The gene is present on 6p21.33 chromosomal locus. There is a pseudogene *CYP21A1* (21OHA) present at the same locus along with major

histocompatibility locus class III genes. The arrangement is as follows centromere - - GLO - - DP - - DQ - - DR - - C2 - - BF - - C4A - - 21OHA - - C4B - - 21OHB - - B - - C – A. Due to sequence homology between the gene and the pseudogene, recombination slippage could occur leading to mutations involving gene-pseudogene conversion, gene-pseudogene fusion products, or gene deletion and pseudogene duplication. Point mutations in the *CYP21A2* genes can also be possible. The classic form of disease occurs due to severe enzyme deficiency or severe type of mutations (null), whereas the non-classic form occurs due to mild enzyme deficiency due to mild mutations (missense, inframe-dels/duplications). The classic form is of two types: type I salt wasting type (75%) which is due to deficiency of aldosterone in which patients present with dehydration and electrolyte disturbance, and virilization is also possible due to androgen excess and type II simple virilizing type (25%) in which there is no deficiency of aldosterone, only androgen excess. Patients with non-classic type can present in postnatal age with signs of androgen excess, advanced bone age, early puberty, masculinization, hyperpigmentation, and excessive body hair. Classic CAH is diagnosed in newborn period with elevated 17hydroxyprogesterone and elevated adrenal androgens. The single-nucleotide mutations c.293-13A > G or c.293-13C > G are among the most frequent mutations in classic CAH and create an aberrant acceptor splice site causing a shift in the translational frame. Other mutations accounting for classic CAH are whole gene deletion, large gene conversion, p.Gly111ValfsX21, p.[Ile237Asn; Val238Glu; Met240Lys], p.Leu308PhefsX6, p.Gln319X, and p.Arg357Trp. Among patients having the mutation p.Ile173Asn, 76% had simple virilizing type, while 23% had the salt wasting phenotype. The other mutations causing non-classic disease are p.Pro31Leu, p.Val282Leu, and p.Pro454Ser.

Spinal Muscular Atrophy

Spinal muscular atrophy is an autosomal recessive genetic disease involving progressive degeneration of the alpha motor neurons in the anterior horn cells of the spinal cord leading to muscular atrophy, weakness. The incidence of carrier state is common in general population (2% of all individuals, panethnic carrier rate), and hence the incidence of live birth is around 1 per 10,000. Depending upon the onset and progression of weakness, there are five forms. SMA type O presents at or before birth, survival is less than 6 months, type I presents in the first 6 months, median survival without treatment is 8–10 months, type II presents at 6–18 months, 70% of patients are alive at 25 years of age, SMA type III presents after 18 months, SMA IV presents in adulthood, and type III and type IV have normal life expectancy. The disease occurs due to mutation in SMN1 gene (chromosome locus 5q13). The *SMN* gene has a functional copy *SMN1* (also known as telomeric *SMN1*) and another copy (also known as centromeric *SMN2*) *SMN2* gene which differs from SMN1 gene in five nucleotide positions only. The c.840C > T transition at exon 7 in *SMN2* gene

produces alternative splicing and thereby while *SMN1* gene produces full-length transcripts, *SMN2* gene produces transcripts lacking exon 7 or exon 5 and exon 7 and only minor amounts of full-length transcripts. The majority of patients (95%) of SMA types 0 to IV have homozygous deletion of entire *SMN1* gene (shown as deletion of exon 7 and exon 8 in DNA tests) or exon 7 only due to gene conversion from *SMN1* to *SMN2* (gene is made thus *SMN1* exon 1 to intron 7 followed by *SMN2* exon8). The exon 8 of *SMN1* gene is noncoding. Five percent of patients of SMA may have point mutations or single-nucleotide substitutions. Diagnosis of *SMN1* deletions can be detected by restriction fragment length polymorphism test. PCR using a pair of primers is used to amplify exon 7 of both *SMN1* and *SMN2* together in single reaction; using the restriction enzyme D*ra*I that recognizes the c.840C > T change in exon 7 of *SMN2*, the PCR products are digested; and after polyacrylamide gel (PAGE) electrophoresis, 3 bands are seen a 188 base pair product corresponding to exon 7 of *SMN1* gene and 164 base pair product and 24 base pair products corresponding to cleaved amplified exon 7 of *SMN2* gene. In case the SMN1 gene 188 base pair band is missing, homozygous exon 7 deletion of *SMN1* gene is diagnosed. However, the RFLP test cannot diagnose carriers of *SMN1* gene since they have a single copy of the gene which will produce the 188 base pair products. MLPA technology, commercial kit numbers P021 and P060, can be used to diagnose both patients and carriers of *SMN1* gene-related disease. False negatives can occur in detection of carrier status in 5–8% of population which have [2 + 0] configuration. These individuals have two *SMN1* gene copies on the same chromosome and none on the other making them a carrier for disease; however, the MLPA kit will detect two copies of the *SMN1* gene giving a false-negative diagnosis. The *SMN2* copy number, which varies from zero to five, is also important to correlate with the type of SMA. The presence of two copies of *SMN2* gene is 80% predictive of SMA I phenotype, whereas presence of 4 or more copies of *SMN2* gene is 88 predictive of SMA III/IV phenotype. Nusinersen (Spinraza), a gene therapy for all types of SMA is an antisense oligonucleotide, a single-stranded RNA molecule that specifically binds to ISS-N1 regulatory motif in intron 7 in the SMN2 pre-mRNA, promoting inclusion of exon 7 and production of full-length SMN transcripts from *SMN2* gene. A phase I trial in 15 individuals with SMA type I showed that after 20 months of therapy, all patients survived and did not require permanent ventilation therapy compared to 8% of historical controls. Among the several point mutations that have been observed, the Y272C is a common missense mutation, accounting for 20% of point mutations in one study (Wirth 2000). Detection of point mutations is done by either Sanger sequencing of *SMN1* gene or by next-generation sequencing analysis using customized panels) (Lopez-Lopez et al. 2020). Recently, Chen et al. (2020) accurately identified the copy number of *SMN1* and *SMN2* genes using genome sequencing data based on read depth analysis and eight reference genomic differences between *SMN1* and *SMN2* genes. Zolgensma, approved for SMA type I, is adenoviral vector-induced delivery of *SMN1* gene to the anterior horn cells of the spinal cord and brainstem and is given as a one-time intravenous injection.

Duchenne Muscular Dystrophy (DMD)

Duchenne muscular dystrophy and Becker muscular dystrophy are two allelic disorders linked to DMD gene, chromosomal locus Xp21.2, an X-linked recessive genetic disease. DMD in males presents as delayed walking or gait disturbances around 1–2 years of age with proximal muscle weakness leading to difficulty in getting up from squatting posture with patient using his hands as support to climb up (Gower's sign), calf pseudohypertrophy, depressed deep tendon reflexes, and learning difficulties (Fig. 2). The disease is progressive with most boys wheel chair-dependent in their teens, and life beyond 25 years is uncommon. BMD in males is a milder disease with presentation as proximal muscle weakness and wasting around 30–40 years of age and life expectancy around fifth to sixth decade and death being due to cardiomyopathy. Carrier females can be rarely affected due to loss of the functional X chromosome (monosomy X/Turner syndrome), X-autosomal translocation or nonrandom X chromosome inactivation. Females can have dilated cardiomyopathy as the sole presentation. The *DMD* gene spans 2.2 megabases of DNA, comprising of 79 exons, one of the largest genes in the genome (Kunkel 2005). More than 5000 types of pathogenic variants have been identified in the *DMD* gene (Leiden open variation database). The *DMD* gene was discovered by reverse genetics or positional cloning in a girl with DMD and apparently balanced X: autosomal translocations but with small deletions of the *DMD* gene at the molecular level. Deletions of one or more genes account for 60–70% of mutations. Seventy percent occur between exon 44 and exon 53, whereas 30% of deletions occur between exon 2 and

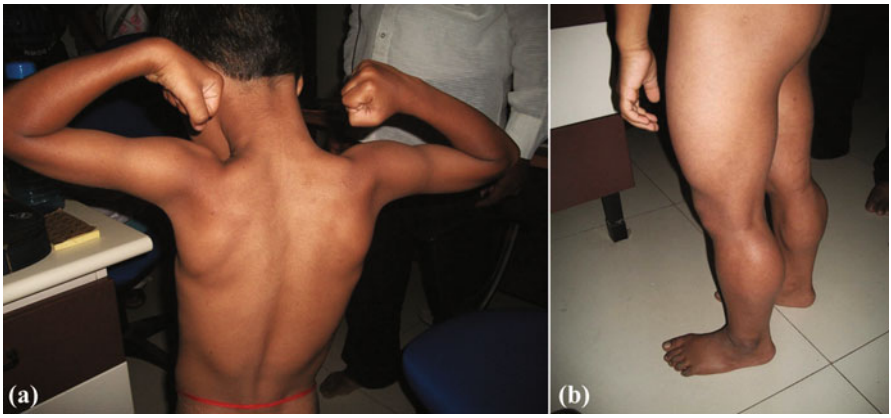


Fig. 2 (**a**) Pradhan sign: patients with Duchenne muscular dystrophy demonstrated a linear or oval depression (due to wasting) of the posterior axillary fold with hypertrophied or preserved muscles on its two borders (i.e., infraspinatus inferomedially and deltoid superolaterally), as if there were a valley between the two mounts. The sign was specific to Duchenne muscular dystrophy with sensitivity of about 90%. It was most remarkable in patients 8–11 years of age; (**b**) calf hypertrophy in the same patient: lower leg muscles, in particular the posterior compartment, paradoxically show an enlarged appearance known as pseudohypertrophy, which is characteristic however not specific of DMD

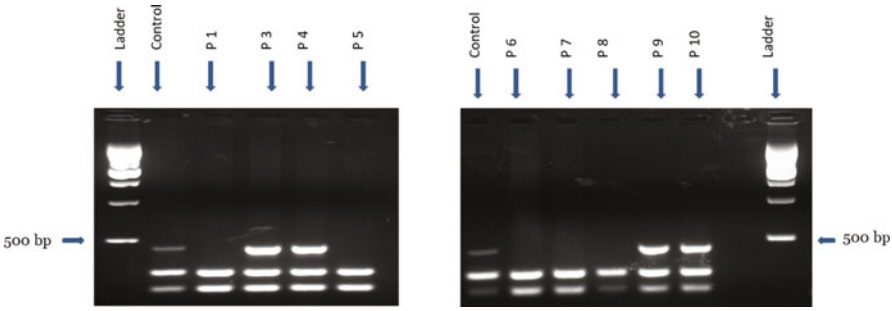


Fig. 3 Diagnosis of DMD by multiplex polymerase chain reaction. Multiplex PCR for exon 49, 41, and 42: lane 1 contains DNA sizing ladder, patient 1 (lane 3), patient 5 (lane 5), patient 6, patient 7, and patient 8 have deletion of exon 49 a common hotspot deletion

exon 20. Duplications which account for 5–10% of mutations in all occur near the 5′ end with exon 2 being the commonest. Single-nucleotide variants account for 25–35% of mutations in DMD disease and 10–20% of mutations in BMD disease. Nonsense mutations account for 20–25% in DMD and around 5% in BMD. Splice site and small indels account for similar percentage in DMD and BMD, whereas missense mutations are rare. Diagnosis of DMD was traditionally by multiplex PCR using Begg and Chamberlain primer sets. The multiplex PCR using different primers sets for different exons in the same tube is followed by gel electrophoresis to detect the multiple bands. Deletions were detected in the gel by the absence of the specific bands (Fig. 3). The Chamberlain set consists of nine primer sets: exon 45 (547 base pair), exon 48 (506 bp), exon 19 (459 bp), exon 17 (416 bp), exon 51 (388 bp), exon 8 (360 bp), exon 12 (331 bp), exon 44 (268 bp), and exon 4 (196 bp) and an extended tenth primer set exon 46 (148 bp). Similarly, there are the Beggs set (exon 1, 3, 43, 50, 13, 6, 47, 60, 52 as they appear in gel), Kunkel set (exon 49, exon 1, exon 16, exon 41, exon 32, exon 42, exon 34), the Abbs 5′ set (exon 1, exon 19, exon 3, exon 8, exon 13, exon 6, exon 4) and the Abbs 3′ set (exon 48, 44, 51, 43, 45, 50, 53, 47, 42, 60, and 52). Using these primer sets, diagnosis was possible in 50–60% of cases. In cases these deletions were negative or as an option directly as first step, MLPA commercial kit containing complementary primer sets P034 and P035 (DMD1 and DMD2) together covering all 79 exons could be used (Fig. 4). These could detect duplications in addition to deletions. Note, carrier females could only be diagnosed on MLPA. Patients who were negative on MLPA could be advised DMD gene sequencing using next-generation sequencing technology. Patients negative on MLPA and NGS test could be advised muscle biopsy to diagnose dystrophin protein deficiency demonstrated by Western blot or immunofluorescence staining techniques.

Inborn Errors of Metabolism (IEMs)

IEMs are group of disorders that occur due to deficiency of a specific enzyme that converts a specific substrate to a product leading to either accumulation of substrate

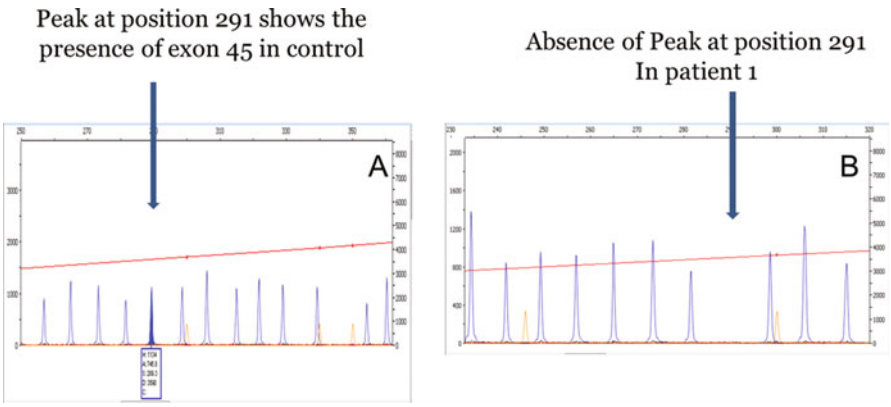


Fig. 4 Diagnosis of DMD by MLPA. Chromatogram showing the presence of deletion in DMD gene (**a**) peak showing presence of exon 45 in control. (**b**) Absence of peak indicating deletion of exon 45 in patient 1 (P1)

or deficiency of the product. The prototype disorders are storage diseases or large molecule diseases that occur due to organ/system dysfunction due to accumulation of substrates or small molecule diseases that occur due to accumulation of toxic byproducts of catabolism. Examples of storage diseases include lysosomal storage diseases a group of around 50 diseases which include mucopolysaccharidoses, lipid storage diseases, glycoproteinoses, mucolipidoses, glycogen storage disease type II, and cystinosis. Mucopolysaccharidoses are 14 types MPS I (MPSIH, Hurler syndrome; MPSIS, Scheie syndrome; MPSIH/S, mixed) to MPS IX (except MPSV, MPSVIII); MPS III is further of types A to D and MPS IV type A and B. Lipid storage diseases include sphingolipidoses (Gaucher d, Niemann-Pick type A/B), gangliosidoses (GM) type I, type II (Tay-Sachs d. (TSD) and Sandhoff d., GM2 AB variant), and leukodystrophies (metachromatic leukodystrophy, Krabbe d.). Glycoproteinoses include mucolipidoses type II and type III, aspartylglucosaminuria, fucosidosis, mannosidosis, and sialidosis. Clinical features include coarse facial features, gum hypertrophy causing gargoyle like facies (Hurler facies), enlarged spleen, liver, skeletal dysostoses (oar-shaped ribs, enlarged sella, proximal pointing of phalangeal bones, splaying/broad metaphyseal ends, hip dysplasia), corneal clouding, and mental retardation. Specific findings such as cherry red spot in fundus can occur in Tay-Sachs disease, fucosidosis, sialidosis, galactosialidosis, Niemann-Pick d., GM1 gangliosidosis, and angiokeratomas in Fabry disease. Fabry disease can lead to renal dysfunction. Diagnosis by biochemical testing is by demonstrating accumulation of substrates such as mucopolysaccharides (dermatan sulfate, heparan sulfate, keratin sulfate, chondroitin sulfate) in urine, demonstrating deficiency of specific enzyme in leucocyte enzymes such as hexosaminidase A and B in Tay-Sachs disease (deficiency of hex A) and Sandhoff disease (deficiency of both hex A and hex B). Molecular diagnosis relies on sequencing the specific gene after biochemical diagnosis or by panel sequencing (NGS). The genes vary in size,

for example, *HEXA* gene (TSD) and *HEXB* (SD) are of 14 exons each, and *GM2A* (GM2 AB variant) gene is of 4 exons. The molecular profile in Indian patients is different from those seen in other countries. In TSD, which is a classic disease seen in Ashkenazi Jews, the three mutations c.1277_1278insTATC, c.1421 + 1G > C, and c.805G > A (p.Gly269Ser) have been reported in more than 90% of cases. In Indian patients of TSD, varied mutations are observed with very few recurrent mutations such as the p.Glu462Val observed from patients from Gujarat. Seventy percent of mutations clustered within exons 5–12 (Mistri et al. 2012; Ankala et al. 2015). Regarding Gaucher disease, mutations in GBA gene c.84dupG, IVS2 + 1G > A, account for 96% percent of mutations in Ashkenazi Jews. In Indian patients, p. Leu483Pro accounted for 62% cases. Rapid diagnostic tests for lysosomal storage diseases include the Vienna StripAssay based on reverse hybridization of biotinylated PCR products. It combines probes for the eight common mutations c.84dupG, IVS2 + 1G > A, p.Asn370Ser (c.1126A > G), p.Val394Leu (c.1297G > T), p.Asp409His (c.1342G > C), p.Leu483Pro (also known as p. Leu444Pro) (c.1448 T > C), p.Arg463Cys (c.1504C > T), and p.Arg496His (c.1604G > A) and two recombinant alleles rec NciI, rec TL. These mutant probes are arranged in parallel array along with control probes on a test strip and generates test results by enzymatic color reaction easily visible to the naked eye. Results can be ready within 6 h. Prototypes of IEM with small molecular diseases are organic acidemias and aminoacidopathies. The prototype of isolated methylmalonic acidemia is caused by partial/complete deficiency of methylmalonyl-CoA mutase (*MMUT* gene), defect in transport or synthesis of its cofactor adenosyl cobalamin (*MMAA, MMAB, MMADHC*), or deficiency of the enzyme methylmalonyl-CoA epimerase (MCEE). Presentation can occur in neonatal period to adulthood depending on severity of enzyme/cofactor. Symptoms include lethargy, vomiting, ketoacidosis, respiratory distress, hyperammonemia, neutropenia, thrombocytopenia, hepatomegaly, and encephalopathy. Secondary complications can include renal dysfunction, basal ganglia injury producing choreoathetosis, immune dysfunction, and optic nerve atrophy. Biochemical diagnosis includes urine organic gas chromatography mass spectrometry which shows elevated methyl malonic acid, 3 hydroxy propionate, 2 methyl citrate, tiglylglycine, and plasma acyl carnitine analysis by tandem mass spectrometry which shows elevated propionyl carnitine, C4 dicarboxylic, or methylmalonic/succinylcarnitine (C4DC). Some patients respond to vitamin B12 treatment and some do not. Mutation analysis can be done by tiered approach. *MMUT* (60% of cases) and *MMAB* genes (12% cases) are analyzed in vitamin B12 nonresponsive individuals, *MMAA* (25% cases) analyzed in vitamin B12-responsive individuals, and lastly *MCEE* (<5% cases) and *MMADHC* (<5% cases) genes. Common mutations are reported in *MMUT* gene, p.Arg108Cys (Spanish mutation) and p.Asn219Tyr mutations are frequently associated with severe mutase deficiency, and p.Gly717Val mutation is associated with moderate type. Treatment includes management of acute sickness episodes with intravenous glucose, insulin, and high-calorie diet with restriction of propionogenic amino acids, hydroxocobalamin intramuscular injections, carnitine supplementation, and antibiotics such as neomycin/

metronidazole to reduce propionate production from gut flora. Other therapies include N-carbamylglutamate for treatment of acute hyperammonemic episodes, liver/ kidney/combined transplantations, and antioxidants for optic nerve atrophy. Another organic acidemia which can present insidiously like dyskinetic cerebral palsy is glutaric aciduria type I caused by mutation in GCDH gene encoding the glutaryl-CoA dehydrogenase enzyme. Clinically it is characterized by acute or subacute encephalopathy precipitated by intercurrent illness, dystonia, macrocephaly, and cognitive decline. GA I is associated with accumulation of upstream byproducts of L-lysine, L-hydroxylysine, and L-tryptophan degradation: glutaric acid, 3-hydroxy glutaric acid, glutaryl carnitine (C5DC acylcarnitine), and glutaconic acid. Neurological injury is characterized by cerebral edema, basal ganglia hyperintense lesions, and frontotemporal atrophy visualized as batwing appearance on CT/MRI scan of brain. Mutation analysis of GCDH gene which comprises 11 exons can be done by Sanger sequencing or by panel testing. More than 300 mutations have been described, but most of them are missense mutations (95%) which cause disturbance of the stability of the homo-tetramer structure of the GCDH protein. The common genetic variants include p.Arg402Trp (European, Russian, Egypt), p.Ala421Val (Amish community, Pennsylvania), IVS1 + 5G > T (Canadian OjiCree Indians), p.Glu365Lys (Irish travelers in Ireland), IVS10-2A > C (China and Taiwan), p.Glu181Gln (Iran and Turkey), p.Ala293Thr (South African Xhosa peoples), and p.Glu414Lys (Lumbee native Americans of North Carolina). In studies from India, around 47 unique mutations have been identified in GCDH gene with exons 9, 10, and 11 containing around 58% of all mutations. Patients with higher residual enzyme activity (30%) have low excretion of urinary glutaric acid, and sensitivity of detection on newborn screening is 85%. Patients with lower enzyme activity (less than 3%) have high urinary glutaric acid excretion, and sensitivity on detection in newborn screening is close to 100%. There is no difference in clinical presentation between the high excretors and the low excretors. Mutations associated with low excretors include p.Arg227Pro, p.Pro286Ser, p.Val400Met (Spanish), p.Met405Val (African American), and IVS1 + 5G > T (Canadian OjiCree Indians).

Genodermatoses

These are a group of genetic diseases of the skin with/without multisystem involvement. One of the commonest genetic skin diseases are tuberous sclerosis and neurofibromatosis type I. Tuberous sclerosis is a multisystem autosomal dominant genetic disease caused by mutations in either TSC1 (chromosome 9q34.13) gene encoding hamartin protein or TSC2 gene (chromosome 16p3.3) encoding tuberin protein both of which are among the key regulators of the AKT/mTOR signaling pathway. The disease involves the skin, brain, kidney, heart, and lungs. The diagnosis is achieved by clinical criteria: two major clinical features or one major and two or more minor features. The major features include hypomelanotic macules (more than three, size more than 5 mm), angiofibromas (>/= 3) or fibrous cephalic plaque,

cortical dysplasia (tubers and cerebral white matter migration lines), lymphangioleiomyomatosis, multiple retinal nodular hamartomas, renal angiomyolipoma, shagreen patch, subependymal giant cell astrocytoma, subependymal nodules, and ungual fibromas (>/=2). Minor features include confetti skin lesions, dental enamel pits, intraoral fibromas (>3), multiple renal cysts, nonrenal hamartomas, and retinal achromic patch. The diagnosis can also be achieved by identification of heterozygous pathogenic mutation in either *TSC1*$_{TSC1}$ or *TSC2* gene. Mutation analysis can be achieved by sequencing analysis on NGS panel inclusive of *TSC1* or *TSC2* gene, followed by deletion – duplication analysis by MLPA if no mutation is identified on NGS. The *TSC1* gene is approximately 50 kb in size and comprises 23 exons, whereas the *TSC2* gene is approximately 50 kb in size and comprises 42 exons. Most mutations occur in *TSC2* gene (69%) rather than *TSC1* gene (26%); 5% cases have no mutation identified. In both genes point mutations/small indels account for more than 90% of cases, whereas deletion/duplication mutations account for 5–10% cases. Cases with *TSC2* gene variants are likely to have more severe phenotype that those with *TSC1* gene variants. About two third of the cases are result of a sporadic or de novo mutation (simplex cases), whereas a third of the cases have a positive family history. Simplex cases (single occurrence in family) are more likely to have *TSC2* gene variant, while those with family history are equally likely to have either *TSC1* or *TSC2* gene variant. More than 1000 mutations have been identified in *TSC1* and *TSC2* genes, and these are loss of function type of mutations. Some *TSC2* gene variants such as p.Arg622Trp, p, Arg605Gln, p.Ser1036Pro, p.Arg1200Trp, p.Gln1503Pro, p.Gly1579Ser, and p. Arg1713His may be associated with mild phenotypes. Approximately 33% of *TSC2* gene mutations occur in exons 32–41 that encode the carboxy domain of tuberin consisting of important motifs such as GAP domain, estrogen receptor, and calmodulin-binding domains a multiple signal pathway kinase target. Missense variants at the Tyr1571 residue in *TSC2* gene are the target of tyrosine kinase. A contiguous gene deletion syndrome including *TSC2* and the adjacent *PKD1* gene has been reported with the phenotype of tuberous sclerosis and polycystic kidney disease which may be evident in utero or in infancy. Rarely, in some cases of autism, *TSC1* or *TSC2* gene variants have been identified proving them as autism susceptibility genes. Albinism is a group of diseases of melanin biosynthesis characterized by generalized reduction in pigmentation of the skin and hair and choroid part of the eye. The four types of oculocutaneous albinism are (I, II, II, IV) respectively caused by mutations in *TYR, OCA2, TYRP1*, and *MATP* gene. The clinical spectrum varies with oculocutaneous albinism type IA (OCAIA) being the most severe and vision being impaired. The differentials include Chediak-Higashi syndrome, Griscelli syndrome, Waardenburg syndrome, Hermansky-Pudlak syndrome (HPS). The diagnosis can be assisted by hair microscopy. Light microscopy of the hair in Griscelli syndrome shows small and large irregular clumping of the melanin. Melanin clumps in hair are smaller, uniform in size, and regularly arranged in Chediak-Higashi syndrome. Griscelli syndrome is of three types: type I caused by mutations in *MYO5A* (less than 5% cases), type II caused by *RAB27A* gene (more than 90% cases), and type III caused by *MLPH* gene (less than 5% cases). Type I is accompanied by primary neurological impairment but without the immunological impairment. Type II patients

are neurological impairment but with immunological dysfunction predisposing to hemophagocytic syndrome which occurs in most patients before 2–3 years of age. Treatment is possible by chemotherapy and bone marrow transplantation is curative. Mutation analysis of OCAIA can be done by hotspot analysis in Indian patients. The commonest mutation in Indian patients with OCAIA is c.832C > T or p.Arg278X initially identified in the Tili community in Bankura district of West Bengal. The other examples are with the Brahmin pedigrees from South India having c.1255G > A or p.Gly419Arg, Vysya pedigrees from South India having c.976C > T or p.Gln326X, and Garai community from East India having c.124G > A or p.Asp42Asn (Chaki et al. 2006). In those patients without the hotspot mutation, all five exons of the *TYR* gene can be sequenced by Sanger sequencing. Mutation analysis of *RAB27A* gene can also be done by hotspot analysis for c.550C > T or p.Arg184X in exon 6 (accounts for 30% cases) and if negative, sequencing all six exons of the gene. Hermansky-Pudlak syndrome is rare except in countries such as Puerto Rico. In addition to oculocutaneous albinism leading to vision impairment, these patients have accumulation of ceroid material in lysosomes, severe immunological deficiency, interstitial lung fibrosis, granulomatous colitis, and bleeding diathesis due to deficiency of dense bodies in platelets. HPS can be caused by mutations in either of the following genes: *AP3B1, AP3D1, BLOC1S3, BLOC1S5, BLOC1S6, DTNBP1, HPS1, HPS3, HPS4, HPS5*, or *HPS6*. The common mutation in Puerto Rico country is c.1470_1486dup16 in *HPS1* gene and g.339_4260del3904 variant in *HPS3* gene (3.9 kilobase deletion), c.1163 + 1G > A in Ashkenazi Jews, and c.1066_1067insG in Israeli Bedouin ancestry patients. Xeroderma pigmentosum is characterized by cutaneous and ocular sun sensitivity with skin changes such as hyperpigmentation, freckling, telangiectasis, ocular changes such as keratitis, atrophy of skin of lids, and increased risk for cutaneous neoplasms such as basal cell carcinoma, squamous cell carcinoma, and melanoma. Some patients may have neurological involvement such as microcephaly, neuropathy, hearing loss, and spasticity. The syndrome is autosomal recessive with mutations in either of the following genes: *XPA, DDB2, XPC, ERCC1, ERCC2, ERCC3, ERCC4, ERCC5*, and *POLH*. (Kraemer and DiGiovanna 2003) Founder mutations may be found in different parts of the world such as in *XPA* gene c.335_338del (India), c.390-1G > C (Japan) and p.Arg228Ter (Tunisia); *XPC* gene, c.1643_1644del (North Africa); *ERCC2* gene, p.Arg638Gln (Iraqi Jews); and *POLH* gene, del exon 10 (Tunisia), c.490G > T (Japan), and c.764 + 1G > A (North Spain/ Basque) (Tamhankar et al. 2015).

Skeletal Dysplasias

These are genetic disorders of the bone leading to either malformation or deformity of the bone, short stature, and kyphoscoliosis, with or without multiple system involvement. The manifestation could be in fetal life or after birth. One of the commonest short stature skeletal dysplasia is achondroplasia which in 99% of cases is caused by heterozygous mutation c.1138G > A or c.1138G > C (p.Gly380Arg) in exon 8 of *FGFR3* gene. The syndrome is characterized by short

limbs with large head, rhizomelic limb shortening. Presentation can be in late pregnancy on antenatal scan detection or soon after birth/infancy. Sanger sequencing for the hotspot can diagnose the disease. There are other syndromes linked to *FGFR3* gene. Hypochondroplasia is a similar milder disease caused by mutation c.1620C > A or c.1620C > G or p.Asn540Lys which occurs in 80% of cases. Craniosynostosis syndromes or syndromes that are caused by premature fusion of skull bones can be caused by *FGFR3* or *FGFR2* gene mutation. Some examples are as follows. Apert syndrome has brachyturricephaly, midface hypoplasia, and mitten hand defect (cutaneous fusion of all fingers and toes). Crouzon syndrome has craniosynostosis, hypertelorism, parrot beak nose, short upper lip, relative mandibular prognathism, and hypoplastic maxilla. Apert syndrome is caused by mutations p.Ser252Trp and p.Pro253Arg in *FGFR2* gene, whereas Crouzon syndrome is caused by mutations such as p.Tyr328Cys and p.Ser347Cys. The syndromes Crouzon syndrome with acanthosis nigricans and Pfeiffer syndrome can be caused by the mutation p.Ala391Glu in *FGFR3* gene, whereas the Muenke syndrome can be caused by p.Pro250Arg mutation in *FGFR3* gene. Another severe dysplasia is thanatophoric dysplasia which is a lethal skeletal dysplasia presenting in utero with micromelic dwarfism, severely short thorax, and protuberant abdomen; X-rays show severe platyspondyly or flat vertebrae. Thanatophoric dysplasia (TD) type I is characterized by curved short femurs or telephone receiver appearance with/without clover leaf skull, whereas TD type II has straight femur with clover leaf skull. TD type I is associated with p.Arg248Cys mutation in *FGFR3* gene, whereas TD type II is associated with p.Lys650Glu mutation in *FGFR3* gene. Then there are skeletal dysplasias that affect bone density: osteopetrosis are characterized by increased bone density or marble bone disease; osteogenesis imperfecta is characterized by diminished bone density leading to frequent fractures. Osteopetrosis can present in infancy with signs of bone marrow insufficiency due to the increased bone density, leading to pancytopenia, hepatosplenomegaly, and signs of nerve compression such as VII nerve deafness and optic nerve atrophy. Treatment is possible by bone marrow transplantation, but the neurological complications may not respond. Infantile types of osteopetrosis are predominantly caused by TCIRG1 gene mutation; other genes include *CLCN7, OSTM1, RANKL*, and *RANK* and are recessive in inheritance. Adult varieties are mostly caused by *CLCN7* gene mutation and are dominant in inheritance. Osteogenesis imperfecta (OI) is caused by decreased bone density and is of dominant and recessive types. OI type I to IV are caused by mutations in *COL1A1* or *COL1A2* genes, type V (*IFITM5* gene, autosomal dominant), type VI to type XX (are autosomal recessive except type XIX which is X-linked recessive (*MBTPS2* gene) genes for type VI to type XVIII are *SERPINF1, CRTAP, P3H1, PPIB, SERPINH1, FKBP10, SP7, BMP1, CREB3L1, SPARC* and *TENT5A*, gene for XX type is MESD. Most of the patients who are of dominant types belong to type I or type III/IV category. OI type II is the lethal intrauterine type. Mutations in *COL1A1* or *COL1A2* genes are usually substitutions of the glycine, the smallest amino acid which is present at the third position to stabilize the collagen triple helix (Gly-X-Y-Gly-X-Y…). Treatment for OI disease is possible by bisphosphonate therapy which increases the bone mineral density. Molecular

diagnosis of OI or osteopetrosis is by NGS panel testing; since very few hotspots are reported, the genes are large, and there is genetic heterogeneity. Some of the skeletal dysplasias that have been characterized in India include short rib thoracic dysplasia type III with or without polydactyl or SRTD3 (MIM#613091) also known as Verma-Naumoff syndrome for which the gene *DYNC2H1* was subsequently identified. Dr. SS Agarwal and Dr. Shubha Phadke at SGPGIMS, Lucknow, identified a skeletal dysplasia affecting several individuals in the Chanangi and Chaluvadi communities of Karnataka and named it Handigodu late onset spondyloepimetaphyseal dysplasia (SEMD), an autosomal dominant disease with variable penetrance. The gene is yet to be identified.

Endocrine Disorders

The classic genetic endocrine disorders are congenital adrenal hyperplasia, growth hormone deficiency, congenital hypothyroidism, multiple endocrine neoplasias, and hyperinsulinism syndromes. The genetics of CAH has been discussed. Isolated growth hormone deficiency can be autosomal dominant/recessive in inheritance. Type IA is a recessive disease due to severe or null mutations in *GH1* gene (chromosome 17q23.3) leading to severe deficiency of growth hormone. There is proportionate short stature present since birth or soon after birth, micropenis, and fasting hypoglycemia. Later on, there could be truncal obesity, high-pitched voice, small hands and feet, and a doll-like appearance. There is tendency to develop antibodies despite an initial good response. There could be large deletions such as whole gene, 6.7 kilobase or 7.6 kilobase, or null variants/splice variants. Large deletions can be detected by Southern blot or MLPA. Type IB is a recessive disease caused by splice mutations that can activate cryptic splicing in *GH1* gene, whereby some levels of growth hormone can be detected. There is no tendency to antibody formation to growth hormone. Examples are the mutation IVS4 + 1G > C or IVS4 + 1G > T in Saudi Arabian population. GHD type II is an autosomal dominant disease with low but detectable levels of GH, which respond well to GH replacement without development of antibodies. These patients have mutation in the first six positions of intron 3 leading to production of 17.5 kilodalton isoform that has a negative effect on secretion of growth hormone. These patients may have small pituitary on MRI and may develop combined pituitary hormone deficiency later on. These patients can be diagnosed by hotspot mutation analysis. Congenital hypothyroidism can be caused by thyroid gland dysgenesis which can be associated with variants in the following genes *TSHR, PAX8, NKX2-1, FOXE1*, and *NKX2-5*. Or the cause could be thyroid hormone synthesis defect such as those caused by variants in the following genes *SLC5A5* (iodine transport defect), *TPO, DUOX2, DUOXA2, SLC26A4* (iodine organification defect), *TG* (thyroglobulin synthesis or transport defect), or *IYD/DEHAL1* (iodotyrosine deiodinase defect). Multiple endocrine neoplasias (MEN) are autosomal dominant syndromes and are of three types. MEN type I is characterized by tumors of parathyroid glands (95% cases), endocrine gastroenteropancreatic tract (up to 80% cases), and anterior pituitary (up to 90% cases).

MEN2A includes tumors of adrenal gland (pheochromocytoma), parathyroid hyperplasia, and medullary thyroid carcinoma (which is most common initial presenting diagnosis). MEN2B includes pheochromocytoma, medullary thyroid carcinoma, Marfanoid habitus, and mucosal neuromas. MEN2A and MEN2B occur due to mutations in the *RET* proto-oncogene. Mutations affecting cysteine residues were found to lead to MEN2A, examples being p.Cys611Trp, p.Cys618Ser, p.Cys618Gly, p.Cys620Arg, p.Cys620Tyr, and p.Cys634Arg. A common recurrent mutation p.Met918Thr is found in most patients with MEN2B. Congenital hyperinsulinism is characterized by hypoglycemia neonatal/childhood onset, large baby, inappropriately elevated insulin levels, C peptide levels and inappropriately low free fatty acids, beta hydroxybutyrate, and acetoacetate. Genetic cause is identified in only 60% cases, with rest having no identified molecular basis. Mutations in *ABCC8* account for most cases (45%). In Finland, founder mutation p.Val187Asp in *ABCC8* gene renders the potassium channel nonfunctional in homozygous state/compound heterozygous state (recessive inheritance). In some patients there could be focal islet cell hyperplasia due to imprinting defect in *ABCC8* gene, an inherited mutation on the paternal allele, and a somatic mutation on the maternal allele (Knudson two hit hypothesis). Dominant missense mutations in ABCC8 gene such as p.Glu1506Lys also can lead to dominant negative effect leading to impaired opening of potassium channel.

Cardiac Disorders

The cardiac genetic disorders can be broadly classified into structural heart defects, cardiac muscle disorders or cardiomyopathies, and cardiac conduction diseases/arrhythmias. The structural heart defects can be the septal defects (ventricular or atrial), valvular defects or vascular defects such as anomalous pulmonary vein drainage, truncus arteriosus, or a combination of them. Chromosomal defects such as Down syndrome (trisomy 21) (atrioventricular septal defects), Edward syndrome (trisomy 18) (ventricular septal defect), and Turner syndrome (monosomy X) (coarctation of aorta) are known to be associated. Microdeletion syndromes such as DiGeorge syndrome (22q11.2 deletion syndrome) (conotruncal defects) and Williams syndrome (pulmonic stenosis) are also frequently associated with heart defects. Single-gene defects such as Noonan syndrome (NS) (pulmonic stenosis) which are part of the RAS-MAPK gene defects are well studied. NS is characterized by characteristic facies: low set posteriorly rotated ears, fleshy helices, hypertelorism, downslanted palpebral fissures, fullness or droopiness of eyelids (ptosis), short stature, congenital heart defect, developmental delay (25%), broad, webbed neck, pectus, cryptorchidism, coagulation defects, lymphatic dysplasia, and ocular defects. Pulmonary valve stenosis is found in 20–50% children. Hypertrophic cardiomyopathy (20–30% cases) develops in infancy or childhood. Mutation can occur in any of the following genes: *PTPN11* (50%), *SOS1* (13%), *RAF1* (5%), *RIT1* (5%), *KRAS* (5%), and other genes (*BRAF, LZTR1, MAP2K1, NRAS*) (less than 5%). *PTPN11* gene encodes a tyrosine protein phosphatase non-receptor type XI (SHP2),

and the mutations causative of NS cause catalytic gain of function and affect the interface of SH2 domain-PTP domain. Syndromes that are closely related are cardiofaciocutaneous syndrome wherein intellectual disability is more severe, there is a higher chance of brain defects, skin pathology (palmoplantar hyperkeratosis, ichthyosis) is more florid, and gastrointestinal problems are more severe. The causative genes are *BRAF* (75%), *MAP2K1* and *MAP2K2* (25%), and *KRAS* (2%). Costello syndrome facial features are more coarse (full lips, large mouth, full nasal tip), curly or sparse hair, papillomata of face and perianal region besides having brain defects such as Chiari malformation, and hydrocephalus. Most (95%) mutations affect p.Gly12 (Gly12Ala, Gly12Ser, Gly12Cys) or p.Gly13 (Gly13Cys, Gly13Asp) in *HRAS* gene. The cardiomyopathies are a group of diseases characterized by progressive weakness of the cardiac muscle. Syndromic cardiomyopathy can occur as part of Noonan syndrome, Pompe disease, Fabry disease, etc. The non-syndromic cardiomyopathy can be hypertrophic (HCM), dilated (DCM), restrictive (RCM), or arrhythmogenic right ventricular (ARVC) in nature, and these are caused by genes expressing in the cardiac muscle primarily. HCM, increased thickness of myocardium due to fibrosis and disarray of myocytes. This disease is caused by sarcomeric proteins such as *MYH7* and *MYBPC3* (both together account for 70% cases) and *TNNI3, TNNT2, TPM1* (30%), with a dominant inheritance. Mutations of MYBPC3 gene occurring in children are usually missense mutations (dominant negative effect), whereas adult cases have null variants (haploinsufficiency effect). DCM, thinned wall of the myocardium, is due to (1) mutations in Z disk proteins such as *ACTN2, CSRP3, LDB3, TCAP, DES, VCL, MYPN*, and *NEXN* leading to "loose sarcomeres" or (2) defect in energy production or regulation by affecting genes such as *FHL2, CRYAB*, or *DMD*. Mutations in LMNA and *SCN5A* gene can lead to DCM with conduction defects (Fig. 5). ARVC, involving myocyte loss and replacement by fibrofatty tissue, is due to disease of desmosomes which forms cell to cell junction and mutations occur in plakoglobin (*JUP*), desmoplakin (*DSP*), desmocollin-2 (*DSC2*), desmoglein-2 *(DSG2*), and plakophilin-2 (*PKP2*). Naxos disease, identified first in Greek islands, is a recessive syndrome caused by mutation in *JUP* gene leading to woolly hair (from birth), palmoplantar keratoderma (first few years of life), and ARVC by adolescence. Carvajal syndrome, identified typically in Ecuador, is a similar recessive syndrome caused by mutation in *DSP* gene with similar hair and skin to Naxos disease, but heart disease leads to predominant arrhythmogenic left ventricular cardiomyopathy (ALVC). RCM is characterized by increased stiffness of the ventricular chambers; thickness may be in normal limits. Mutations can be found in *TNNI3* (18%), *MYH7* (14%), *MYBPC2* (2%), and *DES* genes. Cardiac conduction disorders are broadly long QT syndrome, Brugada syndrome, catecholaminergic polymorphic ventricular tachycardia (CPVT), short QT syndrome (SQTS), idiopathic ventricular fibrillation, and progressive cardiac conduction system disease (PCCD). The heart is structurally normal. Mutations occur in genes encoding channels pertaining to transport of sodium, potassium, and calcium (channelopathies). They can present as sudden cardiac death in young, and autopsy is typically negative. The prototype disorder is long QT syndrome which is defined as corrected QT interval more than 480 milliseconds in an asymptomatic patient and

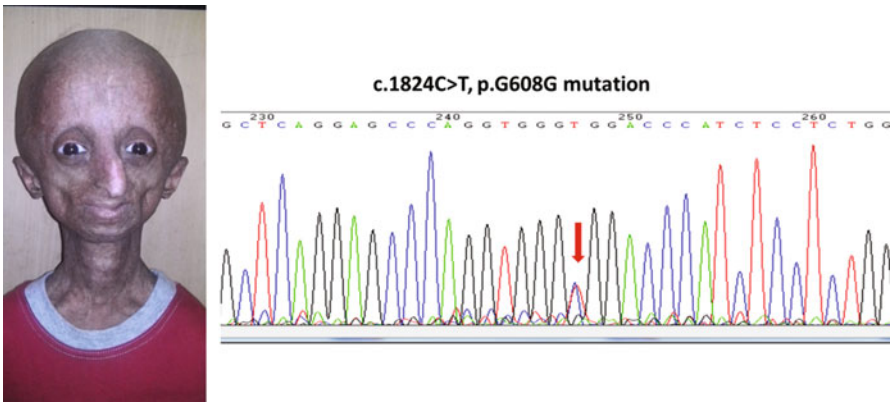


Fig. 5 Mutation analysis of LMNA gene in Hutchinson-Gilford progeria. Child with Hutchinson-Gilford progeria showing pinched nose, decreased subcutaneous fat in cheeks, small jaw, alopecia, absent eyelashes, prominent subcutaneous veins. Sequence chromatogram showing heterozygosity for c.1824C > T mutation in *LMNA* gene

more than 460 milliseconds in presence of unexplained syncope. Around 16 genes are for dominant syndromes including the following: *KCNQ1* (45%) (LQT1, HI/DN), *KCNH2* (35%) (LQT2, LOF), *SCN5A* (10%) (LQT3, GOF), *ANK2* (1%) (LQT4), *KCNE1* (LQT5), *KCNE2* (LQT6), *KCNJ2* (LQT7), *CACNA1C* (LQT8), *CAV3* (LQT9), *SCN4B* (LQT10), *AKAP9* (LQT11), *SNTA1* (LQT12), *KCNJ5* (LQT13), *CALM1* (LQT14), *CALM2* (LQT15), and *TRDN* (LQT16). The recessive syndromes Jervell and Lange-Nielsen lead to severe long QT with deafness and are caused by mutations in *KCNQ1* (JNL1) and *KCNE1* (JNL2). Prophylactic lifelong treatment with beta blockers is advised for LQT1 patients, LQT2 patients best respond to nadolol, and LQT3 patients can be treated with sodium channel blockers.

Disorders of Sex Development (DSD)

These include the anomalies in the development of internal and external genital organs. The commonest are the chromosomal disorders Turner syndrome or monosomy 45,X; Klinefelter syndrome 47,XXY; and mixed gonadal dysgenesis. Turner syndrome presents as short stature, delayed puberty, streak ovaries, and primary amenorrhea in females. Growth hormone and estrogen/progesterone hormone replacement therapy is instituted. Klinefelter syndrome presents in males as delayed puberty, decreased hair on face, body and pubic region, and small testes with azoospermia. Testosterone replacement is given from puberty, and later fertility treatments by microsurgical testicular sperm extraction and using fresh sperm for in vitro fertilization are available. Mixed gonadal dysgenesis are produced by individuals with 46,XX/XY chromosome complement. They may be true hermaphrodites with both ovaries and testes present. Then there are disorders of steroid hormone synthesis that can lead to DSD. Commonest is *CYP21A2* gene defect which

has been dealt with previously. Other steroid hormone synthesis defects include deficiency of 11-beta hydroxylase deficiency (*CYP11B1* gene), 3 beta hydroxysteroid dehydrogenase (*HSD3B2* gene), P450 oxidoreductase (*POR*), and aromatase (*CYP19A1*). Male infants with *HSD3B2* and *POR* mutations may present with undervirilization due to impaired testosterone biosynthesis. Mutations in *CYP11B1* gene lead to androgen excess causing virilization of females (even presenting with penis, scrotum with absent testes), cortisol deficiency, and corticosterone excess leading to hypertension. Founder mutations in *CYP11B1* gene include p. Gln356X (African Americans), and a hotspot is codon 318 mutations (Merke et al. 1998; Paperna et al. 2005). Mutations in *HSD3B2* gene lead to severe defect of cortisol, aldosterone, estrogen, progesterone, and androgens. Males will be undervirilized (micropenis to perineoscrotal hypospadias), and females may show normal differentiation. Founder mutation p.Ala10Glu has been identified in patients with French Canadian origin in *HSD3B2* gene.(Alos et al. 2000) Mutations in POR gene lead to congenital adrenal hyperplasia with or without Antley-Bixler syndrome (ABS). *POR* gene defect leads to affected girls born with ambiguous genitalia due to intrauterine androgen excess, and affected boys are sometimes born undermasculinized. ABS phenotype includes craniosynostosis, femoral bowing, and radiohumeral synostosis. Very low maternal serum estriol in pregnancy can be an indication of *POR* gene defect in the fetus. In 46,XY DSD, mutations occur in testicular differentiation pathway. Mutations in *SOX9* gene present with sex reversal and camptomelic dysplasia (lethal skeletal dysplasia with short, bowed long bones, club feet, Pierre Robin sequence with cleft palate). Heterozygous mutations in *WT1* gene can lead to Denys-Drash syndrome (male pseudohermaphroditism, Wilms' tumor, hypertension, degenerative renal disease), Frasier syndrome (XY male with sex reversal, streak gonads, gonadoblastoma, progressive renal disease), and Meacham syndrome (XY male, female external genitalia, retention of Mullerian structure: double vagina, absent uterus, cardiac and pulmonary malformations). Mutations in androgen receptor gene (AR) can produce sex reversal with XY males having normal female external genitalia but absence of uterus. Urogenital anomalies can be produced by mutations in *HNF1B* gene (maturity onset diabetes of young, renal cysts) and *HOXA13* gene (urogenital anomalies with limb malformations).

<!–ITerm82--?>Adult Onset/Late Onset Diseases

These include spinocerebellar ataxia, Parkinson's disease, and hereditary cancers such as breast cancer and colon cancer. The adult-onset neurological conditions spinocerebellar ataxias are typically triplet repeat anomalies. Certain regions of the genome contain repeats that are trinucleotides such as CGG, CAG, CTG, etc., tetranucleotides, and pentanucleotides. Abnormal expansion or contraction of these repeat regions in the DNA can lead to abnormal gene function and diseases. The commonest groups of disorders are hereditary ataxias. They are characterized by gait disturbance, incoordination of eye movements, and speech and hand movements with atrophy of the cerebellum. There are more than 40 types; most are dominant in

inheritance. The commonest types are SCA types 1, SCA2, SCA3, SCA6, and SCA7. Pyramidal signs (hyperreflexia and spasticity) can occur in SCA1 and SCA3; cognitive impairment can occur in SCA2, SCA12, SCA13, and SCA17; chorea can occur in SCA17 and dentatorubro-pallidoluysian atrophy (DRPLA); and Parkinsonism and motor neuron disease can occur in SCA2. The genes for SCA are SCA1 (*ATXN1* gene, normal repeat size 6–39, expanded size 39–83), SCA2 (*ATXN2* gene, normal repeat size 15–29, expanded size 34–59), SCA3 (*ATXN3* gene, normal repeat size 13–36, expanded size 55–84); SCA6 (*CACNA1A* gene, normal repeat 4–16, expanded size 21–30), SCA7 (*ATXN7* gene, normal repeat size 4–35, expanded size 34 to more than 300), and SCA17 (*TBP* gene, normal repeat size 25–44, expanded size 45–66) (Bird TD 1998). The SCA1, SCA2, SCA3, SCA6, SCA7, and SCA17 are CAG trinucleotide repeats present in coding region forming polyglutamine repeats. Fragile X syndrome occurs due to expansion of CGG repeats in the 5′ untranslated region of the *FMR1* gene on chromosome Xq27.3. Normal repeats are up to 50; carrier females have permutation range of 50–200 which can expand to their sons, while transmission to more than 200 repeats gives full-size mutation that can give rise to Fragile X mental retardation syndrome (Gold et al. 2000). Males who have permutation range CGG repeats can have Fragile X tremor ataxia syndrome characterized by adult onset (around 60 years of age) cerebellar ataxia, intention tremor, and cognitive impairment. Facioscapulohumeral dystrophy is a dominant genetic muscle disorder presenting with weakness of facial weakness, stabilizers of the scapula, and dorsiflexors of the foot. The diagnosis of FSHD type I is established by showing heterozygous contraction of the D4Z4 repeat array on subtelomeric region of chromosome 4q35. The diagnosis of FSHD type II is established by showing hypomethylation of the D4Z4 repeat array. DNA diagnostic methods for trinucleotide repeats consisted of demonstrating the size of the repeats. Southern blot was the conventional diagnostic method. Southern blot analysis for Fragile X is done by first digesting the DNA by restriction enzymes EcoRI and EagI and later blotted on to a nylon membrane followed by hybridization of radiolabeled or chemiluminescent suitable probes such as the StV12.3 probe cloned in recombinant plasmid and then visualization of the signals. However, this method is cumbersome and requires large amount of DNA. The triple primer polymerase chain reaction developed by Warner et al. uses a locus-specific primer pair (one of the primers is fluorescent) flanking the repeat region along with a third primer amplifying from the repeat region itself (Fig. 6). Following the PCR, the product is run on a capillary sequencer, and normal sized peaks are seen as single peak for each allele, and expanded repeats are observed as stuttering peak ladder pattern in the sequence chromatogram. Example is Asuragen Amplide PCR FMR1 kit.

Newborn Screening

Newborn screening is a program to screen asymptomatic newborns for common and treatable birth defects in a cost-effective way for early diagnosis and prompt institution of treatment. The common disorders that are screened in newborn period are hemoglobinopathies using high-performance liquid chromatography for

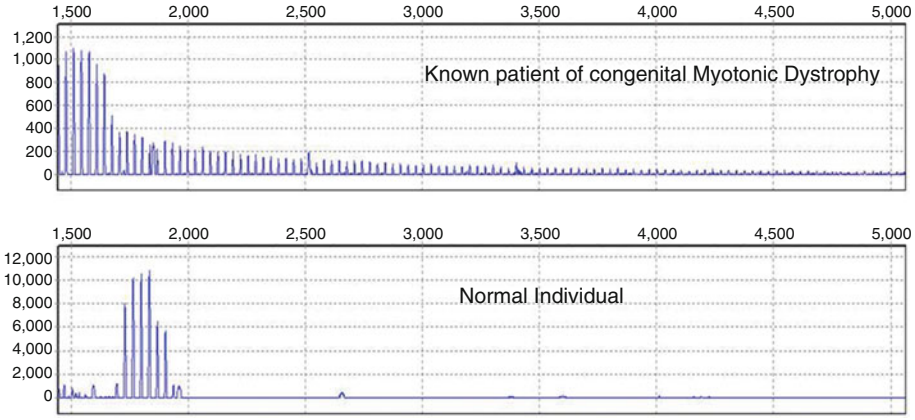


Fig. 6 Triple primer PCR in diagnosis of myotonic dystrophy. Fluorescent triple primer PCR showing presence of unstable CTG repeats more than 150 base pair in size

hemoglobin variants, organic acidemias, fatty acid oxidation defects, and aminoacidopathies using tandem mass spectrometry and gas chromatography and DELFIA (dissociation-enhanced lanthanide fluorescence immunoassay) or ELISA (enzyme-linked immunosorbent assay) for TSH (congenital hypothyroidism) or immunoreactive trypsinogen (for cystic fibrosis). DNA tests have also aided in newborn screening. In countries such as the United States, Netherlands, Germany, etc., newborn screening for spinal muscular atrophy (SMA) has been added. In a summary of SMA screening program from Australia, starting from New South Wales and Australian capital territory, 103,903 newborns were screened using first tier method of Perkin Elmer real-time PCR 4-plex assay on 96 well format and Thermostats Quant Studio DX platform. To calculate SMN2 copy numbers, digital droplet PCR was used. The genetic analysis was multiplexed with T-cell receptor excision circle (TREC) and kappa deleting recombination excision circle (KREC) as markers for primary immunodeficiencies. In those who showed two copies of SMN1 exon 7 deletion and less than four copies of SMN2 gene (called as screen positive newborns), MLPA kit P060 was used as confirmatory test (Fig. 7). In the study around ten newborns were screened positive after which nine confirmed positive. The parents were counseled that untreated presymptomatic newborns with two SMN2 copy numbers would have type I SMA (would never sit) or type II (would never walk) in approximately 82% and 17.2% cases, respectively. Untreated presymptomatic newborns with three SMN2 copies would have severe phenotypes (SMA type I and II) in 17% and 70% cases, respectively. Out of the ten newborns, nine opted for some form of disease modifying therapy (gene therapy) such as intrathecal nusinersen (Spinraza), and one family opted for supportive care only. The newborns were followed, and most patients show significant maintenance of motor function rather than the natural history of disease which includes progressive muscular deterioration and respiratory paralysis. Thus, newborn screening program

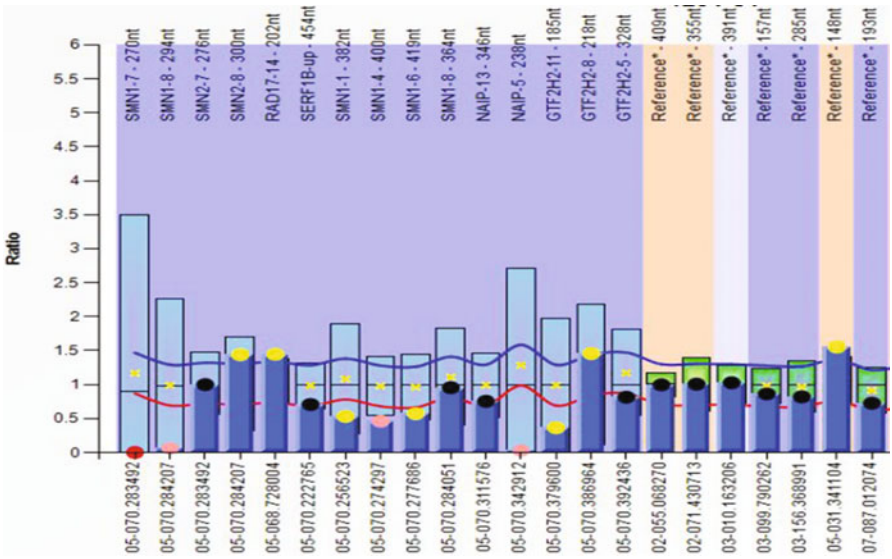


Fig. 7 Diagnosis tool of spinal muscular atrophy: MLPA. Fragment chromatogram of patient showing deletion of exon 7 and exon 8 of *SMN1* gene (log ratio = 0) and two and three copy numbers of exon 7 and 8 of *SMN2* gene (log ratio = 1 and 1.5) in patient with type II spinal muscular atrophy

for SMA in developed countries can change the prognosis and outcomes for the SMA affected children.

Premarital/Preconception/Prenatal/Preimplantation Genetic Screening/Diagnosis/Cancer Genetic Screening

As in newborn screening, screening for genetic diseases at various stages of life has been used as a voluntary tool to prevent transmission of genetic diseases to future generation or for early detection in presymptomatic period. Premarital screening, i.e., screening prospective couples for carrier status of recessive diseases, has been implemented in several countries. These countries include the Mediterranean regions such as Cyprus, Italy, Greece, and Middle East countries such as Turkey, Iran, Palestinian territories, Jordan, Saudi Arabia, Bahrain, Iraqi Kurdistan, and the United Arab Emirates. These programs were primarily aimed at screening prospective couples for beta-thalassemia minor/carrier status. In those couples being both carriers, counseling was done and countries such as Saudi Arabia were prenatal diagnosis, and therapeutic abortions are illegal, and marriages were cancelled. The Dor Yeshorim foundation screens prospective couples from Jewish background for various panels: the Ashkenazi Jewish diseases panel (Tay-Sachs disease, cystic fibrosis, Canavan disease, Niemann-Pick disease type A/B, familial dysautonomia,

Fanconi anemia type C, glycogen storage disease type Ia, mucolipidosis type IV, spinal muscular atrophy) and Sephardic/Mizrahi Jews panel (achromatopsia A and B, ataxia telangiectasia, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy), congenital adrenal hyperplasia, cystic fibrosis-Sephardic mutations, nephropathic cystinosis, Fanconi anemia type A, Friedreich ataxia, hereditary inclusion body myopathy, infantile cerebral and cerebellar atrophy, Leigh syndrome type I, megalencephalic leukoencephalopathy with subcortical cysts, metachromatic leukodystrophy, methylenetetrahydrofolate reductase, methylglutaconic aciduria type III, myoneurogastrointestinal encephalopathy, pontocerebellar hypoplasia type IIe, progressive cerebello-cerebral atrophy, Roberts syndrome, spinal muscular atrophy, Tay-Sachs disease (Sephardic mutations), and Usher syndrome type II. These are autosomal recessive diseases with high risk of mortality or morbidity wherein no treatment is available. Diseases with X-linked recessive inheritance or dominant inheritance are not screened for in premarital screening. Individuals are given an ID; once two individuals submit their ID, their compatibility check is done, and if they are not simultaneous carrier for any disease, they are informed to be compatible. This practice has almost led to elimination of genetic diseases of certain diseases such as Tay-Sachs disease from the Jewish community. In India, couples with a family history of genetic disease or consanguineous couples are increasingly being offered exome sequencing test to match themselves for genetic diseases. However, due to stigma, fear of social ostracisization, the cost of these tests, and the lack of knowledge of the mutation profile in several genetic diseases in the Indian context, these tests have not been uptaken at the mass level. Prenatal screening involves screening for the common chromosomal aneuploidies in the fetus. Maternal serum marker screening such as dual/triple/quadruple/penta marker screening during the 12–18 weeks of pregnancy has sensitivity of 60–80% in detection of Down syndrome. Recently, with the identification of cell-free fetal DNA in maternal circulation by Dennis Lo and colleagues and with the advent of next-generation sequencing technologies, it has been possible to detect chromosomal aneuploidies and microdeletion/duplication syndromes from the cell-free fetal DNA from maternal circulation. This test called as noninvasive prenatal test (NIPT) is a sensitive and accurate screening tool which is slowly replacing the maternal serum marker tests. Since, the tests reduce the need for invasive tests such as chorionic villus sampling and amniocentesis for fetal karyotyping, thereby reducing the risk of abortions due to these methods, these tests are increasingly becoming popular. Based on recent meta-analysis, trisomies 21, 18, and 13 have a false-positive rate and detection rate as follows: 0.05% and 99.5%, 0.04% and 97.7%, and 0.06% and 96.1%, respectively. For sex chromosomal aneuploidies, the detection rate was 90.3–93%, and false-positive rate was 0.14–0.23%. Since the cell-free DNA arises from the placenta, hence, in cases of placental confined mosaicism, vanishing twins, maternal chromosomal mosaic abnormalities, and maternal blood-borne cancers, there could be false positives or negatives. NIPT can also be used to screen all chromosomes for aneuploidies, but the diagnostic yield, specificity, and sensitivity are not accurately derived due to the rarity of such aneuploidies. NIPT is also used to screen the following microdeletion/duplication syndromes: Wolf-Hirschhorn syndrome

(terminal 4p deletion), Cri-du-Chat syndrome (terminal 5p deletion), Langer Giedion syndrome (8q24 deletion), Jacobsen syndrome (terminal 11q deletion), Prader-Willi and Angelman syndrome (15q11.2-q13 deletion), and DiGeorge syndrome (22q11.2 deletion). Cancer genetic screening is especially instituted for hereditary cancers such as breast cancer and colon cancer. In the breast cancer, 5–10% cases are inherited, and in ovarian cancer 10–15% cases are inherited. The most commonly mutated genes in hereditary breast cancer are the *BRCA1* and *BRCA2* genes. In those females with *BRCA1* gene mutation, the risk of breast cancer by age 70 is around 55–72%, while in those with *BRCA2* gene mutation, the risk is 45–69%. For ovarian cancer, the risks for *BRCA1* are 39–44% and for *BRCA2* are 11–17%. The lifetime risks of breast cancer and ovarian cancer in females in general population are 13% and 1.2%, respectively. Thus in those with the mutations, they can opt for more frequent screening such as breast examination, mammography, or by choosing for risk reducing surgeries such as prophylactic bilateral mastectomy and salpingo-oophorectomy. Angelina Jolie a popular Hollywood actress wrote about her experience with *BRCA1* gene mutation identification in herself and her history of her mother, grandmother, and aunt all having died of breast cancer. She underwent the risk reduction surgeries around 37–39 years of age. Following her example, there was an increased awareness about the disease in the general population which has helped the cause of preventive cancer screening in such families.

Conclusion

In conclusion, DNA diagnostic tools have revolutionized modern medicine. Rapid and cost-effective technologies have changed the way genetic diseases are being recognized and diagnosed. Doctors are adopting these tests increasingly as a genotype – first approach to comprehensively screen patients instead of multiple costly, invasive tests or tests requiring anesthesia in children such as MRI, etc. In cases without a diagnosis such as fever of unknown origin, chronic illness without a diagnosis doctors are increasingly adopting these technologies to find answers. The challenges lay in betterment of software technologies for genotype and phenotype correlation so as to minimize the uncertainty in diagnosis and availability of functional analysis of novel variants for betterment of understanding of molecular pathophysiology of these diseases.

References

Affymetrix (2017) website. Chromosome microarray. http://www.affymetrix.com/products/index.affx

Alos N, Moisan AM, Ward L, Desrochers M, Legault L, Leboeuf G, Van Vliet G, Simard J (2000) A novel A10E homozygous mutation in the HSD3B2 gene causing severe salt-wasting 3beta-hydroxysteroid dehydrogenase deficiency in 46,XX and 46,XY French-Canadians: evaluation of gonadal function after puberty. J Clin Endocrinol Metab 85(5):1968–1974

Ankala A, Tamhankar PM, Valencia CA, Rayam KK, Kumar MM, Hegde MR (2015) Clinical applications and implications of common and founder mutations in Indian subpopulations. Hum Mutat 36(1):1–10

Bird TD (1993–2021) Hereditary Ataxia overview. 1998 Oct 28 [Updated 2019 Jul 25]. In: Adam MP, Ardinger HH, Pagon RA et al (eds) GeneReviews® [Internet]. University of Washington, Seattle. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1138/

Buyse IM, McCarthy SE, Lurix P, Pace RP, Vo D, Bartlett GA, Schmitt ES, Ward PA, Oermann C, Eng CM, Roa BB (2004) Use of MALDI-TOF mass spectrometry in a 51-mutation test for cystic fibrosis: evidence that 3199del6 is a disease-causing mutation. Genet Med 6(5):426–430

Castellani C, Massie J, Sontag M, Southern KW (2016) Newborn screening for cystic fibrosis. Lancet Respir Med 4(8):653–661

Chaki M, Sengupta M, Mukhopadhyay A, Subba Rao I, Majumder PP, Das M, Samanta S, Ray K (2006) OCA1 in different ethnic groups of India is primarily due to founder mutations in the tyrosinase gene. Ann Hum Genet 70(Pt 5):623–630

Chen X, Sanchis-Juan A, French CE, Connell AJ, Delon I, Kingsbury Z, Eberle MA (2020) Spinal muscular atrophy diagnosis and carrier screening from genome sequencing data. Genetics in Medicine 22(5):945–953

Dahm R (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Human Genetics 122(6):565–81

Devyser website (2021) Principles of QF PCR. https://devyser.com/

Eaton WA (2020) Hemoglobin S polymerization and sickle cell disease: a retrospective on the occasion of the 70th anniversary of Pauling's science paper. Am J Hematol 95(2):205–211

El-Fadaly N, Abd-Elhameed A, Abd-Elbar E, El-Shanshory M (2016) Accuracy of reverse dot-blot PCR in detection of different β-globin gene mutations. Indian J Hematol Blood Transfus 32(2): 239–243

Farrell PM, White TB, Ren CL, Hempstead SE, Accurso F, Derichs N, Howenstine M, McColley SA, Rock M, Rosenfeld M, Sermet-Gaudelus I, Southern KW, Marshall BC, Sosnay PR (2017) Diagnosis of cystic fibrosis: consensus guidelines from the Cystic Fibrosis Foundation. J Pediatr 181S:S4–S15.e1

Gold B, Radu D, Balanko A, Chiang CS (2000) Diagnosis of fragile X syndrome by Southern blot hybridization using a chemiluminescent probe: a laboratory protocol. Mol Diagn 5(3):169–178

Harteveld CL, Kriek M, Bijlsma EK, Erjavec Z, Balak D, Phylipsen M, Voskamp A, di Capua E, White SJ, Giordano PC (2007) Refinement of the genetic cause of ATR-16. Hum Genet 122 (3–4):283–292

Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC (1989) Identification of the cystic fibrosis gene: genetic analysis. Science 245 (4922):1073–80

Kraemer KH, DiGiovanna JJ (1993–2021) Xeroderma Pigmentosum. 2003 Jun 20 [Updated 2016 Sep 29]. In: Adam MP, Ardinger HH, Pagon RA et al (eds) GeneReviews® [Internet]. University of Washington, Seattle. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1397/

Kunkel LM (2005) 2004 William Allan award address. Cloning of the DMD gene. Am J Hum Genet 76(2):205–214

Lejeune J, Gauthier M, Turpin R (1959) Les chromosomes humains en culture de tissus [The human chromosomes in tissue culture]. Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. 248:602–603

Lopez-Lopez D, Loucera C, Carmona R, Aquino V, Salgado J, Pasalodos S, Miranda M, Alonso Á, Dopazo J (2020) SMN1 copy-number and sequence variant analysis from next-generation sequencing data. Hum Mutat 41(12):2073–2077

Mendonça RH, Matsui C Jr, Polido GJ, et al (2020) Intragenic variants in the SMN1 gene determine the clinical phenotype in 5q spinal muscular atrophy. Neurol Genet 6(5):e505. Published 2020 Sep 1

Merke DP, Tajima T, Chhabra A, Barnes K, Mancilla E, Baron J, Cutler GB Jr (1998) Novel CYP11B1 mutations in congenital adrenal hyperplasia due to steroid 11 beta-hydroxylase deficiency. J Clin Endocrinol Metab 83(1):270–273

Mistri M, Tamhankar PM, Sheth F, Sanghavi D, Kondurkar P, Patil S, Idicula-Thomas S, Gupta S, Sheth J (2012) Identification of novel mutations in HEXA gene in children affected with Tay Sachs disease from India. PLoS One 7(6):e39122

Online Mendelian (2021) Inheritance in Man. www.omim.org

Orkin SH, Kazazian HHJ, Antonarakis SE, Goff SC, Boehm CD, Sexton JP, Waber PG, Giardina PJV (1982) Linkage of β-thalassaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. Nature 296:627–631

Paperna T, Gershoni-Baruch R, Badarneh K, Kasinetz L, Hochberg Z (2005) Mutations in CYP11B1 and congenital adrenal hyperplasia in Moroccan Jews. J Clin Endocrinol Metab 90 (9):5463–5465

Paracchini V, Seia M, Raimondi S, Costantino L, Capasso P, Porcaro L, Colombo C, Coviello DA, Mariani T, Manzoni E, Sangiovanni M, Corbetta C (2012) Cystic fibrosis newborn screening: distribution of blood immunoreactive trypsinogen concentrations in hypertrypsinemic neonates. JIMD Rep 4:17–23

Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, et al (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245(4922):1059–65.

Tamhankar Parag M et al (2020) Clinical characteristics, molecular profile, and outcomes in Indian patients with glutaric aciduria type 1. J Pediatr Genet. https://doi.org/10.1055/s-0040-1715528

Tamhankar PM, Agarwal S, Arya V, Kumar R, Gupta UR, Agarwal SS (2009) Prevention of homozygous beta thalassemia by premarital screening and prenatal diagnosis in India. Prenat Diagn 29(1):83–88

Tamhankar PM, Iyer SV, Ravindran S, Gupta N, Kabra M, Nayak C, Kura M, Sanghavi S, Joshi R, Chennuri VS, Khopkar U (2015) Clinical profile and mutation analysis of xeroderma pigmentosum in Indian patients. Indian J Dermatol Venereol Leprol 81(1):16–22

Tjio JH, Levan A (1956) The chromosome number of man. Hereditas 42:1–6

Welsh MJ, Smith AE (1993) Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. Cell 73(7):1251–1254

Wirth B (2000) An update of the mutation spectrum of the survival motor neuron gene (SMN1) in autosomal recessive spinal muscular atrophy (SMA). Hum Mutat 15(3):228–37

Part V

Modifications in Routine Methodologies to Solve Challenging Cases

Challenges in DNA Extraction from Forensic Samples

32

Anna Barbaro

Contents

Abstract

DNA extraction from forensic samples is the most critical step of all the analytical process since it impacts on the quantity and quality of DNA obtained from a sample: good DNA produces good interpretable profiles.

Forensic evidences are highly variable and unpredictable; thus the choice of the DNA extraction procedure depends on some factors such as the type of the biological sample, the wide ranges of substrates where the sample may be placed, the amount of the biological evidence being examined, exposure to varying

A. Barbaro (✉)
Department of Forensic Genetics, Studio Indagini Mediche E Forensi (SIMEF), President Worldwide Association of Women Forensic Experts (WAWFE), Reggio Calabria, Italy

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_36

environmental conditions, variable degradation, and the presence of PCR inhibitors.

The use of an appropriate method able to optimize DNA recovery from forensic samples is therefore essential in order to obtain high DNA quantity free of inhibitors useful for downstream applications.

Several DNA extraction procedures (manual and automatic) have been developed during years in order to improve DNA extraction especially from challenging samples.

Keywords

DNA extraction · Challenging samples · Inhibition · Degradation · Contamination

Introduction

DNA is the hereditary material of humans and other organisms. Most DNA is contained in the nucleus of the cells, and it is named "nuclear DNA" Nonetheless, a small amount of DNA is placed inside the mitochondria, the organelles in cellular cytoplasm responsible for the energy production. It is called "mitochondrial DNA" or mtDNA. Nuclear DNA is tightly packed into chromosomes, which consist of long chains of DNA and proteins (called histones) which, together with chaperone proteins, condense the DNA molecule to maintain its integrity. In humans, somatic cells are diploid, and they normally contain 23 pairs of chromosomes (46 in total). 22 chromosomes (called autosomes) are the same in male and females, while one pair (sex chromosomes) differ between them. Males have XY chromosomes, while females have XX.

On the contrary, gametes are haploid cells which contain 23 chromosomes.

The amount of DNA per cell is about 6 pg. Except in case of pathologies. DNA profile is the same in every cell of the body. This means any biological sample may be useful for DNA typing. The term DNA extraction refers to DNA isolation from biological samples. DNA may be extracted from a wide range of samples belonging to a person or found at the crime scene. Most common evidence consists in biological fluids (e.g., blood, saliva), bloodstains, oral swabs, hairs, human remains (bones, teeth, cadaveric tissues), and sometimes contact traces.

For the success of DNA typing, it is necessary a good preparation of genomic DNA.

DNA Extraction History

The Swiss physician and biologist Johannes Friedrich Miescher is reported as the first researcher able to extract in 1869 nucleic acids, opening the way to DNA identification.

At that time, he was working on a project aimed to determine the chemical composition and the protein components of cells, using leucocytes, which are the white blood cells.

He observed that the addition of an acid solution to leukocytes produced the precipitation of a substance, which disappeared after the addition of an alkaline solution. He observed that the chemical composition of the substance was different from that of proteins and of other known molecules. He named it *nuclein* (corresponding now to DNA) because he supposed it was contained in the nuclei of the cells. Accidentally, he had developed a method to isolate cellular nuclei from cytoplasm to obtaining DNA (Dahm 2004, 2008).

After this discovery, Miescher developed an extraction method from salmon sperm in order to obtain larger quantities of purified nuclein, which was later named "nucleic acid" by his student, Richard Altman.

Unfortunately at his time, the relevance of his discovery was not appreciated, and his name fall into obscurity during 50 years before his research was recognized by the international scientific community.

In the decades following Miescher's discovery and by the twentieth century, other scientists continued to investigate the chemical nature of the "nuclein." For example, the Russian biochemist Phoebus Leven and the Austrian biochemist Erwin Chargaff carried out a series of research and discovered the primary chemical DNA components (such as the three major components (phosphate-sugar-base) of a single nucleotide, the carbohydrate component (deoxyribose) of DNA, and the ways in which they joined with one another.

These pioneer studies opened the way to Watson and Crick discovery in 1953 about the three-dimensional double helix structure of the DNA molecule (Pray 2008; Tan 2009).

Forensic Samples

Extraction of DNA from a forensic sample is the most critical step of all the analytical process since it impacts on the quantity and quality of DNA obtained from a sample: good DNA produces good interpretable profiles.

Forensic evidences are highly variable and unpredictable; thus the choice of the DNA extraction procedure depends on some factors such as the type of the biological sample, the wide ranges of substrates where the sample may be placed, the amount of the biological evidence being examined, exposure to varying environmental conditions, variable degradation, and the presence of PCR inhibitors.

The intrinsic characteristics of forensic samples mainly consist in the type of biological fluids (e.g., blood, saliva, semen), nature of the tissues (i.e., bones, teeth, hairs), sample quantity (trace or large quantities), and source (i.e., evidence or reference).

Evidence extrinsic characteristics are mainly related to the type of substrate where the sample is placed (e.g., cotton, denim, swabs) and the type of the crime (e.g., homicide, sexual assault), from which they are derived (Lee et al. 2017).

Table 1 Example of DNA amounts extracted from biological samples (according to Butler 2012)

Sample	Amount of DNA (average)
Liquid blood	20,000–40,000 ng/ml
Liquid semen	150,000–300,000 ng/ml
Liquid saliva	1000–10,000 ng/ml
Urine	1–20 ng/ml
Bloodstain	250–500 ng/cm^2
Plucked hair with root	1–750 ng/root
Shed hair with root	1–10 ng/root
Oral swab	100–1500 ng/swab
Postcoital vaginal swab	10–3000 ng/swab
Bone	3–10 ng/mg
Tissue	50–500 ng/mg

Because of this, forensic evidence may exhibit unlimited combinations of sample and substrate types, sample and environmental conditions sample quantity and quality, sample and contaminants/inhibitors, etc.

Some elements are visible and they may be observed or detected, whereas many others unfortunately remain unknown to the forensic examiner.

This renders the choice of the DNA extraction method the most critical step in forensic DNA typing.

The use of an appropriate method able to optimize DNA recovery from forensic samples is therefore essential in order to obtain high DNA quantity free of inhibitors useful for downstream applications (e.g., PCR, sequencing, micro-array analysis).

For example, the procedure used for processing a blood sample is different from that used for DNA extraction from a bone or a hair.

Obviously, the quantity and the quality of DNA recovered from forensic samples is variable depending on the type of material (Table 1).

DNA Extraction Methodologies

DNA extraction from an evidence sample is the most critical step in all processing since it influences the ability to obtain interpretable genotype.

In addition, loss of DNA may occur during this step due to the substrate sustaining the sample.

A biological sample contains a number of substances besides DNA: thus, DNA extraction consists in the purification of the genetic material from all those superfluous substances contained in the sample that could interfere with the subsequent processing steps.

These substances include proteins and cellular components that package and protect DNA, RNA, enzymes, organic residues, salts, etc.

To be effective, a DNA extraction procedure should (Lee et al. 2017).

- Allow acid nucleic extraction from a wide range of biological samples
- Allow high DNA recovery in sufficient quantity and quality for downstream analysis
- Enable DNA isolation from trace samples
- Reduce DNA loss of available
- Remove inhibitors with PCR
- Not introduce any contaminants
- Not alter/degrade DNA
- Be simple and rapid
- Be amenable for automation
- Not use hazardous reagents

DNA extraction procedure generally involves the following steps:

- Cellular lysis (by physical, chemical, or enzymatic methods) in order to release DNA
- Purification (by organic extraction, precipitation, or binding) in order to remove proteins from other cellular components and any contaminants
- Elution of DNA suitable for downstream applications

Organic Extraction

Organic extraction, often called phenol chloroform extraction, during many years was the most widely used method in DNA laboratories (Butler 2012). This procedure is complex; it involves the use of several chemicals and requires many steps.

Firstly, the sample is incubated in a lysis solution, containing sodium dodecylsulfate (SDS) and proteinase K, useful to break the cell membranes and proteins that package DNA in the chromosomes.

Then a phenol-chloroform-isoamyl alcohol mixture (25:24:1) is added to denature proteins and to facilitate the separation of the organic and aqueous phase, where DNA become concentrated.

This is because phenol denatures proteins rapidly, but it does not completely inhibit RNAse activity. In addition, the use of isoamyl alcohol reduces foaming during extraction.

When the mixture phenol-chloroform-isoamyl alcohol is added, a biphasic emulsion is formed: by centrifugation of the sample, the hydrophilic layer stands on the top of the emulsion, while the hydrophobic layer remains on the bottom.

DNA contained in the upper phase is then precipitated by adding 100% ethanol, and the pellet is washed with 70% ethanol to remove residual salts and small organic molecules.

Extracted DNA is then dissolved in distilled water or into T.E. buffer at a concentration suitable for further analysis (Sambrook et al. 1982).

Some posterior protocols involved a concentration step in place of the ethanol precipitation by special column membrane (e.g., Centricon 100 or Microcon

100, Millipore) to remove inhibitors (Comey et al. 1994). This method is low cost and highly efficient since it allows the recovery of high molecular weight DNA. Anyway, it is time-consuming; it requires the use of several hazardous chemicals (Phenol is a toxic, corrosive, and flammable carbolic acid) and many handling steps that increase the risk of error, contamination, and sample loss, especially when working with trace samples.

Because of this, even if it was the most popular method during many years, it was replaced by other methods more suitable for forensic laboratories.

Chelex Extraction Method

This method was introduced in 1991, and it rapidly became popular among forensic laboratories in replacement of the organic extraction. Chelex is a reagent composed of styrene divinylbenzene copolymers, and it contains paired iminodiacetate ions that act which as chelators for polyvalent metal ions. The extraction method consists in adding an aliquot of the Chelex resin to the sample and after an incubation step, in boiling it for several minutes in order to break cells and release DNA (Walsh et al. 1991). Chelex resin binds cellular components and contaminants, while nucleic acids remain in the aqueous solution above the resin. During boiling, Chelex acts as chelating polyvalent metal ions (e.g., magnesium) that catalyze DNA degradation in low ionic strength solutions.

The procedure is simple and rapid; it does not require organic solvents or multiple tube transfers, reducing the risk of contamination. Furthermore, it is cheaper than other DNA extraction methods.

Anyway, the method may be not enough to remove inhibitors (dye, hematin) especially when working with large bloodstains or colored substrates (e.g., denim), due to the lack of a purification step.

In addition, Chelex procedure cannot be used to extract DNA for restriction fragment length polymorphism (RFLP) analyses because it produces extracted DNA at single-strand due to the sample exposure to the high temperature during boiling and to the alkalinity of the method.

Chelex is commercially available as a powder or as a ready-to-use reagent in several formulations (Fig. 1).

Magnetic Beads Extraction Method

The interest for the use of a magnetic separation method to purify biological material (nucleic acids, proteins, etc.) greatly has increased in the last decades. The first method for DNA purification and isolation using magnetic particles was reported and patented in the USA in 1998 (Hawkins 1998). Several simple and reliable procedures have been developed during years. Silica matrices purification is based on the high affinity of DNA backbone (negatively charged) toward the silica particles (positively charged).

Fig. 1 Example of Chelex commercial formulation (Author's property)

The method in fact involves the use of polymer-coated magnetic beads or particles functionalized with silica surfaces that under optimized conditions (e.g., presence of chaotropic salts) bind DNA (negative charged) while contaminants remain in solution. Salt concentration influences the strength of the bonds allowing reversible binding between magnetic particles and DNA (Berensmeier et al. 2006).

When an external magnetic field is applied and maintained, particles bound with DNA are attracted to it, and they precipitate down at the edge of the tube. This allows to discard the liquid phase containing contaminants. Several washing steps are then performed to purify nucleic acid in the magnetic particle pellet. DNA is released from the magnetic particles using an elution buffer (Table 2). This method allows rapid extraction and purification of high-quality DNA from a variety of biological samples. The amount of material genetic extract depends on the number and capacity of the magnetic particles.

Magnetic particles with a large surface area are generally preferred for binding nucleic acids.

Table 2 DNA extraction method using magnetic beads (Author's Property)

Step 1

Lyse sample and releaseDNA

(Remove any substrate)

Step 2

Bind DNA to Magnetic Beads

(Apply a magnetic field to collect beads)

Step 3

Wash Beads and discard the liquid phase containing contaminants

Step 4

Elute DNA from magnetic beads

Step 5

Recover Purified DNA

The development of magnetic beads technology for solid-phase-based extraction greatly improved DNA extraction. The main advantage consists in the method simplicity and in the ability of the magnetic separation of the particles to isolate DNA and to remove debris and contaminants.

The major disadvantage of the method is that contaminants (if in excess) may sterically prevent the DNA adsorption on to the beads. In addition, the silica method preferentially recovers high-molecular-weight DNA, and so it may fail to recover highly fragmented molecules (e.g., $<$100 bp) (Lee et al. 2017).

During years, different modified protocols based on magnetic-beads extraction have been developed in order to improve DNA extraction especially from challenging samples such as hairs, human remains, embedded tissues, and adhesive surfaces (Barbaro 2009, 2011b).

Most widely diffused commercial systems include DNA IQ (Promega), PrepFiler (Applied Biosystems), and QIAmp (Qiagen).

Spin Columns Extraction Method

This procedure relies on the DNA binding to a solid phase which is placed inside the column tube. When the lysed sample is added to the column, nucleic acids bind to the silica gel or silica beads in the spin columns under selective pH and salt conditions created by the chaotropic salts which are the major component of the spin column buffers. The presence of high concentrations of chaotrophic salts (e.g., guanidine isothiocyanate, guanidine hydrochloride, sodium perchlorate) is critical since it allows disruption of the hydrogen bonding between strands and it favors binding of around 90–95% DNA to silica surface (Duncan et al. 2003).

The addition of isopropanol will further enhance and influence the binding of nucleic acids to the silica.

Proteins and other contaminants are eliminated by repeated washing of the silica membrane followed by centrifugation.

Generally, the first wash is made with a buffer at low concentration of chaotropic salts to remove residual proteins and pigments, while the second wash step removes any trace of chaotropic salt, and it is followed by a further wash with ethanol. In this way, around 95% of all the other molecules are filtered away through the column, allowing a high contaminants removal.

DNA is then separated from the silica matrices by elution in low-salt buffer or distilled water. When the elution buffer is added to the spin column, DNA becomes rehydrated, and it dissolves into the buffer (Table 3).

The method is relatively easy and rapid even if it requires several steps and the use of buffer solutions and solvents (Greenspoon et al. 1998).

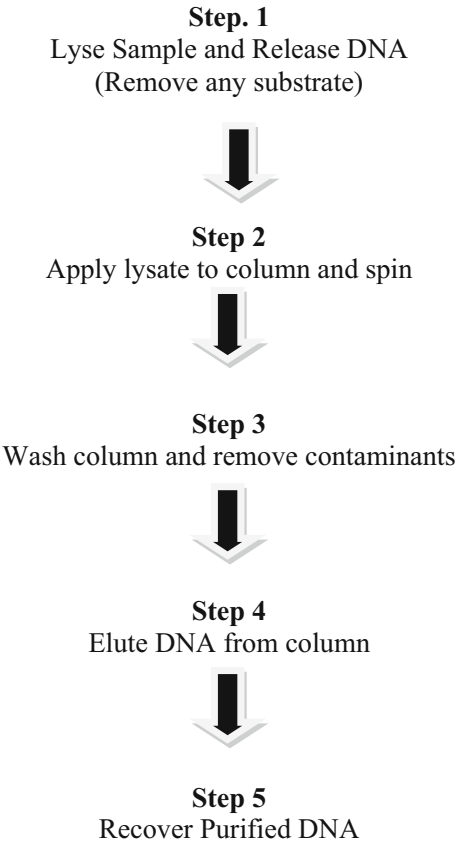
A disadvantage may consist in a low DNA recovery due to the presence of silica-binding inhibitors that may compete with DNA binding during processing.

Depending on the sample (liquid fluids, stains, swabs, etc.), different extraction protocols are available to improve DNA recovery. Most widely diffused commercial systems include Nucleo-Spin (Macherey-Nagel) and QIAmp DNA kit (Qiagen).

Automated Extraction

In the last years, several robotic workstations have been developed that allow automated DNA extraction enabling high-throughput sample processing and efficient and reproducible recovery of DNA (Nagy et al. 2005; Montpetit et al. 2005;

Table 3 DNA extraction method using spin column method (Author's Property)



Witt et al. 2012). Automated procedures not only reduce labor time and cost but also minimize contamination issues eliminating human error in reagent/sample preparation and handling. All reagents are included in special prefilled, sealed cartridges that are placed on the instruments. After an initial sample pre-lysis, subsequently all purification and elution steps are performed automatically reducing the risk of contamination. Either resin or spin column method may be automatizable.

Automated instruments were initially expensive, complex, and large, but during years cheaper, compact, and easy-to-use instruments have been produced by various companies, with different properties and ability to process a variable number of samples.

The mechanism, the procedure, and use modality are almost similar (Fig. 2).

Most diffused biorobots are EZ1 XL (Qiagen), QIAcube (Qiagen), and Maxwell 16 (Promega). Automate Express (Life Technologies). Depending on the platform used, different DNA extraction kits based on resin or column extraction are

Fig. 2 Example of automated robotic workstation (Author's Property)

commercially available together with several protocols for different types of biological samples (Kishore et al. 2006; Liu et al. 2012; Pajnič et al. 2016; Barbaro et al. 2019).

Main advantages of automated DNA extraction may be resumed in the following points:

- Reduced contamination risk
- Minimized manual errors
- High-throughput
- High reproducibility
- Consistent performance

– Reduce processing time
– Workflow integration
– Compatibility with Laboratory Information Management System (LIMS)

Differential Extraction

Sexual assault samples often contain a mixture of sperm cells from the perpetrator in an excess of female epithelial cells of the victim, resulting in an unbalanced ratio between components.

In these cases, it may be useful to perform a differential extraction that allows selective isolation of male and female DNA from the mixed samples: the separation of the two fractions greatly helps in the deconvolution of the donors. This is obviously because the genotype of the perprator is of primary importance in resolving the case.

This procedure was described for the first time in 1985 (Gill et al. 1985). Several protocols are available, and they are based on the preferential lysis of the female epithelial cells using a solution containing sodium dodecyl sulfate (SDS) and proteinase K.

Sperm cells remain intact during this step, and they are subsequently lysed by the addiction of dithiothreitol (DTT). This latter acts by breaking the protein disulfide bridges of the sperm nuclear membranes. Although useful, unfortunately differential extraction may result in loss of some male DNA, rising 90% according to some authors (Vuichard et al. 2011). This is obviously a relevant problem when analyzing evidences containing little amounts of male DNA, such as the ones collected several hours or days after the event. In case of azoospermic individual, due to the absence of sperm cells it is not possible to perform a differential extraction. Hence, it is recommended to analyze directly Y chromosome markers that allow to identify the male DNA profile even in a mixture containing high quantity of female DNA (Sibille et al. 2002).

Other diffused strategies in processing sexual assault samples consist in separate epithelial cells by different techniques such as flow cytometry (Schoell et al. 1999), laser microdissection (Elliott et al. 2003), and microfluidics (Bienvenue et al. 2010).

Direct PCR

In the last years, next-generation PCR amplification kit has been developed containing inhibitor-resistant PCR reagents which enable direct amplification without DNA extraction. An aliquot or a portion of the biological sample or a punch of an FTA card is added directly to the PCR reaction mixture without submitting the sample to prior DNA extraction, purification, or quantification (Wang et al. 2009; Barbaro 2011a). This single-step method has the great advantage to avoid loss of DNA material during extraction steps and reduce the risk of contamination, which is very useful especially when working with small traces.

This method has also the benefit for forensic laboratories of reducing cost and saving time: because of this, it has become a popular procedure for processing reference samples (oral swabs, bloodstains) more rapidly. Several kits are commercially available either in the regular formulation than in the one for direct PCR amplification (Identifiler Direct (Life Technologies), Global Filer Express (Life Technologies), PowerPlex 16 HS (Promega), PowerPlex21 (Promega), Investigator 24plex GO (Qiagen), etc.

Storing Condition

Storage conditions may be variable depending on the type of DNA sample, the duration of storage, and the downstream applications.

Biological material to be used for DNA typing should be adequately stored in a manner to preserve its integrity (NIST, NIJ 2013). Best practices include freezer storage at −20 °C or −80 °C, which should prevent the nucleases activity, avoiding DNA degradation. Nucleases are enzymes that degrade DNA allowing the cells to recycle the nucleotide components. They require magnesium to work properly. A method for preventing their activity in blood samples consists in collecting blood in tubes containing EDTA that acts by chelating the free magnesium.

Also drying conditions preserve DNA and prevent base hydrolysis by DNase. A simple common method for storing blood or saliva samples consists in the use of cellulose-based filter paper (Coble et al. 2008). Some cards are commercially available (e.g., FTA (Whatman), Nucleicard (Copan), etc.) for the collection of biological material (Fig. 3). These cards are treated with chemicals that inactivate bacteria and preserve nucleic acid from degradation. When the biological material is deposited onto the card, the color indicator changes (e.g., pink to white). A punch of this card area is suitable for direct amplification. In the last years for the collection of saliva samples (especially for databases purposes), special devices consisting in a spongy swab that is then pressed against the card to transfer the sample have been developed. This allows sample better preservation. When the collected material is transferred to the card, a change in the indicator color occurs confirming the sample has been transferred to the card. Cards are left air-drying and then stored at room temperature.

A correct post-extraction storage is essential to the maintenance of high-quality DNA to be used for downstream applications. DNA degrades at room temperature, generally frozen better preserve DNA integrity. Extracted DNA is typically stored at 4 °C during laboratory activity and at −20 °C, or 80 °C for long-term storage, to prevent nuclease activity (Butler 2012). Nonetheless, since multiple freeze-thaw cycles can degrade DNA, thus depending on the DNA volume available, when performing multiple analysis, it is recommended to split the DNA sample in aliquots to be thawed only when necessary. For DNA elution, it is suggested to use a solution containing T.E. (Tris-EDTA) that avoids acidification (which leads to depurination) and prevents DNases activity by chelating metal ions.

Fig. 3 Example of cards for biological samples preservation (Author's property)

Main Challenges

Challenges in analyzing casework samples are due to the wide range of biological samples, high degree of heterogeneity of the substrates where samples are placed, great variation in the amount of available sample (often present in limited quantity), exposure to variable environmental conditions (heat, humidity) that favorite bacterial growth and DNA degradation, sample contamination, and the presence of PCR inhibitors (Lee et al. 2017).

DNA Quantity and Quality

A very challenging factor in forensic DNA typing is the wide variability of forensic evidences. DNA in forensic samples is often in limited quantity or shows high degradation due to the several factors.

After cellular death, endonucleases start to break down the bonds between the nucleotides; the digestion of chromatin by lysosomal proteases favorite endonuclease activity.

The enzymatic reaction is followed by other persistent spontaneous and random DNA degradation processes. In addition, exposition to harsh environmental factors

(e.g., UV, humidity, heat, water) or to bacterial, fungal, and insect actions speeds the degradation.

The glycosidic base sugar bond is the most preferred cleavage target: the breakage leads to bases loss that result in DNA fragmentation.

Due to the small fragments (average range 80–200 bp) produced often, PCR primers may fail in annealing with the target regions.

Furthermore, exposition to UV irradiation (e.g., direct sunlight) produces crosslinking between adjacent thymine nucleotides which prevent the passage of DNA polymerase during PCR (Butler 2012).

Inadequate DNA quantity or quality may result in partial DNA profiles that unfortunately cannot provide the same discrimination power than full profiles, since often they do not give enough information able to include or exclude an individual as a sample donor.

Because of this, the use of correct sample collection procedures and adequate storage conditions together with an efficient extraction method strongly influences the quality and quantity of DNA that might be recovered from evidences and consequently the success of the analysis.

Human remains, formalin-fixed and paraffin-embedded tissue, are among the most challenging samples.

Bones and teeth often represent the only sources of DNA in case of mass disaster, missing persons identification, and other compromised scenarios.

DNA in bones and teeth is better preserved than in soft tissues due to the presence of hard connective tissue containing high levels of calcium. The choice of an efficient DNA extraction procedure is crucial to avoid the sampling of minerals and to remove polymerase chain reaction (PCR) inhibitors (Barbaro et al. 2021).

Because of this, some protocols have been developed during years in order to increase the quantity and quality of DNA extracted from human remains (Stray et al. 2010; Lee et al. 2010; Barbaro 2011; Builes et al. 2013).

Formalin-fixed or paraffin-embedded tissue sections are a valuable tool not only for medical studies (e.g., to study diseases such as cancer) but also for forensic application especially retrospective investigations were suitable biological samples may be limited.

In fact, they often represent the unique samples available in situations where individuals are no longer alive or no other biological samples are available for comparison (e.g., paternity testing, missing persons identification, criminal caseworks, evaluation of mismatched specimens).

Fixation prevents the cellular changes, but it produces an extensive modification that affects both the quality and quantity of DNA.

It has been shown that fixatives produce DNA hydrolyzation and also the creation of methylene bridges between the amino-groups of pyrimidine and purine bases. In addition, it favorites the formation of cross-linking between bases and histones and between different DNA strands. Furthermore, scission of the phosphor-diester backbone of DNA may also occur, resulting in recovery of fragmented low-molecular-weight DNA (Ullah et al. 2017).

DNA extraction from these samples along with its amplification is a challenging task because DNA is often highly cross-linked, degraded, and fragmented. Obviously like the quantity, the degree of fragmentation also varies according to the preservation method and the storage time.

Many strategies and procedures have been developed to address these problems in order to generate a successful DNA profile (Jianghai et al. 2009; Janecka et al. 2015; Reid 2017; Agnieszka et al. 2019).

Contamination

DNA extraction is probably the step where the sample is more susceptible to a contamination. Although during years new protocols and kits have been developed for an efficient DNA extraction and purification, anyway contamination risk is still high during this step (Butler 2012). Because of this, a DNA laboratory must have specific procedures for reducing the risk of contamination and for monitoring it (NIST, NIJ 2013). All samples must be carefully handled and stored to avoid the introduction of extraneous DNA (e.g., from laboratory personnel) or sample-to-sample contamination. Contamination with biological samples belonging to the laboratory personnel may be avoided using PPE (personal protective equipment) such as lab coats, masks, and gloves.

In addition, in the laboratory separate work areas must be available for the different activities, and the access to the different areas must be limited only to authorized personnel. DNA laboratories should have a "DNA elimination database" that consists in a database including either DNA profiles of laboratory personnel or all contaminated DNA profiles found during analysis. The use of automatic platform offers the benefit of standardization, and it reduces the potential risk of contamination due to manual handling. Furthermore, dedicated instruments and disposable material must be used, and it is necessary to clean accurately all equipment and instruments. Sample-to-sample contamination may be reduced if evidences are properly handled and stored. Furthermore to reduce any contamination risk, it is necessary to maintain physically separated evidence and reference samples and to process them in different frames and laboratories, when possible (NIST, NIJ 2013).

Anyway, contamination of DNA extraction reagents may be sometimes an issue. Testing all extraction reagents before the use may be useful for monitoring contamination; appropriate control reactions are useful in determining whether DNA contamination has occurred, and they should be included in each analysis batch. The "blank" reaction control includes all reagents except the DNA sample. It is used to detect any DNA contamination of the reagents used for analysis. The negative control includes sterile water instead of the DNA sample, and it is useful to detect any procedural errors (ENFSI 2017). Sometimes it may be useful to collect and process an unstained portion of the item, as control. This may help in case of mixture to detect any substrate contamination.

Furthermore, it is relevant to take in consideration that samples may become contaminated at the crime scene, during collection, packaging, or transportation to

the storage location or to the forensic laboratory. Obviously, the use of precautions and accurate collection and preservation procedures strongly reduce the risk of contamination.

Because of this, either crime scene personnel or laboratory staff should receive training for the minimization of contamination (NIST, NIJ 2013).

Finally, it must be considered that, sometimes, before the arrival of the examiners, the crime scene may have been contaminated by rescues, family members, or curious who can adventitiously leave their traces (touch DNA). Invisible cells are easily transferred with the contact and impossible to be detected.

Due to this even if stringent protocols are applied, unfortunately casual contamination may occur, and it is often very difficult to distinguish contaminated samples from real mixtures.

Inhibitors

The problem of inhibition is common in forensic DNA typing. In fact, biological samples found at crime scenes often contain substances which may act as PCR-inhibitors.

Inhibitors may be co-extracted with DNA, and they interfere with the interaction between DNA and Taq polymerase by binding the active site of the Taq DNA polymerase and reducing its activity during PCR. This can lead to incomplete DNA profiles or complete PCR amplification failure (Alaeddini et al. 2011). Missing relevant DNA profiles consequently can result in an unsolved case or wrong suspect. Because of this, when extracting DNA from biological evidence, an extensive DNA purification is required to remove these inhibitors, although these procedures may increase the risk of DNA loss.

The best way to avoid PCR inhibition is to prevent the inhibitor from being processed with the sample.

This is difficult because some inhibitors are present in the samples themselves; others can be contained in the substrate where the biological material is placed, while others are introduced during the analytical process (Schrader et al. 2012). Inhibitors commonly found in forensic samples include hemoglobin in blood, indigo dyes in denim, humic acid and tannic acid in soil, melanin in hairs, calcium and collagen in bone and tissues, and bile salts and catabolic substances in feces and urine (Rådström et al. 2004). Other important inhibitors are materials and reagents that come into contact with samples during the analysis (e.g., salts, metal ions, organic solvents, etc.) (Table 4).

Commercial kits nowadays available (based on spin columns or resin systems) provide a sufficiently pure DNA extract, removing inhibitors effectively while simultaneously concentrating DNA.

In addition in the last 10 years, great improvements have been made also in the preparation of PCR multiplexes, and in fact, many commercial kits available and commonly used include improved buffer components with greater tolerance for inhibitors eventually present in the samples (GlobalFiler PCR Amplification Kit,

Table 4 Common Inhibitors (according to Rådström et al. 2004)

Inhibitor	Source
Hemoglobin, heme, lactoferrin	Blood
Myoglobin	Muscle tissue
Immunoglobin G (IgG)	Blood
Melanin and eumelanin	Hairs, skin
Indigo dye	Denim
Calcium ions	Milk, bones
Proteinases	Milk
Bile salts	Feces
Complex polysaccharides	Feces, plant material
Collagen	Tissues
Humic acid	Soil, plant, material
Urea	Urine
Polysaccharides	Plants

Thermo Fisher Scientific; Investigator 24plex, Qiagen; PowerPlex® 21 System, Promega).

This underscores the importance to combine an efficient DNA extraction method with a sensitive and robust multiplex STR amplification systems in order to improve the ability of forensic laboratories to obtain a better DNA profile from challenging samples.

References

Agnieszka K et al (2019) DNA extraction from FFPE tissue samples – a comparison of three procedures. Contemp Oncol (Pozn) 23(1):52–58

Alaeddini A et al (2011) Forensic implications of PCR inhibition-a review. Forensic Sci Int Genet 6 (3):297–305

Barbaro A et al (2009) Validation of PrepFiler™ forensic DNA extraction kit (applied biosystems). Forensic Sci Int Genet Suppl Series 2(1):176–177

Barbaro A et al (2011a) Direct PCR by the AmpF lSTR NGM™ kit for database purpose. Forensic Sci Int Genet Suppl Ser 3:1

Barbaro A et al (2011b) Validation of BTA lysis buffer for DNA extraction from challenging forensic samples. Forensic Sci Int Genet 3:61–62

Barbaro A et al (2019) DNA extraction from soil by EZ1 advanced XL (Qiagen). Forensic Sci Int 299:161–167

Barbaro A et al (2021) Highly efficient and automated extraction of DNA from human remains using a modified EZ1 protocol. Forensic Sci Res. https://doi.org/10.1080/20961790.2020.1848138

Berensmeier S et al (2006) Magnetic particles for the separation and purification of nucleic acids. Appl Microbiol Biotechnol 73:495–504

Bienvenue JM et al (2010) DNA extraction on microfluidic devices. Forensic Sci Rev 22:187–197

Builes JJ et al (2013) Teeth as a source of DNA for forensic identification of human remains: a review. Sci Justice 53:433–441

Butler JM (2012) Advanced topics in forensic DNA typing: methodology. Academic Press, Elsevier, USA. ISBN 978-0-12-374513-2

Coble MD et al (2008) A ten year study of DNA blood references collected on untreated filter paper and stored at room temperature. Proc Am Acad Forensic Sci 14:90–91
Comey C et al (1994) DNA extraction strategies for amplified fragment length polymorphism analysis. J Forensic Sci 39(5):1254–1269
Dahm R (2004) Friedrich Miescher and the discovery of DNA. Elsevier, Amsterdam
Dahm R (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Hum Genet 122(6):565–581
DNA Contamination prevention guidelines, ENFSI DNA Working Group (2017) European Network of Forensic Science Institutes. Available at http://enfsi.eu/wp-content/uploads/2017/09/DNA-contamination-prevention-guidelines-v2.pdf
Duncan E et al (2003) Isolation of genomic DNA. In: Bowien B, Dürre P (eds) Nucleic acids isolation methods. American Scientific Publishers, Stevenson Ranch, pp 7–19
Elliott K et al (2003) Use of laser microdissection greatly improves the recovery of DNA from sperm on microscope slides. Forensic Sci Int 137(1):28–36
Gill P et al (1985) Forensic application of DNA fingerprints. Nature 318:577–579
Greenspoon SA et al (1998) QIAamp spin columns as a method of DNA isolation for forensic casework. J Forensic Sci 43:1024–1030
Hawkins T (1998) DNA purification and isolation using magnetic particles. http://www.google.com/patents/US5705628
Janecka A et al (2015) Comparison of eight commercially available kits for DNA extraction from formalin-fixed paraffin-embedded tissues. Anal Biochem 476:8–10
Jianghai L et al (2009) High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. Anal Biochem 395(2):265–267
Kishore R et al (2006) Optimization of DNA extraction from low-yield and degraded samples using the BioRobot EZ1 and BioRobot M48. J Forensic Sci 51:1055–1061
Lee HY et al (2010) Simple and highly effective DNA extraction methods from old skeletal remains using silica columns. Forensic Sci Int Genet 4:275–280
Lee SB et al (2017) DNA extraction methods in forensic analysis, encyclopedia of analytical chemistry. Wiley & Sons, Ltd., USA, online
Liu JY et al (2012) AutoMate express™ forensic DNA extraction system for the extraction of genomic DNA from biological samples. J Forensic Sci 57(4):1022–1030
Montpetit SA et al (2005) A simple automated instrument for DNA extraction in forensic casework. J Forensic Sci 50(3):555–563
Nagy M et al (2005) Optimization and validation of a fully automated silica-coated magnetic beads purification technology in forensics. Forensic Sci Int 152:13–22
Pajnič IZ et al (2016) Highly efficient automated extraction of DNA from old and contemporary skeletal remains. J Forensic Legal Med 37:78–86
Pray L (2008) Discovery of DNA structure and function: Watson and Crick. Nature Education 1 (1):100
Rådström P et al (2004) Pre-PCR processing: strategies to generate PCR-compatible samples. Mol Biotechnol 26:133–146
Reid KM (2017) A review of the optimisation of the use of formalin fixed paraffin embedded tissue for molecular analysis in a forensic post-mortem setting. Forensic Sci Int 280:181–187
Sambrook J et al (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, USA. ISBN 978-0879693091
Schoell WM et al (1999) Separation of sperm and vaginal cells with flow cytometry for DNA typing after sexual assault. Obstet Gynecol 94:623–627
Schrader C et al (2012) PCR inhibitors – occurrence, properties and removal. J Appl Microbiol 113: 1014–1026
Sibille I et al (2002) Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa. Forensic Sci Int 125:2–3
Stray JE et al (2010) Extraction of DNA from human remains. Forensic Sci Rev 22(2):177–185
Tan SC (2009) DNA, RNA, and protein extraction: the past and the present. J Biomed Biotechnol 2009:574398

Technical Working Group on Biological Evidence Preservation (2013) The biological evidence preservation handbook. National Institute of Standards and Technology (NIST, NIJ)
Ullah S et al (2017) DNA perspectives of fixed and paraffin embedded human tissues as resource materials for the identification. Egypt J Forensic Sci 7:23
Vuichard S et al (2011) Differential DNA extraction of challenging simulated sexual-assault samples: a Swiss collaborative study. Investig Genet 2:11
Walsh PS et al (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10(4):506–513
Wang DY et al (2009) Direct amplification of STRs from blood or buccal cell samples. Forensic Sci Int: Genet Suppl Ser 2:113–114
Witt S et al (2012) Establishing a novel automated magnetic bead-based method for the extraction of DNA from a variety of forensic samples. Forensic Sci Int Genet 6(5):539–547

DNA Extraction in Human Bodies: From Fresh to Advanced Stages of Decomposition

33

Venusia Cortellini, Lorenzo Franceschetti, Heitor S. D. Correa, and Andrea Verzeletti

Contents

Abstract

From the beginning of its application in forensics, DNA analysis has been considered the gold standard in human identification: it has become an indispensable tool in criminal investigation, disaster victim identification, and paternity testing. DNA typing can be divided into several steps that go from DNA extraction to PCR amplification and genotyping. The DNA extraction step is crucial because it determines whether there are any intact DNA molecules to be analyzed in the first place. Improvements in DNA typing success may be attributed to advances in PCR performance, which have been constantly increasing its sensitivity, tolerance to inhibitors, and multiplexing capability. However, DNA analysis will fail if the DNA is not efficiently extracted and purified from the biological sample. According to the ease of collection and decomposition state of the corpse, there are different biological substrates to choose from. Depending on

V. Cortellini · L. Franceschetti · H. S. D. Correa · A. Verzeletti (✉)
Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, Forensic Medicine Unit, University of Brescia, Brescia, Italy
e-mail: venusia.cortellini@unibs.it; h.correa@unibs.it; andrea.verzeletti@unibs.it

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_37

the expected number of nucleated cells, biochemical composition of the tissue, and presence of inhibitory substances, each biological substrate requires a specific DNA extraction method. Therefore, the objective of this chapter is to give the reader a theoretical as well as practical basis for performing efficient DNA extraction from different biological substrates collected from fresh or decomposing bodies. The advantages and disadvantages of different methods as well as full protocols are presented so that the reader may choose one that is best suited for them.

Keywords

Forensic genetics · DNA extraction · Legal medicine · Human identification · Decomposed bodies

Highlights

- Depending on the stage of decomposition, different types of biological specimens may be sampled for DNA analysis.
- DNA extraction is a critical step in human identification.
- Without efficient DNA isolation and purification, DNA typing may fail.
- Each type of biological sample may require a specific pretreatment or approach.
- DNA extraction protocols are described for the most used samples of fresh and decomposed bodies.

Introduction

The field of forensic science is constantly changing as new methods and techniques are available and validated for use in the medico-legal system. In this context, genetic science applied to issues of human identification after a disaster, during criminal investigations and in resolving parentage disputes involving deceased persons, has experienced an especially rapid rate of advancement during the last decades (Butler 2005).

Nowadays, DNA profiling has emerged as the gold standard for human identification purposes providing an ability in identifying persons, in paternity testing, and providing investigative leads at a relatively low cost, rather quickly, and with a high degree of discrimination. In particular, recently, DNA profiling has replaced anthropological evaluation and other identification techniques of human remains. In fact, genetic methods give a quantitative result: statistics are performed on the specific traits studied providing the probability of two individuals having similar characteristics for the distribution of different alleles within a population is related.

On the contrary, forensic anthropology and odontology methods, which compare the status and shape of teeth and bones, are less costly methods, but they suffer from the nonquantifiability of the results, since a unanimous and clear agreement on the

quality and quantity of the characters necessary in order to achieve personal identification does not exist: the recurrence of discordant characters settles the case by excluding the identity; some or many concordant characters can only permit expression of a judgment of compatibility or possibility, while a combination of rare characters among the population will allow one to express a judgment of high probability or certainty (Cattaneo et al. 2005a; Cattaneo and Gibelli 2017).

For the majority of "fresh" samples, DNA profiling can be achieved in a fast, streamlined, and high-throughput manner. However, a large number of postmortem biological samples could be extremely compromised, making DNA examinations difficult. Genotyping of postmortem samples was often reported unsuccessful due to low DNA yield, the presence of inhibitors in DNA preparations, and DNA fragmentation/degradation brought about by exposure to harsh environmental conditions and microbial nucleases (Loreille et al. 2007).

New developments in technology and in forensic biology have expanded the options for the collection, sampling, preservation, and processing of samples for DNA typing. Moreover, recent advances currently permit a massive array of novel genetic markers and genotyping techniques to extract the maximum amount of genetic information from a sample, ensuring that identification is not only precise but also possible where material is degraded, or limited.

In fact, while initially DNA analysis could only be carried out on living subjects, using biological samples containing large amount of DNA, the advent of new techniques has made it possible to obtain excellent results even on samples containing extremely degraded and/or quantitatively limited DNA. Furthermore, if initially DNA analysis was carried out purely for parental purposes, today the fields of application have greatly expanded: DNA obtained from a dead corpse can be essential in personal human identification of missing persons and mass disaster victims, as well as in criminal investigations, but also for evolutionary genetic studies (Higgins and Austin 2013). Facing cadavers in different stages of decomposition can be very challenging for the forensic pathologist which has the duty to collect biological samples from deceased persons with the purpose of resolving parentage or kinship issues, as well as for criminalistic purposes. Hence, forensic autopsy is an important process for personal identification and investigation of cause and manner of death, but also fundamental for the collection of biological samples for genetic analyses (Madea et al. 2010).

The choice of matrices collected from cadavers as a source of DNA depends mostly on the preservation of the corpse. When dealing with fresh or partially decomposed cadavers, biological fluids (specially blood), soft tissues (muscle), and hair as a source of DNA are the samples of choice, as they contain large amounts of DNA and require short, common, easy, and cheap laboratory procedures (Schwark et al. 2011). Over time, soft tissues suffer from an accelerated postmortem degradation making DNA testing difficult (Hansen et al. 2014). Little is known about the influence that individual signs of putrefaction may have on the analyzability of DNA. Therefore, results can be hindered by DNA degradation as the quantity and the quality of DNA extracted from the remains is affected by the decay process of the human body, leading to the dissolution of soft tissues (Nouma et al. 2016). The pace

of degradation varies considerably depending on the tissue types, thus resulting in variable DNA degradation (Van den Berge et al. 2016).

However, when the corpse shows severe signs of degradation, it is standard procedure to collect samples from tissues which can guarantee better DNA availability, that is, hard tissues, such as bones and teeth that remain intact, given that they are resistant to autolysis and putrefaction (Rubio et al. 2013). For this reason, hard tissues may be the only sources of DNA in many forensic cases (Cattaneo et al. 2005b). Bone tissues are preferred for DNA-based identification in circumstances such as corpses in a very poor state of preservation, mutilated bodies, mass disasters, and exclusive presence of skeletal remains in cases of either war, fire, explosions, or criminal cases (Putkonen et al. 2010).

The present chapter aims at describing different DNA extraction techniques from fresh and decomposed/skeletonized human bodies. In the following pages, the reader will find recommendations in order to extract DNA from different matrices (saliva, blood, soft tissues, organs, hair, nails, teeth, and bones) and references to validated protocols already widespread in the forensic community.

DNA Extraction from Fresh Bodies

Overview

Before diving into the technical aspects of the DNA extraction process, it is worth discussing the particularities of extracting DNA from fresh versus decomposed human bodies. Fresh bodies are those in which decomposition and putrefactive processes have not started and still possess tissue and organ integrity. That way, commonly used samples are those that are easy to collect and provide high amounts of DNA, such as blood and buccal swabs. Alternatively, urine and hair can also be used. The practitioner should keep in mind that the integrity of the corpse should be preserved whenever possible, collecting samples, and employing sampling techniques that will not be visually perceptible.

DNA extraction of postmortem samples often involves optimized conditions to improve sample lysis and purification, since the human body progressively deteriorates after death. Considering that it is impossible to return to a prior state of less degradation, the amount of sample collected should be enough for eventual reanalysis. Also, decomposition is a continuous process and, even though visual changes may not be readily observed on a corpse, complex biochemical reactions are already occurring, producing unknown by-products that may potentially degrade DNA and hamper its analysis. Therefore, the sooner the collection of biological samples is performed, the highest the chances for successful DNA typing.

Any biological sample contains an enormous number of all four classes of biomolecules: proteins, lipids, carbohydrates, and nucleic acids. Forensic DNA analysis is dependent on the effective nucleic acid's amplification, which may fail if any lipids, carbohydrates, or proteins are not washed away. That means DNA extracts should contain, roughly, only DNA and water molecules in a solution capable of

stabilizing it while minimizing its degradation (Butler 2009). The objectives of the DNA extraction process can be classified in (a) cell lysis, which will release the DNA molecules, (b) separation of the DNA molecules from other cellular material, and (c) DNA isolation into a format compatible with downstream applications (Butler 2011).

The term DNA extraction comes from organic chemistry, since the first method described was by means of a liquid/liquid extraction using phenol and chloroform as organic solvents (Cattaneo et al. 2006). From then onwards, several techniques and protocols have been described that may be employed for DNA extraction depending on available costs, time, technology, sample substrate, degree of purity needed, among other factors. Below, we will discuss the most employed DNA extraction techniques when dealing with samples collected from fresh bodies, their advantages, and limitations. Depending on the expected DNA quantity, degradation, matrix digestibility, and presence of inhibitory substances, DNA extraction can be easily the most laborious phase of the forensic DNA typing process.

Biological Samples

High-quality DNA profiles are routinely obtained from any biological source. However, the following samples (blood, saliva, urine, and hair) are more commonly collected from fresh bodies due to their convenience, as discussed above.

Blood is a heterogeneous tissue composed of a liquid phase (plasma) and solid components: red blood cells (RBC), white blood cells (WBC), and platelets. Blood plasma is mostly made up of water, as well as circulating proteins and other solutes. Since only WBC possess a nucleus, these are the main sources of DNA in human blood.

Saliva is the biological fluid secreted in the buccal cavity that aids in the digestion process. The major constituent of saliva is water, where proteins/enzymes and immunoglobulins are present in low concentrations. It also contains a high quantity of epithelial and WBCs that compose the DNA source of the fluid. Saliva generates less DNA than blood, but it is very convenient to sample through buccal swabs.

Urine, on the other hand, is seldom used in forensic medicine, only in cases where blood or saliva are not collected. It is mostly water, urea, creatinine, and other organic and inorganic compounds. The scarce nucleated cells found in urine are generally epithelial cells and WBCs.

Hairs grow from dermal follicles and are primarily composed of keratin. They have two structures: the follicle or bulb, which is beneath the dermis, and the shaft, which is the scaly, hard, filamentous structure above the skin surface. The layers of the hair fiber include the cuticle, the cortex, and the medulla. Of the many places where hairs can grow on the human body, scalp, axillary, and pubic hairs may be useful in forensic DNA analysis when other matrices are not available.

The general public assumes hair to be an ideal sample for DNA examination, but the truth is that hair is the most overestimated sample in forensic genetics. Successful DNA extraction from hair is usually only possible when roots and adhering tissues

are present. During the keratinization process, nuclear cellular material is degraded in the hair shaft. Therefore, it is advised that a microscopic examination of hairs is performed prior to DNA extraction (Dash et al. 2020).

DNA Extraction

Depending on the biological matrix, expected DNA quantity, degradation, matrix digestibility, and presence of inhibitory substances, DNA extraction protocols can vary widely. On the other hand, advances in PCR buffer technology now allow direct PCR in which DNA extraction can be skipped altogether (Elkins 2012; Prinz and Lessig 2014). Direct PCR protocols are usually employed on fresh buccal swabs or dried stain punches and reduced number of cycles.

Another new technology worth mentioning is Rapid DNA, which consists of a fully automated sample-to DNA profile system in less than 2 h. DNA extraction, STR amplification, electrophoretic separation, fluorescent detection, and data analysis are performed in one instrument after a few minutes of hands-on sample preparation (Kitayama et al. 2020). On the following subheadings, the most common techniques employed in DNA extraction will be described. After a brief introduction, for each biological sample, a detailed protocol will be provided.

Chelex

One of the simplest techniques to extract DNA from biological samples is by using Chelex® 100 resin (Bio-Rad, Hercules, CA, USA). Chemically, Chelex is an ion exchange resin composed of styrene divinylbenzene copolymers that contain paired iminodiacetate ions. This resin chelates polyvalent metal ions, such as magnesium, which inactivate DNases, protecting DNA from being degraded. In DNA extraction protocols, Chelex is used as a 5% suspension, which has a pH between 10 and 11. Being alkaline, when heated to 100 °C, cell membranes are disrupted, releasing DNA. The released DNA, as well as proteins, is denatured. Denatured proteins lose water solubility and can be pelleted by centrifugation along with other cellular debris. The supernatant containing dissolved DNA may be used directly in PCR or other molecular biology applications (Walsh et al. 1991).

The advantages of using Chelex based protocols are that it is fast, simple, and inexpensive. The disadvantages are that it is not as effective in samples containing inhibitory substances, which are common in forensic cases. Therefore, blood, oral mucosa, and urine can be successfully extracted with Chelex in samples collected from fresh cadavers due to an elevated amount of high molecular weight DNA and lower concentration of decomposition by-products that may inhibit downstream analyses (Phillips et al. 2012). For each biological substrate, sample preparation is illustrated on Table 1.

Chelex Protocol

1. Resuspend the sample in 170 μL of 5% Chelex suspension
2. Incubate at 56 °C for at least 30 min

Table 1 Sample preparation for Chelex extraction from each type of biological matrix

Biological matrix	Preparation
Blood	Add 1 mL of deionized water to a microtube Add 3 μL whole blood Vortex briefly Incubate at room temperature for 15 min Centrifuge at 10,000 g for 3 min Discard supernatant except final 50 μL with pellet
Saliva	Cut a third of a buccal swab and add it to a microtube
Urine	Add 1 mL of urine to a microtube Centrifuge at 10,000 g for 3 min Discard supernatant except final 50 μL with pellet
Hair	Cut 0.5–1 cm of the root end of a strand of hair with sheath Place the hair root in a microtube

3. Vortex at high speed for 5 s
4. Incubate at 100 °C for 8 min
5. Vortex at high speed for 5 s
6. Centrifuge at 10,000 g for 3 min to pellet the resin and cellular debris

Notes: For short-term storage (up to 1 month), store extracts at 2–8 °C with the Chelex resin. For long-term storage (>1 month), the extracts should be centrifuged, and the supernatant should be removed from the Chelex resin and stored in a new tube at −20 °C

Whatman FTA® Paper

Another extraction method that is easily employed in biological samples collected from fresh corpses is based on Whatman FTA® paper. FTA paper is a cellulose-based absorbent substrate that contains additives which lyse cells, protect DNA molecules from degradation by nucleases, as well as prevent the growth of microorganisms. After the sample has dried, DNA on FTA paper is stable for several years under room temperature. A small punch of the paper is washed to remove unwanted substances while retaining the DNA. The clean punch is added directly to the PCR (Burger et al. 2005). If the above methods are not able to yield good enough results, another extraction method that provides improved purification, such as those described below, may be employed.

Organic Method

Organic, or phenol-chloroform, extraction receives its name due to the use of the organic solvents phenol and chloroform. It requires the addition of certain species of chemicals so it can achieve its purpose. A surfactant, such as sodium dodecylsulfate (SDS), is added to the aqueous buffer containing the sample to disrupt the phospholipidic bilayer that composes cellular membranes and denature proteins. The enzyme proteinase K is added to digest nucleases and proteins that surround the

Table 2 Sample preparation for organic extraction from each type of biological matrix

Biological matrix	Preparation
Blood or saliva	Add 10 μL of whole blood or saliva to a microtube
Swab with blood or saliva	Cut a third of the swab and place it in a microtube
Urine	Add 1 mL of urine to a microtube Centrifuge at 10,000 g for 3 min Discard supernatant except final 50 μL with pellet
Hair	Cut 0.5–1 cm of the root end of a strand of hair with sheath Place the hair root in a microtube

DNA molecules. Next, a phenol-chloroform mixture is added to extract lipids and separate the denatured proteins from the aqueous solution containing the DNA. After centrifugation, two separate phases are observed: the organic lower phase and the aqueous upper phase. DNA is more soluble in the aqueous phase than in the organic phase. Collecting the clean aqueous phase, DNA molecules can be transferred for analysis (Butler 2009; Butler 2011; Cattaneo et al. 2006; Elkins 2012; Prinz and Lessig 2014).

The organic method provides high-molecular-weight DNA with high purity, but currently there are other methods that use chemicals which are far less toxic than phenol. In addition, organic extraction is not yet amenable to automation, is time consuming, and requires multiple tube transfers, increasing the risks of error and contamination. For each biological substrate, sample preparation is illustrated on Table 2.

Organic Protocol

1. Add 400 μL of extraction buffer (SDS 2%, EDTA 0.01 M, pH 8) to the microtube containing the sample
2. Add Proteinase K enzyme (20 mg/mL) in a ratio of 10:1 (extraction buffer: Proteinase K)
3. Add DTT (0.4 M) (equal to the volume of proteinase K)
4. Vortex the sample for 5 s and then centrifuge briefly
5. Incubate the sample at 56 °C overnight
6. Discard the swab, fabric, or substrate, if the case
7. Add an equal volume of phenol–chloroform–isoamyl alcohol (25:24:1) to the lysate. Vortex and centrifuge at 14,000 g for 5 min. Transfer the aqueous supernatant to a new microtube. Repeat this step 2–3 times until no impurities are seen in the interphase
8. Transfer the upper aqueous phase to a Microcon® 100 Centrifugal Filter. Centrifuge at 4000 g for 20–30 min, or until the liquid has been reduced to the minimum retained volume
9. Wash the filtrated Microcon® with 500 μL of DNA-free water and centrifuge at 4000 g for 20–30 min. Repeat this step 2– 3 times
10. Elute the DNA in 50–100 μL of DNA-free water by inverting the Microcon® into a new 1.5 mL tube. Centrifuge at 4000 g for 3–5 min

Solid-Phase

A natural evolution from the organic method would be to use nontoxic chemicals and automate the DNA extraction process. Solid-phase extraction methods appeared with that proposal. In those methods, silica-based technology is involved, producing better DNA recovery and more efficient removal of inhibitors. The most common techniques are based on the same chemical principle: one method works with silica membranes in combination with a chaotropic salt-containing binding buffer. Another one uses silica-covered magnetic beads in combination with a chaotropic salt-containing binding buffer. Other protocols rely on the use of a silica suspension. DNA is absorbed by the silica surface in the presence of a high concentration of salt and a pH $\leq$7.5. The method of choice should be decided by the quality of the samples. Depending on the biological substrate, the solid-phase employs a silica membrane having specific characteristics. Basically, all protocols differ in the use of decalcification and digestion buffer, incubation temperature, and time. Stringent washes purify the bound DNA molecules from proteins and other contaminants, and, in the end, a nonstringent wash elutes highly purified DNA (Greenspoon et al. 1998). The disadvantages are the higher costs, complexity, more hands-on time, and tube changes, which raises risks for error and contamination. Alternatively, there are several biotechnology companies that commercialize solid-phase DNA extraction kits using an automated platform, increasing costs but reducing time, complexity, error, and contamination (Prinz and Lessig 2014; Kitayama et al. 2020).

QIAamp DNA Investigator Kit (QIAGEN, Hilden, Germany) Protocol

Blood and Saliva

1. Pipet 1–100 μL of whole blood or saliva into a 1.5 mL microtube
2. Add Buffer ATL to a final volume of 100 μL
3. Add 10 μL of proteinase K
4. Add 100 μL of Buffer AL and close the lid
5. Vortex for 15 s
6. Incubate at 56 °C for 10 min
7. Briefly centrifuge the microtube to remove drops from inside the lid
8. Add 50 μL of ethanol (96–100%), close the lid, and vortex for 15 s
9. Incubate for 3 min at room temperature
10. Briefly centrifuge the microtube to remove drops from inside the lid
11. Follow the purification and elution (for all samples) step below

Buccal Swab

1. Cut off the cotton head of the buccal swab and place it in a 1.5 mL microtube
2. Add 20 μL of proteinase K and 400 μL of Buffer ATL and close the lid
3. Vortex for 10 s
4. Incubate at 56 °C for 1 h;
5. Add 400 μL of Buffer AL and close the lid
6. Vortex for 15 s

7. Incubate at 70 °C for 10 min with shaking at 900 rpm for 10 min
8. Briefly centrifuge the microtube to remove drops from inside the lid
9. Add 200 μL of ethanol (96–100%) and close the lid
10. Vortex for 15 s
11. Briefly centrifuge the microtube to remove drops from the inside of the lid
12. Follow the Purification and elution (for all samples) step below

Urine

1. Add 1 mL of urine to a 1.5 mL microtube
2. Centrifuge at 10,000 g for 3 min to pellet cells
3. Discard the supernatant except the final 50 μL containing the pellet
4. Add 500 μL of Buffer AE
5. Vortex for 5 s
6. Centrifuge at 6000 g for 2 min
7. Discard the supernatant
8. Add 300 μL of Buffer ATL and 20 μL of proteinase K to the pellet
9. Vortex for 10 s

   Note: Adding 20 μL of 1 M DTT may increase sensitivity, since urine can contain sperm cells which can only be lysed in the presence of reducing agents.
10. Incubate at 56 °C for 1 h
11. Briefly centrifuge the microtube to remove drops from inside the lid
12. Add 300 μL of Buffer AL, close the lid, and vortex for 10 s
13. Incubate at 70 °C for 10 min with shaking at 900 rpm
14. Briefly centrifuge the microtube to remove drops from the inside of the lid
15. Add 150 μL of ethanol (96–100%), close the lid, and vortex for 15 s
16. Briefly centrifuge the microtube to remove drops from inside the lid
17. Follow the Purification and elution (for all samples) step below

Hair

1. Cut off 0.5–1 cm from the root end of a strand of hair with sheath
2. Place the cut off root in a 1.5 mL microtube
3. Add 300 μL of Buffer ATL, 20 μL of proteinase K, and 20 μL of 1 M DTT, close the lid
4. Vortex for 10 s
5. Incubate at 56 °C for 1 h with shaking at 900 rpm
6. Briefly centrifuge the tube to remove drops from the inside of the lid
7. Add 300 μL of Buffer AL, close the lid, and vortex for 10 s
8. Incubate at 70 °C for 10 min with shaking at 900 rpm
9. Briefly centrifuge the microtube to remove drops from the inside of the lid
10. Add 150 μL of ethanol (96–100%), close the lid, and vortex for 15 s
11. Briefly centrifuge the microtube to remove drops from the inside of the lid
12. Follow the Purification and elution (for all samples) step below

Purification and Elution (for All Samples)

1. Carefully transfer the entire lysate to the QIAamp MinElute column (in a 2 mL collection tube) without wetting the rim, close the lid, and centrifuge at 6000 g for 1 min

   Note: If the lysate has not completely passed through the membrane after centrifugation, centrifuge again at a higher speed until the QIAamp MinElute column is empty
2. Carefully open the QIAamp MinElute column and add 500 µL of Buffer AW1 without wetting the rim. Close the lid and centrifuge at 6000 g for 1 min
3. Place the QIAamp MinElute column in a clean 2 mL collection tube and discard the collection tube containing the flow-through
4. Carefully open the QIAamp MinElute column and add 700 µL of Buffer AW2 without wetting the rim. Close the lid and centrifuge at 6000 g for 1 min
5. Place the QIAamp MinElute column in a clean 2 mL collection tube and discard the collection tube containing the flow-through
6. Carefully open the QIAamp MinElute column and add 700 µL of ethanol (96–100%) without wetting the rim. Close the cap and centrifuge at 6000 g for 1 min
7. Place the QIAamp MinElute column in a clean 2 mL collection tube and discard the collection tube containing the flow-through
8. Centrifuge at full speed (20,000 g) for 3 min to dry the membrane completely
9. Place the QIAamp MinElute column in a clean 1.5 mL microtube and discard the collection tube containing the flow-through. Carefully open the lid of the QIAamp MinElute column and incubate at room temperature (15–25 °C) for 10 min or at 56 °C for 3 min
10. Apply 20–100 µL of Buffer ATE or ultrapure water to the center of the membrane
11. Close the lid and incubate at room temperature for 5 min. Centrifuge at full speed for 1 min

    Notes: The extract is ready to be used in PCR-based amplification. For short-term storage (up to 1 month), store extracts at 2–8 °C. For long-term storage (>1 month), the extracts should be stored at − 20 °C

DNA Extraction from Decomposed/Skeletonized Bodies

Overview

Following death, the natural progression of the body's decomposition or postmortem changes starts. At the cellular level, a complex series of biochemical and pathological processes initiates, resulting in considerable alteration of the structure and composition of the entire human body. These changes start immediately after death, occur sequentially, and continue for a prolonged time at different rates for different organs (Zhou and Byard 2011).

The onset, the extent, and the rates of these changes are affected, accelerated, or decelerated, by multiple intrinsic and extrinsic factors, differing among geographical

regions and also in the same geographic region from one season to another (Sutherland et al. 2013). Intrinsic factors primarily include body mass and the surface area of the body, while extrinsic factors primarily include the subject's clothing and insulation, the environment of the death scene, and the storage of the body after death.

Two mechanisms are involved in the postmortem process of decomposition: autolysis and putrefaction (Cockle and Bell 2017). Autolysis is a process that occurs due to leakage of hydrolytic cellular enzymes from cells after death, especially in those organs with high concentration of cellular enzymes like the pancreas. The leakage of cellular contents is a suitable environment for microbes normally composing the human microbiota which grow and degrade surrounding tissues, starting putrefaction (Cattaneo et al. 2005a). Unlike autolysis, putrefactive changes are evident on a macroscopic level: the skin shows discoloration and body parts such as the face, abdomen, breast, and scrotum start to bloat. It can appear in various forms, such as putrefactive fluids and gases.

There are five stages of decomposition, namely: fresh, bloated, active decay, advanced decay, and skeletal stage (Lee Goff 2009). These stages may occur at the same time in different parts of the same body, so for the forensic practitioner it could be challenging to label the decomposition of the corpse within a single stage. The speed of the onset of putrefaction and its rate of progression are affected mostly by ambient temperature. In addition, blowflies and flesh flies are often the first insects to be attracted to a dead body, primarily during the bloated and decay stages of decomposition. In the bloated stage, body parts, including organs and soft tissues, swell due to the accumulation of putrefactive gases or other decomposition products from the putrefaction process (Gebhart et al. 2012). Active decay is the stage after bloating, where postmortem purging takes place and putrefactive body fluids become forced out of body orifices. The detachment of hair or hair sloughing and black discoloration of ruptured skin are seen. Advanced or late decay is a stage where bones begin to get exposed, and the body assumes a "caved in" appearance. Degradation-resistant tissues such as hair (although already sloughed off) and cartilage are spared up to this stage. The skeletal stage, also called dry remains stage or skeletonization, begins when bone exposure is extensive (Swann et al. 2010). Remaining dry skin, cartilage, and tendons are minimal in this stage. Decomposition considerably decelerates at this stage, and it takes years or decades for the skeletal remains to disintegrate. Differential decomposition processes involving mummification or adipocere formation are also reported (Ubelaker and Zarenko 2011).

The macroscopic changes discussed above reflect also on a cellular level. When an organism dies, the cells' DNA start to degrade and become susceptible to damage and chromosomal fragmentation: the higher the degradation of the tissues, the lower the DNA yield. Primarily, two factors (time and condition of exposed environment) influence the decomposition as well as degradation of DNA evidence (Dash et al. 2020). In particular, the process of degradation increases with time. The rate and the intensity of the decomposition process, and consequently the degradation of the samples, are dependent on several other variables: the environmental conditions such as climate, temperature, humidity, pH and ultraviolet radiation, the presence of

scavengers, natural and artificial inhibitor substances, the rate of bacterial growth, and the body mass index of the deceased (Courts et al. 2015).

In forensic biological samples, DNA damage occurs mostly due to hydrolysis, oxidation, and pyrimidine dimer formation. In these cases, humidity plays a crucial role in DNA degradation, as DNA is easily damaged by hydrolysis resulting in deamination, depurination, and depyrimidination. Also, prolonged exposure to heat, ultraviolet radiation, microorganisms, and microbial nucleases further increases the rate of DNA damage and degradation resulting in fragmentation of the DNA molecule into smaller pieces. More than 70% of soil microorganisms contain nucleases, which make them capable of destroying nucleic acids (Milos et al. 2007). On the contrary, desiccation and/or low temperature can slower these chemical processes (Davoren et al. 2007). The genetic material obtained from those samples can be both degraded and chemically damaged (Fattorini et al. 2009), implying that no or low copy number of entire templates are available for successful DNA typing.

Biological Samples

Since tissues of a human cadaver decompose over time, forensic casework encounters a major chance of examining degraded biological samples such as soft tissues, organs, nails, teeth, and bones.

Soft tissues like muscle and organs are the source of DNA of choice dealing with partially decomposed cadavers. In general, they contain large amounts of DNA, although different tissue types are affected differently by the above-mentioned degradation processes. Putrefaction tends to be more active in tissues such as the kidney and liver, resulting in early DNA degradation. The most advised soft tissues for DNA analysis include the brain, muscle, kidney, and heart. The required laboratory procedures are short, simple, and affordable (Schwark et al. 2011). With ongoing putrefaction in soft tissues and organs, DNA integrity will continuously decrease up to a point at which analysis and profiling using standard methods is no longer possible and the collection of other matrices are needed, such as nails and hard tissues (Caputo et al. 2011).

Nails are made, like hair, of a tough protective protein called alpha-keratin with the aim of preserving the nail matrix, which contains nerves, lymph, and blood vessels. This structure ensures nail tissues high resistance to decay. They are a promising alternative as a source of DNA, even if nails appear to be less considered as a tissue of choice from decomposed bodies (Allouche et al. 2008). Although the DNA concentration in nail tissue is fairly low in comparison with hard tissues, it was shown that a certain constant concentration can be obtained regardless of the time elapsed after death (Kaneko et al. 2015).

Skeletal elements are composed by hard tissues, as teeth and bones, which are often the only biological materials remaining after exposure to environmental conditions, intense heat, certain traumatic events, and in cases where a significant amount of time has passed since the death of the individual (Latham and Miller 2019).

Teeth are one of the most useful sources of DNA for genetic analysis. They are a valuable source of DNA due to their unique composition and location protecting them from decay. The good state of preservation is granted by enamel, which is the hardest tissue of human body. Teeth roots, composed of cementum and pulp, are a rich source of DNA with high-quality and low-contamination.

In case of extremely damaged or degraded human skeletal remains, bones are the most suitable samples for genetic analysis. Bone possesses a complex anatomical structure made of many organic as well as inorganic molecules such as hydroxyapatite, collagen, and noncollagen proteins. The combination of organic as well as inorganic substances renders bones their hardness and makes them less prone to degradation. Therefore, the collection of bones can offer higher probability of obtaining preserved DNA since desiccation and/or low temperature can slower decomposition processes (Hochmeister et al. 1991). The goal of skeletal DNA extraction techniques is to maximize DNA yield, minimize any additional DNA damage, and remove any inhibitors that may co-purify with the skeletal DNA and interfere with later genetic analyses. The relationship between the bone type and the generation of an informative DNA profile has been investigated: compact bone of the lower limbs tends to yield greater amounts of DNA than spongy bone. The main disadvantages of collecting samples from bones and teeth are the requirement of tedious and time-consuming extraction procedures due to the need of mechanical fragmentation and tissue decalcification (Madea et al. 2010). Also, the high number of steps required by these processes increases the risk of contamination, which has to be always considered (Vanek et al. 2017).

Exogenous contamination plays a limiting role in the analysis and interpretation of DNA typing results. When the biological samples are contaminated by the above-mentioned contaminating substances, these compounds are co-purified with the nucleic acids in spite of the accuracy of the protocol employed and act inhibiting the DNA polymerases (King et al. 2009). In these cases, it is mandatory to perform a new extraction by freshly prepared solutions and new consumables.

DNA Extraction

DNA profiling of putrefied remains or skeletonized samples may be the only means to identify their human sources. With ongoing decomposition, the damage downgrades both the quality and quantity of DNA extracted from forensic samples and, consequently, minimizes the high discrimination power of routine forensic STR analyses by resulting in partial or no STR profiles (Alaeddini et al. 2010). In fact, although using STRs in DNA typing is currently the gold standard for human identification purposes, highly degraded samples often result in partial STR profiles because the larger loci (>250 bp) commonly fail to amplify due to fragmentation of the DNA structure (Sorensen et al. 2016).

In cases of soft tissues, organs, and nails, DNA extraction protocols are the same for the fresh bodies as previously described. Differently, purifying DNA from bones

and teeth requires modification of the usual DNA extraction techniques utilized for other biological samples (Latham and Miller 2019).

Most of these methods includes a pre-extraction treatment aimed at removing contaminating DNA transferred to the surface of the bone or tooth. Decontamination can be achieved by physically removing the outer bone surface, by immersing the bone or tooth in a bleach solution, or by exposure to ultraviolet radiation (Latham and Madonna 2014).

The hard tissues are often pulverized and subsequently incubated in extraction buffer and proteinase K, which work dissolving both the organic and inorganic portions of the bone tissue (Loreille et al. 2007). Grinding the sample into a powder (mechanically or using freezer mills) breaks the hydroxyapatite mineral matrix, increases the surface on which DNA extraction chemicals operate, and releases a greater amount of DNA. The amount of bone powder varies among protocols, requiring from 0.2 g to 2.5 g of starting material. The amount of powder needed for the extraction process has slowly been reduced thanks to the optimization of the DNA extraction process and the increased sensitivity of DNA kits which allow for minimal destruction of the skeletal samples (Hervella et al. 2015).

Also, hard tissue extraction can require another pre-extraction treatment aimed at clearing mineralization within the bone, which represents physical barrier to the extraction reagents and therefore inhibit the release of DNA molecules (Loreille et al. 2007). Many protocols for bones and teeth are based on the incubation in an ethylene diamine tetra-acetic acid (EDTA)-containing extraction buffer, in order to reach total decalcification (Rothe and Nagy 2016). The EDTA both demineralizes the hard tissues and inactivates DNAses by chelating bivalent cations such as Mg++ or Ca++ (Loreille et al. 2007).

As in the previous paragraphs, the following subheadings will illustrate the most common techniques employed in DNA extraction from advanced decomposed bodies and skeletonized remains. A brief introduction and a detailed protocol will be provided for each biological matrix.

Organic

Muscle, organs, nails, bones, and teeth preparation are presented on Table 3, while the protocol is described below.

Organic Protocol

1. Add 300 µL of extraction buffer (10 mM Tris HCl, 100 mM NaCl, 39 mM DTT, 10 mM EDTA, SDS 2%) and 20 µL of proteinase K to the sample
2. Incubate at 56 °C overnight
3. Add 300 µL of phenol:chloroform:isoamyl alcohol to each tube
4. Mix vigorously until a complete emulsion is formed
5. Centrifuge tubes for 3 min at 4000 g. There should be a clear delineation between the layers
6. Transfer the aqueous layer of each sample to clean tubes

Table 3 Decomposed sample preparation for organic extraction from each type of biological matrix

Biological matrix	Preparation
Muscle and organs	Add 10 mg of tissue to a tube
Nail	Add nail clippings to a tube
Bones and teeth	Cut a piece of bone or tooth roots (e.g., using a hammer or rotational device) Pulverize the bone or tooth piece (e.g., using a mill or blender) Add 0.5 to 1 g of powder to a tube Add 40 mL of EDTA 0.5 M, pH 7.5 to powdered sample for 24 h at 4 °C Shake or invert the tubes gently to completely saturate the bone powder. Continue to gently shake until no dry spots are visible in the powder Centrifuge tubes for 15 min at 2000 rpm and discard the supernatant Wash the pellet with sterile water, vortexing for 15 s Centrifuge tubes for 15 min at 2000 rpm and discard the supernatant

7. Repeat steps 4–7 until the interface is clean (or a minimum of two times)
8. Add 10 μL of NaAc 3 M and mix thoroughly
9. Add 2 times the volume of cold absolute ethanol and incubate at −20 °C for 2 h
10. Centrifuge tubes for 15 min at 4000 g
11. Discard the alcohol supernatant
12. Add 200 μL of 70% ethanol
13. Centrifuge tubes for 2 min at 4000 g
14. Discard the alcohol supernatant
15. Let the pellet dry thoroughly under hood
16. Add 50 μL of sterile water and incubate at 56 °C for 2 h

Solid-Phase

Muscle, organs, and nails preparation is shown in Table 3, while the protocol for these samples is described in "Solid-Phase" For bone and teeth samples with prolonged time elapsed since death, DNA extraction with the silica-membrane method showed to have more DNA yield than with the beads method, since when using the extraction method with the beads, DNA loss could occur. Therefore, the silica-coated beads method should be chosen for more recent skeletal samples or samples with high amounts of inhibitors, while the silica-membrane method is preferred for samples with low amounts of DNA. Skeletal sample preparation for both solid-phase extraction methods is given in Table 4.

QIAamp DNA Investigator Kit (QIAGEN, Hilden, Germany) Protocol

1. Add 360 μL of Buffer ATL, 20 μL of proteinase K, close the lid
2. Vortex for 10 s
3. Incubate at 56 °C overnight with shaking at 900 rpm
4. Briefly centrifuge the tube to remove drops from the inside of the lid
5. Add 300 μL of Buffer AL, close the lid, and vortex for 10 s

Table 4 Decomposed sample preparation for solid-phase extraction from each type of biological matrix

Biological matrix	Preparation		
	Silica-membrane		Silica-coated beads PrepFiler® BTA Forensic DNA Extraction Kit (Thermofisher, Waltham, MA, USA)
Bones	Cut a piece of bone (e.g., using a hammer or rotational device) Pulverize the bone piece (e.g., using a mill or blender) Add up to 100 mg of bone powder to a tube	Cut a piece of bone (e.g., using a hammer or rotational device) Add the bone piece in a solution of EDTA 0.5 M until it is demineralized	Cut a piece of bone (e.g., using a hammer or rotational device) Pulverize the bone piece (e.g., using a mill or blender) Add up to 50 mg of bone powder to a microtube
Teeth	Pulverize one tooth (e.g., using a mill or blender) Add up to 100 mg of tooth powder to a tube	Add the tooth in a solution of EDTA 0.5 M until it is demineralized	Pulverize one tooth (e.g., using a mill or blender) Add up to 50 mg of tooth powder to a microtube

6. Incubate at 70 °C for 10 min with shaking at 900 rpm
7. Briefly centrifuge the microtube to remove drops from the inside of the lid
8. Add 150 µL of ethanol (96–100%), close the lid, and vortex for 15 s
9. Briefly centrifuge the microtube to remove drops from the inside of the lid
10. Carefully transfer the entire lysate to the QIAamp MinElute column (in a 2 mL collection tube) without wetting the rim, close the lid, and centrifuge at 6000 g for 1 min

    Note: If the lysate has not completely passed through the membrane after centrifugation, centrifuge again at a higher speed until the QIAamp MinElute column is empty
11. Carefully open the QIAamp MinElute column and add 600 µL of Buffer AW1 without wetting the rim. Close the lid and centrifuge at 6000 g for 1 min
12. Place the QIAamp MinElute column in a clean 2 mL collection tube and discard the collection tube containing the flow-through
13. Carefully open the QIAamp MinElute column and add 700 µL of Buffer AW2 without wetting the rim. Close the lid and centrifuge at 6000 g for 1 min
14. Place the QIAamp MinElute column in a clean 2 mL collection tube and discard the collection tube containing the flow-through
15. Carefully open the QIAamp MinElute column and add 700 µL of ethanol (96–100%) without wetting the rim. Close the cap and centrifuge at 6,000 g for 1 min
16. Place the QIAamp MinElute column in a clean 2 mL collection tube and discard the collection tube containing the flow-through

17. Centrifuge at full speed (20,000 g) for 3 min to dry the membrane completely
18. Place the QIAamp MinElute column in a clean 1.5 mL microtube and discard the collection tube containing the flow-through. Carefully open the lid of the QIAamp MinElute column and incubate at room temperature (15–25 °C) for 10 min or at 56 °C for 3 min
19. Apply 20–50 μL of Buffer ATE or ultrapure water to the center of the membrane
20. Close the lid and incubate at room temperature for 5 min. Centrifuge at full speed for 1 min

Notes: Extracts may be held at − 20 or − 80 °C for extended storage. In the short term, 4 °C is adequate. It is best to minimize freeze/thaw cycles as freezing may damage the DNA

Silica-Coated Beads Protocol PrepFiler® BTA Forensic DNA Extraction Kit (Thermofisher, Waltham, MA, USA)

1. Add 230 μL of the following lysis buffer solution to each sample: 220 μL PrepFiler® BTA Lysis Buffer, 3 μL 1.0 M DTT and 7 μL Proteinase K
2. Cap the tubes, vortex for 5 s, centrifuge briefly, then gently flick to resuspend any powder or substrate
3. Make sure the tubes are well sealed, then place them in a thermal shaker and incubate at 1000 rpm and 56 °C for 2 h
4. Allow the sample to equilibrate to room temperature
5. Centrifuge the sample tubes at 10000 g for 90 s, then carefully transfer the clear lysate to a new 1.5 mL tube
6. If needed, add PrepFiler® BTA Lysis Buffer to bring the total sample lysate volume to 180 μL
7. Add 300 μL of PrepFiler® Lysis Buffer to the sample lysate tube, vortex it briefly to mix, then centrifuge it briefly
8. Vortex the PrepFiler® Magnetic Particles tube for 5 s, invert the tube to confirm that no visible pellet remains in the bottom of the tube, then centrifuge it
9. Pipet 15 μL of thoroughly resuspended magnetic particles into the sample lysate tube
10. After adding the particles, recap the PrepFiler® Magnetic Particles tube to prevent evaporation
11. Cap the sample lysate tube, vortex it at low speed for 10 s, then centrifuge it briefly to collect any residual tube contents from the sides and cap of the tube
12. Add 300 μL isopropanol (99.5% molecular biology grade) to a sample lysate tube and mix one sample at a time to promote binding
13. Immediately after adding isopropanol, cap the sample lysate tube, vortex it at low speed for 5 s, then centrifuge it briefly to collect any residual tube contents from the sides, and cap of the tube
14. Place the sample lysate tube in a shaker, then mix at 1000 rpm at room temperature for 10 min
15. Vortex the sample DNA tube at maximum speed (approximately 10,000 rpm) for 10 s, then centrifuge briefly to collect any residual tube contents from the sides and cap of the tube

16. Place the sample DNA tube in the magnetic stand and observe that the magnetic particles form a pellet against the back of the tube. Wait until the size of the pellet stops increasing (approximately 10 min)
17. With the sample DNA tube remaining in the magnetic stand, carefully aspirate and discard all visible liquid phase. Do not aspirate or disturb the magnetic particle pellet
18. Perform steps a through e three times:
    (a) Add prepared wash buffer to the sample DNA tubes:
        - First wash: 600 µL Wash Buffer A
        - Second wash: 300 µL Wash Buffer A
        - Third wash: 300 µL Wash Buffer B
    (b) Cap the sample DNA tubes and remove them from the magnetic stand
    (c) Vortex the sample DNA tubes for 15 s, then centrifuge briefly to collect any residual tube contents from the sides and cap
    (d) Place the sample DNA tubes in the magnetic stand for 1 min
    (e) With the sample DNA tubes remaining in the magnetic stand, carefully aspirate and discard all visible liquid phase. Do not disturb the magnetic particle pellet
19. Centrifuge the tubes briefly, place the tubes back on the magnetic stand for 30–60 s, then collect any residual liquid
20. Add 50 µL of PrepFiler® Elution Buffer to the sample DNA tube, then vortex at maximum speed until the pellet is resuspended
21. Place the sample DNA tube in a thermal shaker, then incubate at 70 °C and 1000 rpm for 10 min
22. Vortex the sample DNA tube at maximum speed until there is no visible magnetic particle pellet on the side of the tube (approximately 2 s), then centrifuge briefly to collect any residual tube contents from the sides and cap
23. Place the sample DNA tube in the magnetic stand, then wait until the size of the pellet at the side of the tube stops increasing (approximately 5 min)
24. Carefully aspirate all visible liquid phase in the sample DNA tube
25. Transfer the eluate to a new, labeled 1.5 mL tube for storage

Notes: The isolated DNA can be stored at 4 °C for up to 1 week, or at − 20 °C for longer storage

Conclusions

Successful individual-specific DNA profiles are routinely obtained from a variety of biological samples in the forensic context. The choice of matrices collected from cadavers for DNA extraction depends mostly on the preservation of the corpse. High-quality DNA profiles can be potentially obtained from any biological source. When dealing with fresh or partially decomposed cadavers, biological fluids, soft tissues, and hair are the samples of choice as a source of DNA, requiring routinely laboratory procedures. Blood and saliva are better sources than urine, due to their

high number of nucleated cells, while DNA extraction from hair is usually only possible when roots and adhering tissues are present.

With ongoing putrefaction, samples may be degraded, which contain inhibitors or low DNA content. Soft tissues like muscle and organs are the source of DNA of choice dealing with partially decomposed cadavers. In cases of advanced decay, nails, which are resistant to decomposition, and hard tissues, such as bones and teeth, are alternative sources of DNA. Genetic identification can even be successfully carried out from aged, burned, and fragmented skeletal remains. The main disadvantages of collecting samples from bones and teeth are the requirement of tedious and time-consuming extraction procedures due to the need of mechanical fragmentation and tissue decalcification. Also, the high number of steps required by these processes increases the risk of contamination.

The Chelex protocol for DNA extraction is among the most used due to its low cost and short execution time. It provides high DNA yield from samples that contain a large number of cells and no PCR inhibitory substances. Organic protocols have always been the preferred extraction method for high quantity and purity of DNA from any biological sample. However, since they require use of toxic chemicals, its use has shifted towards silica-based extractions. Silica-based protocols may provide somewhat less DNA quantity or purity than organic ones, but the difference is not believed to be significant. In addition, silica-based protocols are safer, easier, available in commercial kits and amenable to automation.

Purifying DNA from bones and teeth often requires modification of the DNA extraction techniques utilized for other biological samples. The goal of skeletal DNA extraction techniques is to maximize DNA yield, minimize any additional DNA damage, and remove any inhibitors that may co-purify with the skeletal DNA and interfere with later genetic analyses. In challenging forensic scenarios, the collection of more than one sample from different tissues is advised. Also, if DNA is degraded or in low quantity, multiple extractions and amplifications may be necessary.

Furthermore, since PCR is always required to produce DNA profiles in forensic genetics, the risk of contamination must always be considered. Contamination may be introduced at any stage, from recovery of the remains to the laboratory analysis. Precautions must be adopted when handling challenging/critical samples in order to optimize DNA profiling. Moreover, when the size of the sample is small, the risk of mistyping due to exogenous contamination increases exponentially. With the aid of the protocols provided herein, the practitioner should be able to extract high quality DNA from the most diverse range of samples for human identification and criminalistic purposes.

References

Alaeddini R, Walsh SJ, Abbas A (2010) Forensic implications of genetic analyses from degraded DNA-a review. Forensic Sci Int Genet 4:148–157. https://doi.org/10.1016/j.fsigen.2009.09.007

Allouche M, Hamdoum M, Mangin P, Castella V (2008) Genetic identification of decomposed cadavers using nails as DNA source. Forensic Sci Int Genet 3(1):46–49. https://doi.org/10.1016/j.fsigen.2008.07.008

Burger MF, Song EY, Schumm j. W. (2005) Buccal DNA samples for DNA typing: new collection and processing methods. BioTechniques 39:257–261. https://doi.org/10.2144/05392PF01
Butler JM (2005) Forensic DNA typing. Academic Press, London
Butler JM (2009) Fundamentals of forensic DNA typing. Academic, Cambridge, MA
Butler JM (2011) Advanced topics in forensic DNA typing: methodology. Academic, Cambridge, MA
Caputo M, Bosio LA, Corach D (2011) Long-term room temperature preservation of corpse soft tissue: an approach for tissue sample storage. Investig Genet 2:17. https://doi.org/10.1186/2041-2223-2-17
Cattaneo C, Gibelli D (2017) Identification. In: Houck MM (ed) Forensic anthropology. Academic Press Elsevier Inc., London, pp 307–314
Cattaneo C, De Angelis D, Grandi M (2005a) Mass disasters. In: Schmitt A, Cunha E, Pinheiro J (eds) Forensic anthropology and medicine: complementary sciences from recovery to cause of death. Humana Press Inc., Totowa, pp 431–443
Cattaneo C, De Angelis D, Porta D, Grandi M (2005b) Personal identification of cadavers and human remains. In: Schmitt A, Cunha E, Pinheiro J (eds) Forensic anthropology and medicine: complementary sciences from recovery to cause of death. Humana Press Inc., Totowa, pp 359–379
Cattaneo C, Gelsthorpe K, Sokol RJ (2006) DNA extraction methods in forensic analysis. In: Meyers RA, Brettell TA (eds) Encyclopedia of analytical chemistry: applications, theory and instrumentation. https://doi.org/10.1002/9780470027318.a1104m
Cockle DL, Bell LS (2017) The environmental variables that impact human decomposition in terrestrially exposed contexts within Canada. Sci Justice 57(2):107–117. https://doi.org/10.1016/j.scijus.2016.11.001
Courts C, Sauer E, Hofmann Y, Madea B, Schyma C (2015) Assessment of STR typing success rate in soft tissues from putrefied bodies based on a quantitative grading system for putrefaction. J Forensic Sci 60(4):1016–1021. https://doi.org/10.1111/1556-4029.12746
Dash HR, Shrivastava P, Das S (2020) Principles and practices of DNA analysis: a laboratory manual for forensic DNA typing. Springer, New York
Davoren J, Vanek D, Konjhodzić R, Crews J, Huffine E, Parsons TJ (2007) Highly effective DNA extraction method for nuclear short tandem repeat testing of skeletal remains from mass graves. Croat Med J 48:478–485
Elkins KM (2012) Forensic DNA biology: a laboratory manual. Academic, Cambridge, MA
Fattorini P, Marrubini G, Ricci U, Gerin F, Grignani P, Cigliero SS, Xamin A, Edalucci E, La Marca G, Previderé C (2009) Estimating the integrity of aged DNA samples by CE. Electrophoresis 30(22):3986–3995. https://doi.org/10.1002/elps.200900283
Gebhart FT, Brogdon BG, Zech WD, Thali MJ, Germerott T (2012) Gas at postmortem computed tomography – an evaluation of 73 non-putrefied trauma and non-trauma cases. Forensic Sci Int 222(1-3):162–169. https://doi.org/10.1016/j.forsciint.2012.05.020
Greenspoon SA, Scarpetta MA, Drayton ML, Turek SA (1998) QIAamp spin columns as a method of DNA isolation for forensic casework. J Forensic Sci 43:1024–1030
Hansen J, Lesnikova I, Funder AM, Banner J (2014) DNA and RNA analysis of blood and muscle from bodies with variable postmortem intervals. Forensic Sci Med Pathol 10:322–328. https://doi.org/10.1007/s12024-014-9567-2
Hervella M, Iñiguez MG, Izagirre N, Anta A, De-la-Rúa C (2015) Nondestructive methods for recovery of biological material from human teeth for DNA extraction. J Forensic Sci 60(1):136–141. https://doi.org/10.1111/1556-4029.12568
Higgins D, Austin JJ (2013) Teeth as a source of DNA for forensic identification of human remains: a review. Sci Justice 53(4):433–441. https://doi.org/10.1016/j.scijus.2013.06.001
Hochmeister MN, Budowle B, Borer UV, Eggmann U, Comey CT, Dirnhofer R (1991) Typing of deoxyribonucleic acid (DNA) extracted from compact bone from human remains. J Forensic Sci 36:1649–1661
Kaneko Y, Ohira H, Tsuda Y, Yamada Y (2015) Comparison of hard tissues that are useful for DNA analysis in forensic autopsy. Leg Med 17(6):547–552. https://doi.org/10.1016/j.legalmed.2015.07.001

King C, Debruyne R, Kuch M, Schwarz C, Poinar H (2009) A quantitative approach to detect and overcome PCR inhibition in ancient DNA extracts. BioTechniques 47(5):941–949. https://doi.org/10.2144/000113244

Kitayama T, Fukagawa T, Watahiki H, Mita Y, Fujii K, Unuma K, Sakurada K, Uemura K, Sekiguchi K, Mizuno N (2020) Evaluation of rapid DNA system for buccal swab and disaster victim identification samples. Legal Med 46:101713. https://doi.org/10.1016/j.legalmed.2020.101713

Latham K, Madonna M (2014) DNA survivability in skeletal remains. In: Pokines J, Symes S (eds) Manual of forensic taphonomy. CRC Press, Boca Raton, pp 403–425

Latham KE, Miller JJ (2019) DNA recovery and analysis from skeletal material in modern forensic contexts. Forensic Sci Res 4(1):51–59. https://doi.org/10.1080/20961790.2018.1515594

Lee Goff M (2009) Early post-mortem changes and stages of decomposition in exposed cadavers. Exp Appl Acarol 49(1-2):21–36. https://doi.org/10.1007/s10493-009-9284-9

Loreille O, Diegoli T, Irwin J, Coble M, Parsons T (2007) High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet 1:191–195. https://doi.org/10.1016/j.fsigen.2007.02.006

Madea B, Saukko P, Oliva A, Musshoff F (2010) Molecular pathology in forensic medicine-introduction. Forensic Sci Int 203(1-3):3–14. https://doi.org/10.1016/j.forsciint.2010.07.017

Milos A, Selmanovic A, Smajlovic L, Huel RL, Katzmarzyk C, Rizvic A, Parsons TJ (2007) Success rates of nuclear short tandem repeat typing from different skeletal elements. Croat Med J 48:486–493

Nouma Y, Ben Amar W, Zribi M, Bardaa S, Hammami Z, Maatoug S (2016) Forensic examination after exhumation: contribution and difficulties after more than thirty years of burial. J Forensic Legal Med 44:120–127. https://doi.org/10.1016/j.jflm.2016.10.005

Phillips K, McCallum N, Welch L (2012) A comparison of methods for forensic DNA extraction: Chelex-100® and the QIAGEN DNA investigator kit (manual and automated). Forensic Sci Int Genet 6:282–285. https://doi.org/10.1016/j.fsigen.2011.04.018

Prinz M, Lessig R (2014) Forensic DNA analysis. Handbook of forensic medicine. Wiley, New York

Putkonen MT, Palo JU, Cano JM, Hedman M, Sajantila A (2010) Factors affecting the STR amplification success in poorly preserved bone samples. Investig Genet 1:9. https://doi.org/10.1186/2041-2223-1-9

Rothe J, Nagy M (2016) Comparison of two silica-based extraction methods for DNA isolation from bones. Legal Med. https://doi.org/10.1016/j.legalmed.2016.07.008

Rubio L, Santos I, Gaitan MJ, Martin de las Heras, S. (2013) Time-dependant changes in DNA stability in decomposing teeth over 18 months. Acta Odontol Scand 71:638–643. https://doi.org/10.3109/00016357.2012.700068

Schwark T, Heinrich A, Von Wurmb-Schwark N (2011) Genetic identification of highly putrefied bodies using DNA from soft tissues. Int J Legal Med 125:891–894. https://doi.org/10.1007/s00414-010-0537-2

Sorensen A, Rahman E, Canela C, Gangitano D, Hughes-Stamm S (2016) Preservation and rapid purification of DNA from decomposing human tissue samples. Forensic Sci Int Genet 25:182–190. https://doi.org/10.1016/j.fsigen.2016.05.013

Sutherland A, Myburgh J, Steyn M, Becker PJ (2013) The effect of body size on the rate of decomposition in a temperate region of South Africa. Forensic Sci Int 231(1-3):257–262. https://doi.org/10.1016/j.forsciint.2013.05.035

Swann LM, Forbes SL, Lewis SW (2010) Analytical separations of mammalian decomposition products for forensic science: a review. Anal Chim Acta 682(1-2):9–22. https://doi.org/10.1016/j.aca.2010.09.052

Ubelaker DH, Zarenko KM (2011) Adipocere: what is known after over two centuries of research. Forensic Sci Int 208(1-3):167–172. https://doi.org/10.1016/j.forsciint.2010.11.024

Van den Berge M, Wiskerke D, Gerretsen RRR, Tabak J, Sijen T (2016) DNA and RNA profiling of excavated human remains with varying postmortem intervals. Int J Legal Med 130:1471–1480. https://doi.org/10.1007/s00414-016-1438-9

Vanek D, Budowle B, Dubska-Votrubova J, Ambers A, Frolik J, Pospisek M, Al Afeefi AA, Al Hosani KI, Allen M, Al Naimi KS, Al Salafi D, Al Tayyari WAR, Arguetaa W, Bottinelli M, Bus MM, Cemper-Kiesslich J, Cepil O, De Cock G, Desmyter S, El Amri H, El Ossmani H, Galdies R, Grün S, Guidet F, Hoefges A, Iancu CB, Lotz P, Maresca A, Nagy M, Novotny J, Rachid H, Rothe J, Stenersen S, Stephenson M, Stevanovitch A, Strien J, Sumita DR, Vella J, Zander J (2017) Results of a collaborative study on DNA identification of aged bone samples. Croat Med J 58:203–213. https://doi.org/10.3325/cmj.2017.58.203

Walsh PS, Metzger DA, Higuchi R (1991) Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10:506–513

Zhou C, Byard RW (2011) Factors and processes causing accelerated decomposition in human cadavers – an overview. J Forensic Legal Med 18:6–9. https://doi.org/10.1016/j.jflm.2010.10.003

DNA isolation from human remains

34

María Saiz, Christian Haarkötter, X. Gálvez, L. J. Martinez-Gonzalez, M. I. Medina-Lozano, and Juan Carlos Alvarez

Contents

Abstract

The DNA analysis of human remains can be challenging despite the strong matrix in bones and teeth that helps preserve DNA. To isolate DNA, the correct procedures need to be applied. Factors such as temperature, pH, and humidity affect DNA degradation, while polymerase chain reaction inhibitors can affect or even prevent DNA amplification. If a sufficient quantity and quality of genetic material is obtained during DNA extraction, the key stage in DNA typing, a genetic profile can be obtained, thereby helping to identify the human remains. The aim of this chapter is to show various pretreatment strategies for bones and teeth (surface washing, chemical washing, enzymatic predigestion, milling and sanding, and

M. Saiz · C. Haarkötter (✉) · X. Gálvez · M. I. Medina-Lozano · J. C. Alvarez
Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain
e-mail: chaarkotter@ugr.es

L. J. Martinez-Gonzalez
Department of Biochemistry and Molecular Biology III, Faculty of Medicine, University of Granada, Granada, Spain

Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), Granada, Spain

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_38

ultraviolet radiation), as well as pulverization methods (mortaring, freezer milling, or tissue lysis), manual DNA isolation protocols (total demineralization, organic use of Chelex resin, and manual purification), and available commercial kits for DNA extraction from human remains.

Keywords

Ancient DNA · Automated DNA extraction · Human remains · Manual DNA extraction · Sample pretreatment

Introduction

When discussing human remains, we usually refer to corpses or skeletons, and samples are most frequently found in bones and teeth. Both types of samples protect DNA from degradation and biological processes due to their physical and chemical robustness. However, accessing the DNA is not as easy as in other tissues due to these protective characteristics.

Bone tissue consists primarily of proteins (mainly collagen and osteocalcin) and minerals. Approximately 70% of the mineral component of bone comprises hydroxyapatite, which includes calcium phosphate, calcium carbonate, calcium fluoride, calcium hydroxide, and citrate. The DNA in bones is located in the osteoblasts, osteocytes, and osteoclasts. Osteons are the functional unit of bones and include osteocytes (located in spaces within the dense bone matrix called lacunae) and haversian canals, which contain blood vessels and nerves and are formed by concentric layers called lamellae. This structure favors the deposit and storage of mineral salts, which gives bone tissue its strength. Osteoblasts produce the organic components of the bone matrix and are situated at the surface of the bone matrix. Osteoclasts are responsible for bone remodeling and resorption during bone growth and are located on the surface of the bone matrix (Mescher 2018b) (Fig. 1).

Teeth consist of dentin, a calcified material harder than bone that forms a large part of the structural axis of the tooth and surrounds the internal pulp cavity. Dentin consists of 70% hydroxyapatite, type I collagen, and proteoglycans. Dentin in the dental crown is covered by enamel, an extremely mineralized, hard, acellular, avascular tissue. Enamel is the hardest component of the human body and consists of 96% calcium hydroxyapatite, very few proteins, and no collagen. The dentin at the tooth root is covered by cementum, another type of calcified connective tissue that resembles bone. The soft tissue in the dental pulp is highly vascular and innervated and consists of odontoblasts, fibroblasts, endothelial cells, peripheral nerves, undifferentiated mesenchymal cells, and other nucleated blood components, making it rich in DNA (Muruganandhan and Sivakumar 2011; Mescher 2018a). DNA is recovered from the tooth pulp, where it is abundant and unlikely to be contaminated by nonhuman DNA (Girish et al. 2010) (Fig. 2).

Bone and tooth tissues are compact and hard structures that preserve the DNA in their matrix. Isolating DNA from bones and teeth therefore requires several pretreatment steps before the DNA can be recovered from these cells.

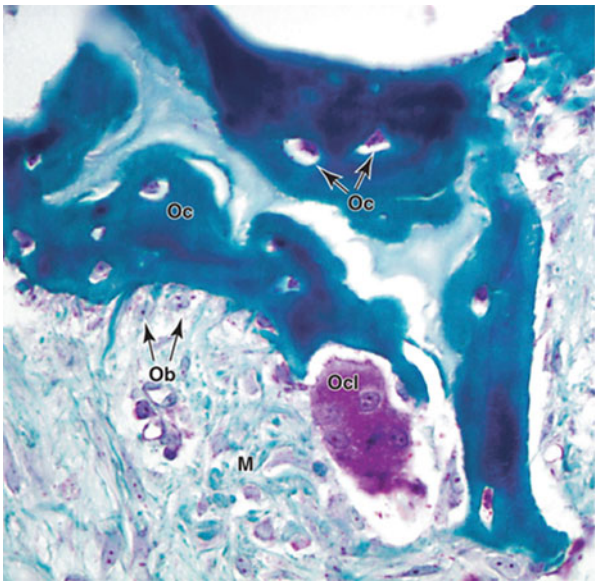


Fig. 1 Bone matrix. (M) Mesenchymal regions, (Ob) osteoblasts, (Oc) osteocytes, and (Ocl) osteoclasts (Mescher 2018b)

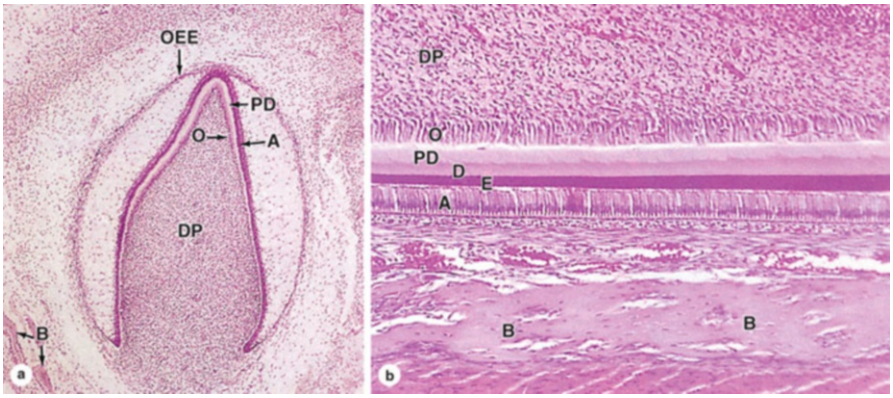


Fig. 2 Tooth formation. (A) Ameloblasts, (B) bone, (D) dentin, (DP) dental papila, (E) enamel, (PD) predentin, (O) odontoblasts, and (OEE) outer enamel epithelium (Mescher 2018a)

There are three main problems to solve at this stage: the introduction of modern DNA, the presence of too few DNA molecules to serve as templates for polymerase chain reaction (PCR) testing, and the co-purification of inhibitory substances that result in false negatives. The isolation of ancient DNA therefore has three requirements: 1) the samples have to be pretreated to reduce contamination, 2) extraction techniques that do not damage the DNA need to be employed, and 3) these techniques need to have a high purification power to reduce or eliminate the presence of inhibitors.

Sample Pretreatment

Decontamination

Bone and tooth samples are typically obtained from excavations and are often improperly and carelessly handled. There is, however, a growing awareness among archaeologists and anthropologists regarding the importance of wearing protective clothes when working with ancient DNA samples. Conducting a pretreatment stage prior to DNA isolation is extremely important for eliminating any possible contamination by exogenous DNA and the remains of putrilage and impurities.

There are several strategies for reducing or eliminating the possible superficial contamination of bones and teeth such as surface washing, acid washing, highly concentrated ethanol washing, bleach washing, hydrogen peroxide washing, milling and grinding, ultraviolet irradiation, and sampling of the inner part of the compact bone and combined techniques (Kemp and Smith 2005).

Surface Washing

Vigorous surface washing can remove the external layer of exogenous material from the bone (Holland et al. 1993) and involves using sterile water prior to cutting (Merriwether et al. 1994) or repeated rinsing of the cut pieces in distilled water, with a final air dry step (Alonso et al. 2001). However, humidity has been widely reported as a factor in damaging DNA because it facilitates mineral dissolution and increases hydrolytic damage. Moreover, the interdependence between the organic and mineral components of bone supports the hypothesis of bone susceptibility to chemical and biological effects due to the increase in porosity (Emmons et al. 2020).

Sodium Hypochlorite Washing

For forensic and ancient DNA samples, one of the most common methods for eliminating exogenous DNA from bone and teeth is washing the surface or even the powdered bone or tooth with bleach. Although washing significantly reduces exogenous human DNA, it also results in a loss of endogenous DNA (Dabney and Meyer 2019). Sodium hypochlorite, an active component of bleach, rapidly attacks nucleic acids in a nonspecific manner, degrading purine and pyrimidine bases through oxidation reactions such as chlorination (Hayatsu et al. 1971).

Given the destructive nature of bleach, the possibility of replacing it with other compounds such as phosphate buffer has been explored. The use of phosphate buffer is based on the competition between free phosphate ions and DNA phosphate groups attached to hydroxyapatite. Phosphate buffer has been shown to be less aggressive toward endogenous DNA than bleach washing and, although it can eliminate some microbial DNA, it is less effective with exogenous DNA (Dabney and Meyer 2019).

Ethanol Washing

An alternative to bleach washing is the cleaning of bone and teeth surfaces with sterile cotton previously moistened with a 95% (Fisher et al. 1993) or 70% ethanol solution (Stone and Stoneking 1998) in ultrapure water. Although ethanol exerts no degradative

activity on DNA, its use for precipitating DNA is well known, and its usefulness for cleaning surfaces in forensic laboratories has long been accepted (Kampmann et al. 2017).

Acid Washing

Similar to bleach and ethanol washing, the use of weak acids to remove exogenous DNA from samples is also common, as 30% acetic acid diluted in ultrapure water (Montiel et al. 2001). This technique is based on the power of denaturation of a low pH solution hydrolyzing the glycosidic and phosphodiester bonds of DNA. However, the effectiveness of this solution is lower if the DNA is rich in guanine and cytosine (Shapiro et al. 1978).

Hydrogen Peroxide Washing

Immersing human remains in a 3% (Ginther et al. 1992) or 3–30% hydrogen peroxide solution for 10–30 min (Merriwether et al. 1994) has also been shown to be an effective method for removing the exogenous component of samples. The decomposition of hydrogen peroxide into water and free oxygen radicals causes oxidative damage to DNA by radical-ionic mechanisms (Mouret et al. 1991).

Enzymatic Predigestion

One of the main problems with chemical decontamination methods is that they also attack the sample's endogenous DNA. A less aggressive method is the predigestion of the samples with ethylenediaminetetraacetic acid (EDTA) lysis buffer and proteinase K, which significantly reduces the contaminating DNA without affecting the sample's endogenous genetic material (Schroeder et al. 2019).

Milling and Sanding

The pretreatment technique of milling and sanding the surface of human remains consists of applying mechanical abrasion to their outermost part to eliminate the adhering exogeneous material. There are various approaches with these techniques such as air abrasion with 100 μm aluminium oxide particles (Richards and Sykes 1995), the use of sandpaper discs (Kalmár et al. 2000), scraping with a sterile scalpel (Lalueza-Fox et al. 2001), and the increasingly widespread use of precision rotary tools such as those manufactured by Dremel® (Gaudio et al. 2019).

The main problem with milling and sanding pretreatment is the formation of bone dust, which can contaminate the working surface, tools, other samples, and even the operator. The Laboratory of Genetic Identification of the University of Granada (Spain) developed a milling, sanding, and cutting methacrylate enclosure with a removable lid and two lateral holes to insert the operator's hands and the milling/sanding tool. The bottom of the enclosure is covered with filter paper, and the entire milling and sanding operation is conducted in a fume hood (Álvarez et al. 2001). The enclosure not only prevents the clogging of the filters of the hoods or cabinets in which the operation is performed but also facilitates the cleaning of the equipment. The external and internal surfaces of bones can be milled and grinded, and the bone can then be cut into fragments, being 0.5–1 cm^2 fragments recommended (Fig. 3).

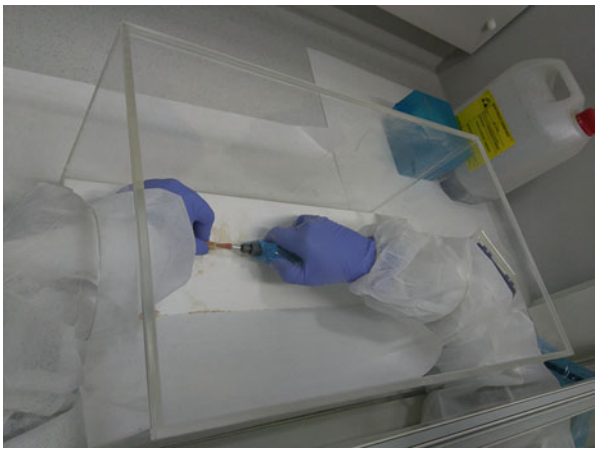


Fig. 3 Methacrylate enclosure

Inner Part of the Compact Bone Sampling

To minimize the risk of contamination with modern DNA, a 2×2 cm section of compact bone from long bones, such as the posterior femoral diaphysis, can be cut, followed by scraping of the inner and outer surfaces with a scalpel (Palmirotta et al. 1997), leaving only the innermost part of the compact bone, which might be free of or less exposed to exogenous DNA.

Ultraviolet Irradiation

Short-wave ultraviolet (UV) light (254 nm) induces the covalent bonding of thymine bases, preventing denaturalization of DNA double strands during PCR, making it inaccessible to polymerase during amplification. Therefore, exposing the surface of the samples to UV light for a few minutes is useful for eliminating exogenous DNA (Latham and Miller 2019).

Variations in the UV exposure time, ranging from 10 min to up to 2 days (Carlyle et al. 2000; Kalmár et al. 2000; Matheson and Loy 2001), have been described and are sufficient to affect exogenous DNA but not endogenous DNA. The distance between the irradiation source and the irradiated surface is also important: the closer the source is to surface, the greater the irradiation power (Champlot et al. 2010; Hall et al. 2014). However, 10 min for both sides of the sample is the most frequently employed exposure time.

Combination of Techniques

Many laboratories combine two or more of the previous decontamination protocols to reduce potential exogenous contamination. The Laboratory of Genetic Identification of the University of Granada first mills, sands, and cuts the bone samples into pieces measuring approximately 0.5 cm^2 and then irradiates them with UV for 10 min on both sides before pulverizing the samples.

Pulverization

After decontaminating the sample, each laboratory employs its own extraction method, processing the bone and tooth samples in small pieces or grinding them into powder through various processes.

Manual Mortar

Manual mortars and pestles have typically been employed to grind bones and teeth. Numerous laboratories still use Teflon pestles in ceramic mortars to generate the fine bone powder needed for extractions (Cafiero et al. 2019).

Freezer Mill

Freezer/Mill® cryogenic grinders are widely used in laboratories to grind samples such as teeth, bones, and other animal and human tissues. Samples are placed in a sealed cryogenic grinding vial in the grinder and then immersed in liquid nitrogen. The samples are cooled to cryogenic temperatures and then pulverized by magnetically shuttling a steel impactor back and forth against two stationary end plugs (Fig. 4).

TissueLyser

To grind and disintegrate the tissues, a TissueLyser II system is recommended. The TissueLyser II grinds bone and teeth samples by shaking them with metallic balls without requiring liquid nitrogen. However, liquid nitrogen can be used with this technique to prevent the samples from heating, thereby facilitating the grinding (Fig. 5).

Fig. 4 Grinding vials and Freezer/Mill ®

Fig. 5 Grinding vials and TissueLyser II sytem

DNA Isolation

DNA isolation is the most important stage in the DNA analysis process, because it will determine the outcome of the entire process. If there is insufficient starting DNA, amplification will fail, yielding no results. Therefore, the key is to obtain as much DNA as possible. Molecules of genetic material have to be isolated from other cell components before the genetic material can be analyzed, because the cell proteins that package and protect the DNA can inhibit the analysis (Butler 2005). The other major problem is the presence of inhibitors that need be eliminated or minimized (Barrio-Caballero 2012), because they either inactivate DNA polymerase or compete with other components of the DNA synthesis reaction. The presence of extrinsic substances from bone such as humic and fulvic acids from the soil and intrinsic substances such as calcium have to been eliminated (Eilbert and Foran 2009).

Manual DNA Isolation

Total Demineralization

The total demineralization method was first developed in 1991 with well-preserved animal and human bones from archaeological sites. The method employs 0.5 M EDTA, proteinase K, and N-lauroylsarcosine at 37 °C for 18–24 h, followed by an extraction with phenol-chloroform. Bone decalcification is made possible by incubating cut samples with EDTA for 72 h, despite this approach reducing the total DNA yield (Hagelberg and Clegg 1991). The basis for this method is the chelating activity of EDTA, which binds to iron and calcium ions. The problem of reduced DNA yield is solved by making EDTA part of the lysis buffer, so that DNA can be purified with phenol-chloroform-isoamyl alcohol in a 25:24:1 proportion after an overnight lysis. The resultant is then concentrated with centrifugal filter units (Edson et al. 2004). A hybrid protocol of these two methods can be used by incubating

0.6–1.2 g of bone powder with 15 mL 0.5 M EDTA, 1% lauroyl-sarcosinate, and 20 mg/μL of proteinase K in a rotatory shaker overnight. The bone powder is thereby completely dissolved. Organic extraction with phenol-chloroform, filtration with centrifugal filter units, and two washes with ultrapure water are then performed (Loreille et al. 2007).

Organic Extraction

The well-known phenol-chloroform-isoamyl alcohol method, also known as organic extraction, was developed in 1991. Using this approach, 5 g of bone are powdered and decalcified with 0.5 M EDTA for 3–5 days, washed three times with ultrapure water, lysed with proteinase K and extraction buffer for 2 h, extracted with phenol-chloroform-isoamyl three times, and concentrated and purified in a Centricon centrifugal filter (Hochmeister et al. 1991). The mixture is typically stabilized with 10 mM of Tris-EDTA, the proteins are unfolded with phenol, empowered augmented by chloroform. The chloroform also denatures lipids, while the isoamyl alcohol stabilizes the interphase and increases DNA purity. DNA will be trapped in the upper aqueous phase.

Modified protocols introduce sodium dodecyl sulphate (SDS) and dithiothreitol (DTT) into the lysis buffer (Ferreira et al. 2013). SDS is an anionic detergent that linearizes the proteins present in chromatin, while DTT is a reducing agent that reduces the disulphide bonds present in proteins.

Although the phenol-chloroform method yields a large amount of DNA, the main issue is that it is a dangerous reagent, both for the analyst and the environment. The solution should therefore be used in a fume hood and its residues properly treated and disposed of. The method is also time-consuming and requires significant hands-on time.

Chelex® Resin

In 1998, a simple, chelating, single-tube, resin-based procedure, using minimal steps was suggested. The Chelex®-100 (Bio-Rad Laboratories, Hercules, CA) is a chelating resin composed of styrene divinylbenzene copolymers and iminodiacetate ions that bind to polyvalent metal ions. In the basic procedure, samples are boiled in a 5% Chelex®-100 suspension (Willard et al. 1998). A prior 30 min incubation at 56 °C is recommended for bone samples. An adapted protocol for ancient bone samples starts with a 3 h incubation at 56 °C of 100 mg of bone powder in Chelex®-100, followed by a 20 min boiling period (Coulson-Thomas et al. 2015). To yield more DNA, proteinase K can be added prior to incubation (Tsuchimochi et al. 2002). Despite being fast and environmentally friendly, Chelex® cannot remove PCR inhibitors.

Manual Purification of DNA Extracts

To maximize the chances of success, DNA extraction protocols need to obtain the largest amount of target DNA possible while reducing or even eliminating the presence of PCR inhibitors. To this end, there are two classical methods for analyzing ancient DNA (Yang et al. 1998): centrifugal filter units and silica particle

columns. The first method uses Centri-con™ filters (Hagelberg and Clegg 1991) that consist of an anisotropic membrane that retains macrosolutes, such as DNA, while letting low-molecular-weight compounds pass through, which can also occur with PCR inhibitors. Other protocols include further washing with approximately 1 mL of distilled water or 2 mL of TE buffer with 0.01 M Tris and 0.001 M EDTA at a pH of 7.5 (Hochmeister et al. 1991). The second method uses silica particles (Höss and Pääbo 1993) with a high binding capacity for DNA molecules and are therefore retained while the inhibitors are washed out; however, the silica particles are themselves potential PCR inhibitors. The silica pellets are therefore washed twice with a 10 M guanidine thiocyanate and 0.1 M Tris-HCl buffer at a pH of 6.4, washed twice with 70% ethanol, and washed once with acetone. Another approach is to precipitate out any material that is nonnucleic by adding saturated sodium acetate (a process known as salting-out [Cattaneo et al. 1995]), adding 1 mL of the solution to the tube, shaking it manually for 30 s and centrifuging it for 10 min at 4000 g.

Several DNA purification commercial kits are available, such us DNA IQ™ System purification (Promega, MA, USA), or QIAquick™ PCR purification kit (QIAGEN, Hilde, Germany) (Ye et al. 2004).

DNA Isolation with Commercial Kits

There are currently numerous commercial isolation kits for bone DNA analysis (see Table 1), although most require mechanical pretreatment. Most of these kits can be automated with the appropriate equipment, which offers several advantages such as maintained sample integrity, increased reproducibility, constant performance, greater throughput, workflow integration, electronic audits, compatibility with laboratory information management systems (LIMS), sample switching and data entry error minimization, reduced hands-on time, and lower repetitive stress injuries (Lee and Shewale 2017). In this section we will describe some of these advantages.

Table 1 Commercial kits for bone DNA analysis

Kit	Format	Lysis time	Protocol time	Automatization
PrepFiler™ BTA (ThermoFisher)	100 reactions	~2 h	~2–3 h	AutoMate™ Express
Bone DNA Extraction (Promega)	100 reactions	~2,5 h + ~2,5 h	~2 h	Maxwell®
QIAamp DNA Investigator (Qiagen)	50 reactions	Overnight	~30-60 min	QIAcube Connect
EZ1 Investigator (Qiagen)	48 reactions	~24–48 h	~20 min	EZ1
Cells and Tissue DNA Isolation (Norgen Biotek)	50 reactions	~2 h	~60 min	–
CrimePrep (Ademtech)	96/48 reactions	~2 h	~60 min	Automag

PrepFiler ™ BTA Forensic DNA Extraction Kit

PrepFiler™ BTA Forensic DNA Extraction Kit (Thermo Fisher Scientific Inc., MA, USA) was developed for isolating DNA from bone, teeth, and other forensic samples with adhesives (cigarette butts, envelope flaps, tape lifts, and chewing gum). The extraction kit uses a format that provides for 100 reactions, and the protocol can be performed in approximately 2–3 h, which reduces the processing time by requiring a shorter lysis time than standard methods. The use of phenol-chloroform with this kit is not necessary. To release DNA from calcified tissues, the kit employs a sequence of washes, with various buffers and filter columns. DNA isolation is performed with a magnetic bead. The protocol is divided into four parts: lysis (PrepFiler® BTA lysis buffer and DTT), DNA binding (PrepFiler® magnetic beads), purification (PrepFiler® wash buffer), and DNA reconstitution (PrepFiler® BTA lysis buffer). The PrepFiler® lysis buffer is composed of a thiocyanic acid compound with guanidine (1:1), while the PrepFiler® BTA lysis buffer is based on sodium hydroxide. There are several variations to the method according to the samples' complexity, ranging from increasing the quantity of powdered sample (and thus the volume of lysis reagents) to extending the lysis time to overnight. The elution volume can also be customized to concentrate the DNA extract.

There is an automated option for these kits: the AutoMate Express™ (Thermo Fisher Scientific Inc., MA, USA), which is based on the above protocol. First, lysis is performed in a thermoshaker and then the lysate is automatically purified. The AutoMate Express™ uses prefilled buffer cartridges that reduce the handling of samples, thereby reducing potential contamination by the operator (Applied Biosystems 2012).

Bone DNA Extraction Kit

The Bone DNA Extraction Kit (Promega Corporation, WI, USA) is actually the joining of two protocols: a preprocessing protocol and a subsequent purification protocol. The kit was developed as a combination of classical purification protocols, created by various genetic identification laboratories, and uses a demineralization buffer (0.5 M EDTA, pH 8.0, 1% lauroylsarcosine) and an organic extraction protocol (proteinase K and 1-thioglycerol) to effectively and efficiently extract DNA from the calcium matrix. The first step can be performed using manual or automated Promega methods (using the Maxwell® extraction instrument), using DNA IQ™ for DNA purification. The kit's format provides for 100 reactions, and the estimated time to completion is more than 7 h due to the demineralization and subsequent digestion requiring an incubation time of 2.5 h each. Performing an extraction within a single working day is therefore problematic (Promega 2019).

QIAamp DNA Investigator Kit

As with the Promega protocols, the Qiagen protocol (Qiagen N.V. Hilden, Germany) has two differentiated phases: a pretreatment protocol for bones and teeth (which describes the decalcification and lysis using EDTA and proteinase K) and a subsequent DNA purification using MinElute spin columns (QIAamp DNA Investigator Kit) to obtain purified genomic DNA. The success of this purification phase depends

on the combination of the selective binding properties of a silica-based membrane. The DNA purification can be automated using QIAcube Connect, an instrument widely used in genetic identification laboratories to fully automate the purification of nucleic acids and proteins.

EZ1 DNA Investigator Kit

Qiagen developed the EZ1 DNA Investigator kit, which uses a similar DNA extraction protocol for powdered bone and tooth to that of the PrepFiler™ BTA Forensic DNA Extraction Kit. The Qiagen protocol is based on a lysis phase performed manually and an automated purification phase in the EZ1 Advanced automated sample preparation system. The lysis requires a decalcification step with 0.5 M EDTA (not included in the kit) for 24–48 h and digestion with proteinase K for 3 h. The lysate is then divided into various aliquots, and buffer MTL is added to load the sample into the device. The protocol supports three different quantities of powered bone or tooth, which can vary the volume of reagents and the protocol on the device. The automated system uses prefilled buffer cartridges, which reduces the handling of samples and thus potential contamination by the operator (QIAGEN 2013).

Cells and Tissue DNA Isolation Kit

The Cells and Tissue DNA Isolation Kit (Norgen Biotek Corp., Ontario, Canada) employs a protocol that purifies DNA from various tissue types using a magnetic bead system. The manufacturer recommends a decalcification step prior to isolating genomic DNA to improve the efficiency of the DNA recovery. To perform decalcification, the bone or teeth are crushed, incubated with EDTA at 4 °C for 24 h, and centrifuged several times. The supernatant is then removed, and 20 mg of the sample is used as the substrate to perform the kit's protocol. The kit's format provides for 50 reactions, and the estimated hands-on time is 1 h and more than 24 h for the incubations (Norgen Biotek Corp 2015).

Crime Prep Adem-Kits for Casework

Ademtech (Pessac, France) commercial kits are based on calibrated particles with high magnetic content and controlled surfaces, specially designed for molecular biology. The Crime Prep protocol starts with 100 mg of bone powder and a 2h lysis, followed by binding, washing, drying, and eluting. Alternative protocols employ an overnight lysis. Crime Prep comes in a 96-sample format and an estimated hands-on time of 1 h plus 2 h of lysis. The process can also be automated with an Automag device for 48 samples with a reduced time cost (Ademtech 2019).

Incorporation of Manual Pretreatment Protocols to Commercial Kits

Over the years, various DNA extraction protocols have been developed to extract DNA from skeletal remains by incorporating manual pretreatment to the commercial kits. This section discusses the protocols capable of extracting DNA without requiring that the samples be pulverized. In 2007, a new DNA extraction procedure that did not require pulverization of samples was described (Kitayama et al. 2010). In their article, the authors presented a new experimental kit that combined a

conventional phenol-chloroform DNA extraction procedure with the QIAamp DNA Mini Kit for DNA isolation kit. In this protocol, mechanical grinding was replaced with gentle stirring during overnight incubation. The results were inconclusive due to the low number of samples and differences in the quality of the extracted DNA with respect to grinding protocols. However, there is certain value in exploring protocols that do not require the pulverization of samples.

In 2020, De Donno et al. described a DNA isolation from a saponified sternum from a limbless human body recovered at sea. The authors extracted DNA using a Macherey-Nagel kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany), modifying the usual procedure described for NucleoSpin® DNA Trace Kit 2. The authors made numerous modifications to the original protocol, halving the quantity of bone material, volumes of reactant, proteinase K, and lysis buffer (B3 buffer, included in the NucleoSpin DNA Trace bones buffer set). The authors also changed the number and type of columns used for binding DNA to the silica membrane (Piglionica et al. 2012; De Donno et al. 2020).

Cartozzo et al. described a similar DNA isolation method for waterlogged bones. The extraction was based on an organic isolation method followed by the use of a Thermo Fisher kit. The extraction began with digestion with proteinase K, purification with phenol-chloroform-isoamyl alcohol and subsequent drying of the aqueous phase using a Speed Vac Concentrator. The dried pellet was reconstituted with deionized water. This eluate was the substrate for starting the DNA extraction using the ChargeSwitch® gDNA Plant Kit (Thermo Fisher Scientific Inc., MA, USA). The protocol was performed according to the manufacturer's procedures (Pagan et al. 2012; Cartozzo et al. 2018). After the experiments, the authors concluded that the magnetic bead technology of the ChargeSwitch® gDNA Plant Kit might be the most efficient method for recovering DNA from waterlogged bones, a surprising statement after using a kit recommended for fungi and plants.

In any case, each bone presents its own set of challenges, requiring manual procedures and commercial kits to be adapted to ensure the success of the DNA extraction. Embalmed bones, for example, not only involve issues with extracting DNA from bones but also bring to the table the exposure to various compounds such as glutaraldehyde and formalin, which can induce molecular cross-linking. In these cases, modifying the existing grinding techniques and combining them with decalcification buffers, phenol-chloroform treatment, and commercial kits will produce efficient methods for extracting sufficient high-quality DNA (Gielda and Rigg 2017).

Conclusions

In conclusion, this review shows how to overcome the drawbacks of isolating DNA from mineralized tissues, in order to identify them through forensic genetics. Bone and tooth tissue consist primarily of proteins and minerals, which are a major inconvenience in the laboratory; however, these tissues protect against degradation of the large DNA molecules. Several techniques aimed at preventing the introduction

of exogenous DNA into the study samples have been reported. These techniques reduce or prevent the loss of the scarce DNA molecules and also reduce or eliminate the co-purification of inhibitory substances.

The complications caused by the time elapsed between death and the laboratory procedures are increased by the characteristics of where and how the body was found. The success of DNA extraction and isolation is affected by variables such as relative humidity, temperature, UV light exposure, and microbiome (amount and type of microorganisms). These factors significantly affect the degradation of the cadavers and their skeletal remains.

These variables that increase degradation and alter the mineral concentration make each bone an enigma. There is therefore no single solution for extracting DNA from bones. Although there are many valid solutions, several of which have been covered in this chapter, there is no ideal protocol for extracting DNA from bones and teeth, as this will depend on the circumstances surrounding each sample.

The most advisable strategy is to use more than one extraction method. There are protocols that eliminate all contaminants and inhibitors from the sample. Due to purification, however, there is an excess of DNA loss, and the final concentration obtained is therefore low. Other methods that obtain more DNA can contain mineral remains or DNA from other organisms. As described in this chapter, the most effective approach is to use different protocols depending on the origin of each bone or, as numerous authors have done, combine stages from different procedures. Nevertheless, this approach should only be taken with a deep understanding of each step and the reactions in each stage. This is the only way to successfully obtain extracted DNA without an excessive number of attempts.

To ensure success, certain tests should be performed before extracting DNA from bones belonging to the same mass grave, same burial type, or similar types of catastrophes. If the skeleton is highly valuable and its identification is imperative, it is highly recommended that the sample be fragmented into 1- to 2-g pieces, so that more than one test can be performed. Problems occurring during the first DNA extraction can thereby be solved in subsequent extractions.

Lastly, the use of commercial kits for the last stages of DNA extraction or for the purification of the isolated DNA is recommended. This approach is very common in laboratories and has been described by numerous authors in the literature. These kits become highly recommended due to their capacity for preventing PCR inhibition and obtaining genetic profiles, which is the ultimate goal in identifying victims.

References

Ademtech (2019) Crime Prep Adem-Kits for casework. In: Forensic casework. https://www.ademtech.com/forensic/casework/. Accessed 2 Jun 2021

Alonso A, Anđelinoviæ Š, Martín P et al (2001) DNA typing from skeletal remains: evaluation of multiplex and megaplex STR systems on DNA isolated from bone and teeth samples. Croat Med J 42:260–266

Álvarez JC, Entrala C, Lorente JA et al (2001) Análisis de ADN mitocondrial. Forénsica 1:40–59

Applied Biosystems (2012) PrepFiler® and PrepFiler® BTA forensic DNA extraction kits. In: Quick ref. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_099065.pdf. Accessed 6 Feb 2021
Barrio-Caballero PA (2012) Revisión de métodos de extracción de ADN a partir de restos óseos en el laboratorio forense. Rev Española Med Leg 39:54–62
Butler JM (2005) Forensic DNA typing, second. Elsevier, Burlington
Cafiero C, Re A, Stigliano E et al (2019) Optimization of DNA extraction from dental remains. Electrophoresis 40:1820–1823. https://doi.org/10.1002/elps.201900142
Carlyle SW, Parr RL, Hayes MG, O'Rourke DH (2000) Context of maternal lineages in the greater southwest. Am J Phys Anthropol 113:85–101
Cartozzo C, Singh B, Boone E, Simmons T (2018) Evaluation of dna extraction methods from waterlogged bones: a pilot study. J Forensic Sci 63:1830–1835. https://doi.org/10.1111/1556-4029.13792
Cattaneo C, Smillie DM, Gelsthorpe K et al (1995) A simple method for extracting DNA from old skeletal material. Forensic Sci Int 74:167–174
Champlot S, Berthelot C, Pruvost M et al (2010) An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PLoS One 5:e13042. https://doi.org/10.1371/journal.pone.0013042
Coulson-Thomas YM, Norton AL, Coulson-Thomas VJ et al (2015) DNA and bone structure preservation in medieval human skeletons. Forensic Sci Int 251:186–194
Dabney J, Meyer M (2019) Extraction of highly degraded DNA from ancient bones and teeth. In: Shapiro A, Barlow A, Heintzman P et al (eds) Ancient DNA methods in molecular biology. Humana Press, New York, p 1963
De Donno A, Mele F, Baldassarra SL et al (2020) Dna extraction from sternum bone for identification of a saponified body: use of a modified protocol. Anthropol Anzeiger 77:235–242. https://doi.org/10.1127/anthranz/2020/1162
Edson SM, Ross JP, Coble MD et al (2004) Naming the dead — confronting the realities of rapid identification of degraded skeletal remains. Forensic Sci Rev 16:64–89
Eilbert KD, Foran DR (2009) Polymerase resistance to polymerase chain reaction inhibitors in bone. J Forensic Sci 54:1001–1007
Emmons AL, Mundorff AZ, Keenan SW et al (2020) Characterizing the postmortem human bone microbiome from surface-decomposed remains. PLoS One 15:e0218636. https://doi.org/10.1371/journal.pone.0218636
Ferreira ST, Paula KA, via Maia FA, Moraes AV (2013) A comparative study between two protocols for DNA extraction from bones. Forensic Sci Int Genet Suppl Ser 4:e374–e375. https://doi.org/10.1016/j.fsigss.2013.10.190
Fisher DL, Holland MM, Mitchell L et al (1993) Extraction, evaluation, and amplification of DNA from decalcified and undecalcified United States civil war bone. J Forensic Sci 38:60–68
Gaudio D, Fernandes DM, Schmidt R et al (2019) Genome-wide DNA from degraded petrous bones and the assessment of sex and probable geographic origins of forensic cases. Sci Rep 9:8226
Gielda L, Rigg S (2017) Extraction of amplifiable DNA from embalmed human cadaver tissue. BMC Res Notes 10:737. https://doi.org/10.1186/s13104-017-3066-y
Ginther C, Issel-Tarver L, King M-CC (1992) Identifying individuals by sequencing mitochondrial DNA from teeth. Nat Genet 2:135–138
Girish K, Rahman FS, Tippu SR (2010) Dental DNA fingerprinting in identification of human remains. J Forensic Dent Sci 2:63–68
Hagelberg E, Clegg JB (1991) Isolation and characterization of DNA from archaeological bone. Proc Biol Sci 244:45–50. https://doi.org/10.1098/rspb.1991.0049
Hall A, Sims LM, Ballantyne J (2014) Assessment of DNA damage induced by terrestrial UV irradiation of dried bloodstains: Forensic implications. Forensic Sci Int Genet 8:24–32
Hayatsu H, Kung PS, Ukita T (1971) Reaction of sodium hypochlorite with nucleic acids and their constituents. Chem Pharm Bull 19:2189–2192. https://doi.org/10.1248/cpb.19.2189
Hochmeister MN, Budowle B, Borer UV et al (1991) Typing of Deoxyribonucleic Acid (DNA) extracted from compact bone from human remains. J Forensic Sci 36:1649–1661. https://doi.org/10.1520/JFS13189J

Holland MM, Fisher DLDL, Mitchel LG et al (1993) Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. J Forensic Sci JFSCA 38:542–553

Höss M, Pääbo S (1993) DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic Acids Res 21:3913–3914

Kalmár T, Bachrati CZ, Marcsik A, Raskó I (2000) A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. Nucleic Acids Res 28:e67

Kampmann M-L, Børsting C, Morling N (2017) Decrease DNA contamination in the laboratories. Forensic Sci Int Genet Suppl Ser 6:e577–e578

Kemp BM, Smith DG (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. Forensic Sci Int 154:53–61

Kitayama T, Ogawa Y, Fujii K et al (2010) Evaluation of a new experimental kit for the extraction of DNA from bones and teeth using a non-powder method. Leg Med 12:84–89. https://doi.org/10.1016/j.legalmed.2009.12.004

Lalueza-Fox C, Luna Calderón F, Calafell F et al (2001) MtDNA from extinct Tainos and the peopling of the Caribbean. Ann Hum Genet 65:137–151

Latham KE, Miller JJ (2019) DNA recovery and analysis from skeletal material in modern forensic contexts. Forensic Sci Res 4:51–59

Lee SB, Shewale JG (2017) DNA extraction methods in forensic analysis. In: Meyers RA (ed) Encyclopedia of analytical chemistry, pp 2–11

Loreille OM, Diegoli TM, Irwin JA et al (2007) High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet 1:191–195

Matheson CD, Loy TH (2001) Genetic sex identification of 9400-year-old human skull samples from Cayönü Tepesi, Turkey. J Archaeol Sci 28:569–575

Merriwether DA, Rothhammer F, Ferrell RE, Ferrel RE (1994) Genetic variation in the New World: ancient teeth, bone, and tissue as sources of DNA. Experientia 50:592–601. https://doi.org/10.1007/BF01921730

Mescher AL (2018a) Digestive tract. In: Junqueira's basic histology: text and atlas, 15e. McGraw-Hill Education, New York, NY

Mescher ALL (2018b) Bone. In: Junqueira's basic histology, 15e Ed. McGraw-Hill, New York, NY

Montiel R, Malgosa A, Francalacci P (2001) Authenticating ancient human mitochondrial DNA. Hum Biol 73:689–713

Mouret JF, Polverelli M, Sarrazini F, Cadet J (1991) Ionic and radical oxidations of DNA by hydrogen peroxide. Chem Biol Interact 77:187–201

Muruganandhan J, Sivakumar G (2011) Practical aspects of DNA-based forensic studies in dentistry. J Forensic Dent Sci 3:38–45

Norgen Biotek Corp (2015) Cells and tissue DNA isolation kit (magnetic bead system) – 50 preps. In: Resources. https://norgenbiotek.com/sites/default/files/resources/Cells-and-Tissue-DNA-Isolation-Kit-Magnetic-Bead-System-Insert-PI59100-1.pdf

Pagan F, Lim C, Keglovic M, McNevin D (2012) Comparison of DNA extraction methods for identification of human remains. Aust J Forensic Sci 44:117–127. https://doi.org/10.1080/00450618.2011.610821

Palmirotta R, Verginelli F, Di Tota G et al (1997) Use of a multiplex polymerase chain reaction assay in the sex typing of DNA extracted from archeological bone. Int J Osteoarcheology 7:605–609

Piglionica M, De Donno A, Baldassarra SL et al (2012) Extraction of DNA from bones in cases where expectations for success are low. Am J Forensic Med Pathol 33:322–327. https://doi.org/10.1097/PAF.0b013e3182519479

Promega (2019) Purification of DNA from bone samples using bone DNA extraction kit, custom and DNA IQ™ chemistry. In: Appl note. https://www.promega.es/-/media/files/resources/application-notes/genetic-identity/an343-purification-of-dna-from-bone-samples.pdf?la=en. Accessed 2 Jun 2021

QIAGEN (2013) Purification of DNA from bones and teeth using the EZ1® advanced or EZ1 advanced XL. In: QIAGEN Suppl Protoc. https://www.qiagen.com/us/resources/resourcedetail?id=00069c66-ce82-4574-8a5c-7c8f7b16e625&lang=en. Accessed 2 Jun 2021

Richards MB, Sykes BC (1995) Authenticating DNA extracted from ancient skeletal remains. J Archeol Sci 22:291–299

Schroeder H, de Damgaard PB, Allentoft ME (2019) Pretreatment: improving endogenous ancient DNA yields using a simple enzymatic predigestion step. In: Shapiro B, Barlow A, Heintzman P et al (eds) Ancient DNA methods and protocols. Humana Press, New York, pp 21–24

Shapiro IM, Moar MH, Ohno S, Klein G (1978) Acetic acid treatment denatures DNA while preserving chromosomal morphology during the in situ hybridization procedure. Exp Cell Res 115:411–414

Stone AC, Stoneking M (1998) mtDNA analysis of a prehistoric oneota population: implications for the peopling of the new world. Am J Hum Genet 62:1153–1170

Tsuchimochi T, Iwasa M, Maeno Y et al (2002) Chelating resin-based extraction of DNA from dental pulp and sex determination from incinerated teeth with Y-Chromosomal alphoid repeat and short tandem repeats. Am J Forensic Med Pathol 23:268–271

Willard JM, Lee DA, Holland MM (1998) Recovery of DNA for PCR amplification from blood and forensic samples using a chelating resin. In: Lincoln PJ, Thomson J (eds) . Humana Press, Forensic DNA Profiling Protocols, pp 9–18

Yang DY, Eng B, Waye JS et al (1998) Technical note: improved DNA extraction from ancient bones using silica-based spin columns. Am J Phys Anthropol 105:539–543. https://doi.org/10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1

Ye J, Ji A, Parra EJ et al (2004) A simple and efficient method for extracting DNA from old and burned bone. J Forensic Sci 49:1–6. https://doi.org/10.1520/JFS2003275

Usefulness of Quantitative PCR in Forensic Genetics

35

Christian Haarkötter, M. J. Alvarez-Cubero, Juan Carlos Alvarez, and María Saiz

Contents

C. Haarkötter · J. C. Alvarez · M. Saiz
Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain
e-mail: chaarkotter@ugr.es; juanca@ugr.es; msaiz@ugr.es

M. J. Alvarez-Cubero (✉)
Department of Biochemistry and Molecular Biology III, Faculty of Medicine, University of Granada, Granada, Spain

Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), Granada, Spain
e-mail: mjesusac@ugr.es

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_39

Abstract

Although most DNA studies in forensic genetics are focused on the sample, including DNA extraction and amplification, sample quantification is an important step that is required to determine the optimal DNA input for amplification. In addition, the effective quantification of DNA can provide information regarding the degradation and inhibition of DNA to optimize the amplification strategy or the extraction method and can be used to inform the decision to analyze another sample of the same specimen. In this chapter, quantitative PCR (qPCR, also known as real-time PCR, RT-PCR) will be described after a brief history of quantification methods, PCR fundamentals, current applications in forensic genetics [including the sequencing of short-tandem repeats (STRs), single-nucleotide polymorphisms (SNPs), Y-chromosomes, mitochondrial DNA (mtDNA), and non-human DNA], and future perspectives for this technique.

Keywords

DNA analysis · DNA quantification · Forensic genetics · Quantitative PCR · Real-time PCR (RT-PCR)

Introduction

Due to the emergence of direct amplification techniques, for which DNA extraction represents an optional step, DNA quantification has become an alternative technique for some forensic cases. However, in some cases, DNA quantification continues to represent a necessary step for the optimization of DNA concentrations required for successful amplification processes. Quantification not only provides information regarding DNA concentration but by applying quantitative polymerase chain reaction (qPCR), during which DNA amplification is monitored during each cycle, DNA degradation, sex determination, the presence of multiple contributors from different sexes, and the presence of PCR inhibitors can be assessed.

All of this information that can be obtained from DNA quantification can assist investigators during the following sample analysis steps. qPCR offers several advantages to the field of forensic genetics, including cost and time saving for laboratories, providing the information necessary to optimize the entire analytical process, including information regarding DNA quantity and degradation.

In this chapter, the qPCR fundamentals will be discussed, including a review of the history of DNA quantification and an exploration of the current applications for qPCR in the field of forensic genetics, such as autosomal and Y-chromosome quantification, degradation assessment, mitochondrial DNA (mtDNA) analysis, non-human DNA analysis, ancient DNA (aDNA) analysis, and SNP analysis applications, ending with future perspectives for this technique.

DNA Quantification

DNA Quantification Importance

DNA quantification continues to be a mandatory process in most international recommendations, including the Federal Bureau Investigation Quality Assurance Standards, in which Standard 9.4 states that every forensic sample must be quantified before amplification unless the laboratory features a validated system for direct amplification and typing, which is a characteristic of some direct amplification kits. This standard is also included in the Scientific Working Group on DNA Analysis Methods Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.

DNA quantification is necessary for two reasons (Butler 2011). First, the majority of available short-tandem repeat (STR) commercial kits requires a narrow range of DNA input, typically 0.5–1 ng of DNA, a range that can be even lower for the current generation of higher-resolution genetic analyzers. The use of insufficient DNA may result in partial profiles, allelic dropout, allelic drop-in, or null profiles, whereas the use of too much DNA may cause artefacts in electropherograms, such as pull-up spikes, although most modern genetic analyzers have bioinformatic algorithms designed to mitigate the effects of large DNA quantities.

Second, quantification saves time and money, as the quantification results can be used to reorient the laboratory workflow to reconsider the planned extraction or purification methods or to attempt the purification and extraction of another sample if one is available. Finally, DNA quantification is crucial when working with compromised samples, such as aDNA, DNA from bones, or potentially mixtures from multiple sources (Carracedo 2005).

DNA Quantification Methods

Forensic DNA quantification has passed through two different stages: initially, forensic analysts quantified the total amount of DNA content in an extracted sample; however, this approach was not specific to human DNA (Nicklas and Buel 2003a). Therefore, additional methods have been developed to improve specificity (Lee et al. 2015):

1. *Total genomic methods*: These methods measure the total amount of DNA, including both human and non-human DNA, which is found in the sample.
   a) *UV spectrophotometry*: Ultraviolet (UV) spectrophotometry is the most common method used to perform DNA quantification, based on the correlation between nucleotide concentrations and absorption at 260 nm, which allows the A260/A280 ratio to serve as a measure of DNA purity. Today, compact UV spectrophotometers, with the Nanodrop™ Spectrophotometer representing the most common device, can quantify a 1-μl sample without requiring a cuvette (Joseph 2010).

b) *Fluorescent assays*: During this method, a fluorochrome selectively binds to double-stranded DNA, and the fluorescence can be measured and compared against a standard curve made from DNA samples with known concentrations. This is a very stable method that allows assay flexibility, detecting approximately at least 250 pg of DNA (Ahn et al. 1996).
c) *Gel-based quantification*: Samples are loaded onto a 0.8% agarose gel in TBE (Tris-Borate-EDTA buffer), and an electric field is applied, resulting in DNA migration toward the positive pole. Ethidium bromide or other fluorochromes, such as GelRed® (which is less toxic), can then be used to visualize the DNA under UV light (Alonso 2012). This method is semiquantitative and used to determine the presence of total genomic DNA. This method has an approximately 1-ng resolution and can be used to assess the degradation state based on the appearance of a DNA smear line.

2. *Human- and higher primate-specific methods.*
a) *Slot blot*: DNA is immobilized on a nylon membrane, and a biotinylated oligonucleotide probe (D17Z1, which is specific for humans and primates) hybridizes to the DNA, and they both conjugate to a peroxidase. Peroxidase then reduces hydrogen peroxide in a chemiluminescent detection reagent, resulting in the oxidation of luminol and the emission of photons that can be detected using autoradiography film. The DNA concentration can be calculated by comparing the sizes and densities of the dots produced on the film against a dilution series containing standard DNA with known concentrations which makes this method capable of detecting as little as 150 pg of DNA (Walsh et al. 1992).
b) *AluQuant*: The AluQuant™ system was designed by Promega, based on 300-bp DNA fragments that are found in humans and primates, referred to as Alu elements because *Arthrobacter luteus* was used in the initial characterization. When the probe hybridizes to the denatured DNA in a sample, a pyrophosphorylation reaction is initiated, which liberates a deoxynucleotide that becomes transformed into adenosine triphosphate through the activity of a kinase. The adenosine triphosphate then activates luciferin, which produces measurable light. This system is capable of detecting DNA levels as low as 100 pg (Mandrekar et al. 2001).

Quantitative PCR (qPCR) Fundamentals

Detection Methods

TaqMan Probes

PCR represented a breakthrough development when it was first invented, and its utility increased further with the introduction of *Thermophilus aquaticus* usage, which allowed the process to be optimized for specificity. One property of *Taq* polymerase is its 5′→3′ exonuclease activity, through which the polymerase degrades any oligonucleotide that has hybridized between the two forward and

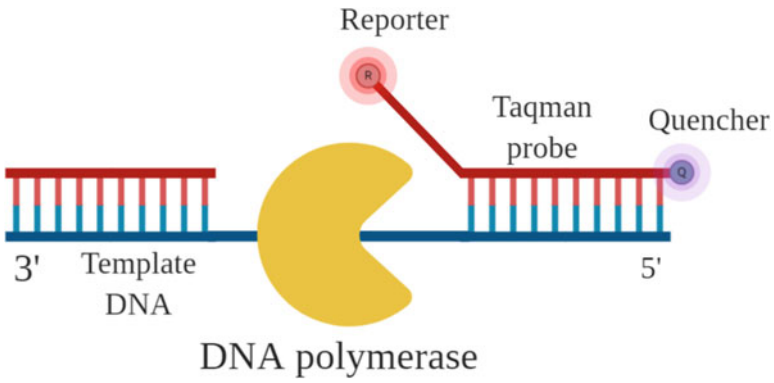


Fig. 1 TaqMan probe diagram (Holland et al. 1991)

reverse primers (Holland et al. 1991). This functionality of the TaqMan probe represents the cornerstone of quantitative PCR (Fig. 1). TaqMan probes are 18–30 bp oligonucleotides designed to feature melting temperatures near the optimal extension temperature of the DNA molecule of interest. Given the short length of probes, minor groove binder molecules are often added to increase the melting temperature of the probe, allowing shorter probes to be used (Kutyavin et al. 2000).

TaqMan probes are bound to two different dyes that transfer energy from one to the other when they are physically close: the reporter (R) dye is typically attached to the 5′-end of the probe, whereas the quencher (Q) dye is attached to the 3′-end of the TaqMan probe. When the DNA polymerase initiates synthesis between the two primers bound to the DNA template, the 5′→3′ exonuclease activity hydrolyses the probe, liberating both the reporter and the quencher, which frees the reporter from the influence of the quenching dye, allowing the reporter to emit fluorescence that can be measured when it is excited by a certain wavelength of light. Table 1 shows the most common reporters and quenchers that are used today.

SYBR® Green qPCR Assays

SYBR Green I is a fluorescent dye that binds only to double-stranded DNA (see Fig. 2) and emits fluorescence proportionally to the amount of bound double-stranded DNA. In a typical real-time PCR reaction, the starting level of input DNA is small; therefore, the only double-stranded DNA that can be detected represents the PCR product. During the exponential phase of the PCR, SYBR® Green fluorescence doubles during each cycle until it reaches the plateau phase; the initial amount of DNA can be calculated according to the difference between the fluorescence cycle threshold of the sample relative to the cycle threshold of a standard (Ponchel 2006).

SYBR® Green is an asymmetric cyanine that is commonly used during gel electrophoresis because it binds between the strands of double-stranded DNA. Therefore, qPCR assays require the inclusion of the DNA template, this dye, and a pair of amplification primers, which can result in the detection of as little as 1 pg of

Table 1 Different reporters and their recommended quenchers (BioCat Real Time PCR Dyes (2009)

Colour	Reporter	Alternative reporter	DYE-5′-T		Recommended quencher	Color
			EX	EM		
	Biosearch Blue™	–	352	447	BHQ-1	
	FAM	–	495	520	BHQ-1	
	TET	–	521	536	BHQ-1	
	CAL Fluor® Gold 540	VIC/TET/JOE	522	544	BHQ-1	
	JOE	–	529	555	BHQ-1	
	VIC	–	538	554	–	
	HEX	–	535	556	BHQ-1	
	CAL Fluor Orange 560	VIC/HEX/JOE	538	559	BHQ-1	
	Quasar® 570	CY3	548	566	BHQ-2	
	ABY®	–	–	580	–	
	Cy™ 3.5	–	581	596	–	
	ROX	–	586	610	BHQ-2	
	CAL Fluor red 610	TEXAS RED/ ROX/ ALEXA FLUOR® 594	590	610	BHQ-2	
	JUN®	–	–	617	–	
	Pulsar ® 650	–	460	650	BHQ-2	
	Cy 5		646	669	–	
	Quasar 670	CY5	647	670	BHQ-2, BHQ-3	
	Cy 5.5		675	694	–	
	Quasar 705	CY5.5	690	705	BHQ-2, BHQ-3	

DNA (Nicklas and Buel 2003b). This alternative quantification method is advantageous due to the reduced need to handle the sample and represents a low-cost assay compared with the slot blot; however, some studies have suggested preferential binding to DNA with higher G + C% and larger-sized amplicons, which, combined with the non-specific binding to any double-stranded DNA, can represent problematic characteristics for forensic contexts (Giglio et al. 2003).

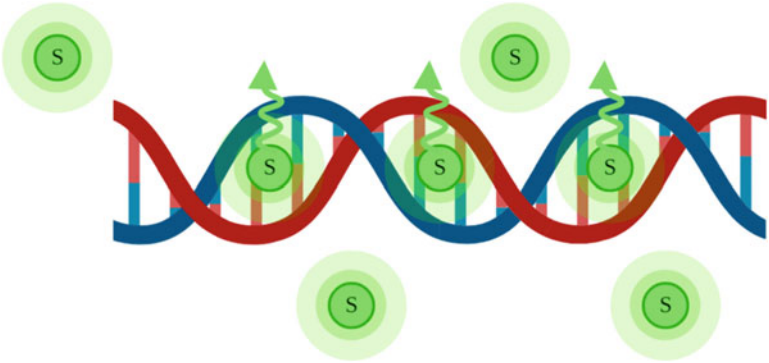


Fig. 2 SYBR® Green binding model

However, this method is particularly advantageous because this quantification assay can be freely designed by any investigator, based on the choice of primer pairs, which can be used to quantify exact regions of interest. Several commercial kits are available, including SYBR™ Green PCR Master Mix (Thermo Fisher, Waltham, MA, USA), SYBR® Green JumpStart™ (Sigma-Aldrich, St. Louis, Mo, USA), and QuantiTect® SYBR® Green PCR (Qiagen, Hilden, Deutschland).

It is important to know if specific products have been obtained or whether dimers or unspecific products are present in qPCR when SYBR Green I is used. That is why it is crucial the presence of a melting curve. DNA melting allows the detection of small amounts of wrong qPCR products and can even detect single-nucleotide differences between amplified targets and copy number. Most of the commercial qPCR devices have a melting rate of 0.01–0.4 °C/s. This range gives enough resolution to discriminate different qPCR products, taking around 5 min per plate the generation of a melting curve. However, the interpretation of this data is not automatized in any qPCR platform, and neither is available an application for automatically evaluating the DNA melting curve of different curves. Nevertheless, there are web applications available to predict melting curves of an amplicon and to identify differences between melting curves (Ruijter et al. 2019). Other methods, less specific and practical, imply agarose gel analysis which implies the visualization of the amplicons in an agarose gel.

Molecular Beacons

Molecular beacons were designed to focus on the annealing specificity between two complementary DNA strands. This method uses specific single-stranded DNA probes that contain a loop with a complementary sequence and two complementary arms, forming a bulb-like shape (see Fig. 3). The stem is attached to both a fluorescence molecule and a quencher, which prevents the emission of fluorescence in its native form. When the molecular beacon finds its complementary sequence, it anneals to the sequence, forming a more stable complex than the complex formed by the annealing of the two complementary arms. The fluorophore and the quencher become detached, allowing the fluorophore to emit detectable fluorescence (Tyagi and Kramer 1996).

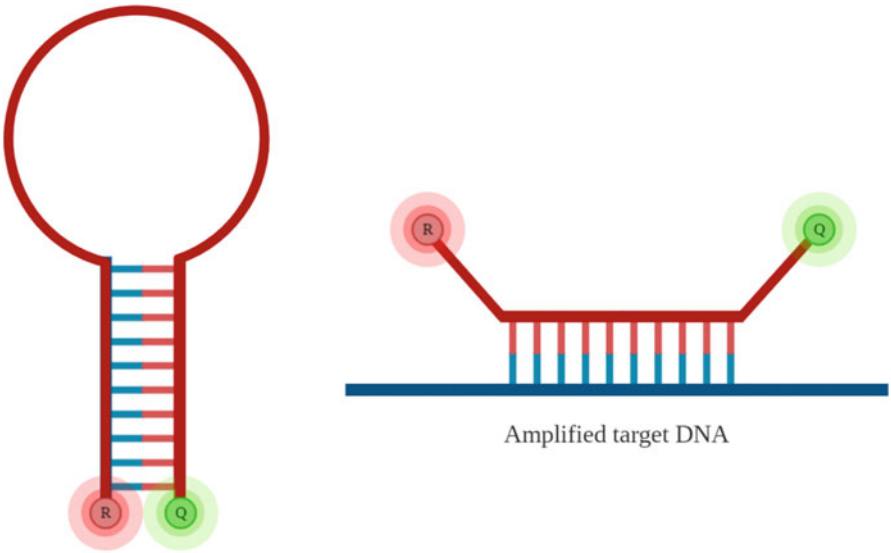


Fig. 3 Molecular beacon hybridization model

Although the primary disadvantage of molecular beacons is a requirement for large complementary sequences to bind (Hopkins and Woodson 2005), several forensic applications have been proposed for these molecules. In the same reaction, multiple targets can be detected using various molecular beacons attached to different fluorophores for each target, limited only by the detection capabilities of the real-time PCR instrument. These characteristics can be useful for the high-throughput genotyping of SNPs (Sobrino et al. 2005) or body fluid identifications using tissue-specific RNA species (Young et al. 2018).

Calibration Curve and DNA Quantity Calculation

Every PCR assay consists of three different stages, based on how the reaction progresses:

1. *Exponential growth*: At the beginning, assuming a 100% PCR efficiency, the number of amplicons will be doubled during each cycle.
2. *Linear growth*: As the PCR assay consumes dNTPs, the primer reaction speed decreases, shifting to arithmetic growth rather than geometric growth.
3. *Plateau phase*: During this stage, the reagents become fully consumed, preventing any additional PCRs from being performed, and the number of amplicons remains stationary.

These phases can be visualized if the detected fluorescence is compared against the cycle number in a graphical representation, which results in a sigmoid function

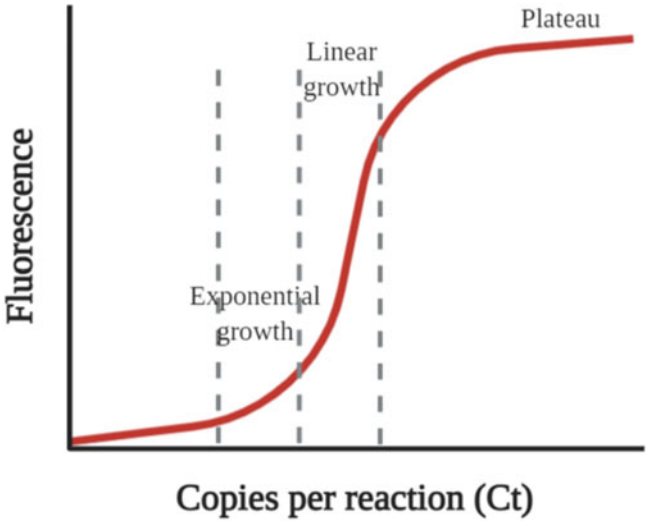


Fig. 4 Plot showing PCR fluorescence vs cycles

(see Fig. 4). The best point to measure the fluorescence is during the exponential growth phase because the number of PCR products in the next cycle, which should be double, or in the previous cycle, which should be half, can be predicted. If we define an arbitrary point in which the level of fluorescence can be detected above baseline noise, larger DNA samples will become detectible after fewer cycles.

The cycle number at which fluorescence above background can be detected is referred to as the threshold cycle, C_T (Heid et al. 1996). If we graphically represent the threshold cycle values for various samples with known concentrations according to the logarithm of the DNA quantity, a linear model can be obtained through linear regression, representing a calibration curve that can be used to calculate the DNA concentrations of the samples, as seen in Fig. 5.

When performing a qPCR assay, determining the coefficient of determination, R squared, or R^2 is very important because this value indicates the goodness of fit of a model by directly measuring how much of the variance observed in one of covariate can be attributed to another covariate. This value is calculated as the square of the correlation coefficient, which ranges from 0 to 1, with values that approach 1, indicating optimal goodness of fit for the calibration curve (Lucy 2005).

Quantitative PCR Instruments

To detect fluorescence during the performance of PCR assays, various real-time PCR systems have been developed, which are typically accompanied by specific software that can automatically calculate the linear regression, the coefficient, and the DNA quantity that is present in each sample.

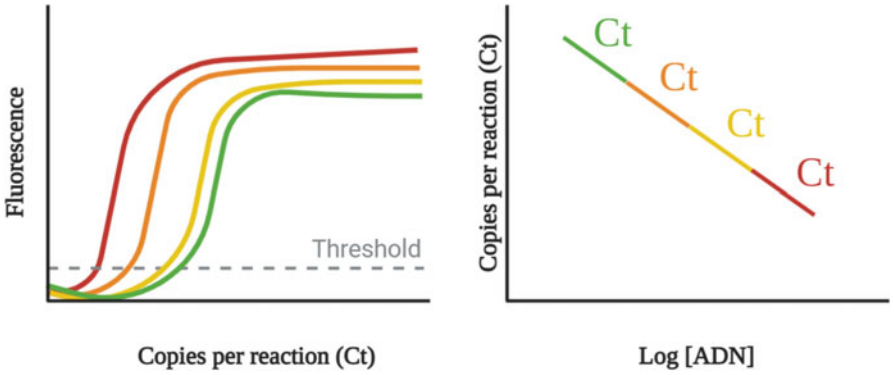


Fig. 5 Real-time PCR fluorescence vs cycle plot; the values of at the C_T threshold can then be plotted against the logarithmic value of the DNA concentration to generate a calibration curve

qPCR systems combine two pieces of equipment that are familiar to the forensic biologist: a thermocycler and a fluorimeter. qPCR systems feature four major components (Shipley 2007): the light source, which determines the range of reporter dyes that can be used by the instrument; the detection system, which determines the spectral range and sensitivity of the; the thermocycling mechanism, which aims to rapidly reach and maintain specific temperatures; and the software, which controls the assay and performs the calculations.

The thermocycler performs the PCR, the reporters become liberated, the light source applies light to the samples, the reporters emit light, and the light is detected by an active pixel sensor, typically a complementary metal-oxide-semiconductor (CMOS) sensor. As shown in Fig. 6, applying an excitation filter and an emission filter facilitates the performance of multiple qPCR assays utilizing different reporters simultaneously (see Fig. 6).

In order to ensure an optimal performance of the PCR instruments, regular calibrations are required that vary according to the instrument. There are different types of calibrations:

i) ROI calibration that allows the software used to locate the wells on the sample block so the increases in fluorescent can be associated to a specific well on the plate.
ii) Background calibration that measures the basal fluorescent generated from background electrical signals, sample blocks, water, and consumables. This calibration is essential to eliminate the background signal, and it increases the instrument accuracy.
iii) Dye spectral calibration that enables the instrument to distinguish the different fluorescent dyes that will be used in the system.
iv) Instrument verification run that verifies that the instrument can generate a standard curve and calculate the quantities of two known samples.

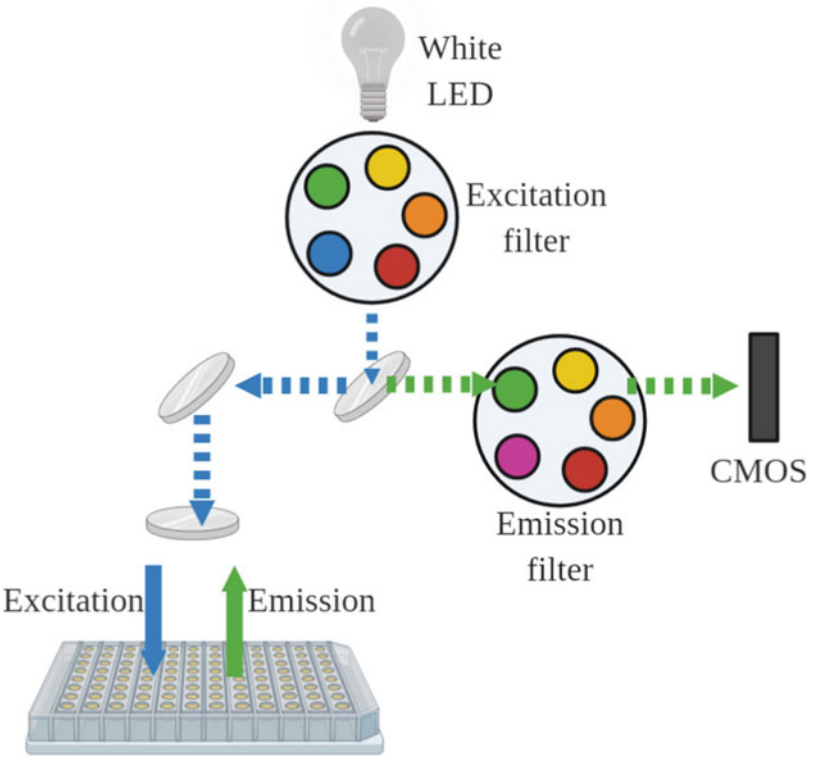


Fig. 6 qPCR instrument diagram

qPCR Customization

There are many commercial kits available to work in qPCR analyses, thanks to the versatility and specificity of the technique; multiple assays can be developed just with the design of the correct primers and probes. Some examples of customized assays developed in qPCR are the assessment of the degree of DNA degradation using primers that amplify fragments of different sizes (Heß et al. 2021); the determination of the degree of inhibitors of the PCR reaction present in the DNA extract; sex determination by the analysis of specific regions of Y and X chromosomes (Ginart et al. 2019). However, not only data from human DNA can be determined but also presence and quantity of fungicidal and bacterial DNA in a human DNA sample (Maggi et al. 2020), the determination of the different species in a crime scene by the analysis of the DNA of common wild animals present in the area, and the identification of the individuals by the determination of their microbiome on different samples (Baggesgaard Sterndorff et al. 2020).

It is crucial an excellent primers and probes design to obtain the best results on the customized qPCR experiments, and there are some fundamental parameters to be considered:

a) *Primer Design*
   - *Primer melting temperatures*: Both primers should have similar melting temperatures to assure that both primers bind simultaneously and amplify efficiently the DNA product.
   - *Primer annealing temperature*: It determines the capacity of primers to anneal to the DNA template. Optimal annealing temperatures will result in the highest product yield with the correct amplicon.
   - *GC content*: An ideal GC content for primers is 50% to allow complexity and maintain a unique sequence.

b) *qPCR Probe Design*
   - *Double-quenched probes* are ideal to lower the background noise and increase the signal.
   - *Location*: Probes should be located close to any of both primers but not overlap the primer binding site.
   - *Probe melting temperature*: Probes should have a melting temperature 6–8 °C higher than primers in order to not compromise the sensitivity of the assay.
   - *Probe annealing temperature*: Should be no more than 5 °C below the lower primer melting temperature.
   - *CG content*: GC content should be 35–65% and avoid a G at the 5′ end to not quench the 5′ fluorophore.

c) *Both Primer and Probe Design*
   - *Complementarity and secondary structure*: Both primer and probe design should analyze the possible development of self-dimers, heterodimers, and hairpins.
   - *Specificity*: It is crucial to assure that selected primers are unique to the target sequence.

d) *Amplicons*
   - *PCR product size*: PCR product size has to be 70–150 bp to guarantee the optimal design of primers and probes (Prediger 2013).

There are several online tools to design and analyze custom design primers and probes: Primer 3 (Untergasser et al. 2012), Primer Blast (Ye et al. 2012), Primer Express Software v3.0.1 and Oligoperfect (ThermoFisher, Waltham, MA, USA), PrimerQuest®, real-time PCR design tool, OligoAnalyzer™ Tool and UNAFold Tool (Integrated DNA Technologies, Coralville, IA, USA), and OligoArchitect and OligoEvaluator (Sigma Aldrich, Saint Louis, MO, USA).

Current Applications of qPCR in Forensic Genetics

Currently, DNA quantification in forensic genetics serves two primary objectives: assessing the quantity, or the DNA amount, and assessing quality, or the degree of degradation, of the extracted DNA. Based on this information, the following steps

associated with the DNA analysis can be modified: the amplification dilution factor can be modified, an initial purification step can be applied to the sample, the analysis protocol or sample handling procedures can be performed with improved efficiency, or other DNA extraction assays can be attempted using a different protocol or another sample if one is available. Therefore, the following five qPCR applications hold promise for the field of forensic genetics (Bunce et al. 2012):

1. *Optimization of DNA extraction*: Several DNA extraction protocols are currently available, including both automated and manual protocols. Understanding the amounts of extracted DNA that can be recovered by various protocols will determine the extraction protocol that should be applied to different types of sample.
2. *Detection of PCR inhibition*: qPCR assays introduce an internal PCR control (IPC), which represents an artificial oligonucleotide that cannot be found in nature; the absence of any IPC result might indicate the presence of a PCR inhibitor in the DNA extracts, which would require different strategies.
3. *Assessment of DNA degradation*: Two fragments are amplified during the qPCR reaction, one longer fragment of approximately 200–300 bp and a second shorter fragment of approximately 70–100 bp, to assess a degradation index – the ratio between the small and long target concentration. Amplification can be optimized according to this index.
4. *Sex determination and the detection of DNA contaminants or mixtures*: Some commercial kits include a human DNA target and a second sample of human male DNA, to allow for the determination of the sex associated with the DNA sample and to calculate the male to female ratio and indicate whether the DNA is mixed or contaminated.
5. *NGS libraries*: When using a next-generation sequencing (NGS) platform, a library for high-throughput sequencing must be built and accurately quantified. Forensic samples are not typically characterized by the availability of large quantities of DNA; therefore, qPCR can be used to quantify DNA samples both before and after enrichment.

Several qPCR strategies have been developed depending on the target, including assays designed for autosomal DNA, sexual chromosomes, and mtDNA (Alonso and García 2007).

Autosomal DNA Quantification

The first published use of qPCR in forensic genetics was the performance of a duplex assay to evaluate both nuclear and mtDNA, using the nuclear-encoded, single-copy gene *RB1*, which is located on chromosome 13, as an autosomal target of 78 bp, using hair, bloodstains, fingerprints, skin debris, and saliva stains as samples (Andréasson et al. 2002). Another study used the human *TH01* STR marker as a target, generating a 62-bp amplicon, using buccal swabs as the sample source (Richard et al. 2003), claiming that a human DNA quantification probe should be sufficiently specific. The next step was the assessment of DNA degradation, through the use of two *TH01*

targets: one that was 170–190 bp in length and another that was 67 bp in length, using different casework samples and prequantified control DNA, inducing artificial degradation through the application of DNase I (Swango et al. 2006).

- **Moderately degraded samples**: All STR alleles were detected with 1 ng DNA input.
- **Highly degraded samples**: Approximately, 80% of STR alleles were detected with 1 ng DNA input.
- **Most degraded samples**: Very low genotyping success.

Currently, commercial kits are available for the quantification of specific human DNA. Quantifiler™ HP and Trio DNA quantification kits (Thermo Fisher Scientific, Massachusetts, USA 2014) both contain an 80-bp, small, autosomal target; a 214-bp, large, autosomal target; and a 130-bp, internal PCR control. Quantifiler™ Trio also contains a 75-bp human male target. Both can detect <1 pg/μL, with a detection range from 0.005 ng/μL to 50 ng/μL of DNA (Holt et al. 2015).

In addition, the PowerQuant® System and Plexor® HY System (Promega, Wisconsin, USA) are also available. The PowerQuant System uses a short 94-bp target, a large 294-bp target, two amplicons of 81 bp and 136 bp that specifically target the Y chromosome, and a long IPC of 436 bp. This system can detect a DNA concentration as low as 0.5 pg/μL, and the use of two Y chromosome targets enhances the sensitivity while minimizing the effects associated with copy number variations that either Y locus could have on the overall Y quantification (Ewing et al. 2016).

The Plexor HY System amplifies a 99-bp sequence from the human *RNU2* locus, on chromosome 17. The Y chromosome amplicon, which is 133 bp, is found in *TSPY*, and this system includes a 150-bp IPC. This system uses Plexor® technology, which measures the reduction in fluorescence that occurs during the PCR reaction, and can detect as little as 3.8 pg of DNA (Krenke et al. 2008).

InnoQuant® and InnoQuant® H-Dye quantification kits (InnoGenomics, New Orleans, LA, USA) rely on *Alu* sequences, using a short 80-bp target and a long 207-bp target, with an internal PCR control of 172 bp, which can assess human DNA at quantities of less than 1 pg/μl (Pineda et al. 2014). Finally, the Investigator® Quantiplex Pro Kit (Qiagen, Hilden, Germany) detects a small human target of 91 bp, a large human target of 353 bp, a human male target of 81 bp, and an IPC of 434 bp and is capable of detecting less than 1 pg of DNA, with a detection limit of 0.25 pg/μl of male DNA mixed into 250 ng/μl of female DNA (Vraneš et al. 2017).

All of these kits perform similarly when using high-quality DNA samples containing low levels of PCR inhibitors, capable of detecting DNA levels under 1 pg; however, the InnoQuant™ HY kit appears to provide the highest precision. For highly inhibited samples, the Investigator® Quantiplex® appears to be more tolerant of the presence of inhibitors, providing accurate quantification results except under conditions of high salt concentrations, when the Quantifiler® Trio appears to perform the best. Occasionally, the use of an STR typing kit in combination with a corresponding qPCR assay kit, such as the combination of Globalfiler® and Quantifiler® Trio, can return better-quality profiles (Holmes et al. 2018).

X and Y Chromosome Quantification

The first X and Y chromosome quantification assays were designed to determine DNA quantity and sex using 106-bp and 112-bp amplicons for the X and Y chromosomes, respectively, and the fluorescent tagging of PCR products, resulting in the amplification of *AMGX* and *AMGY*. The sensitivity of *AMGX* was significantly better than that for *AMGY* (Alonso et al. 2003). Other assays used SYBR Green I and a 70–73-bp amplicon from the amelogenin gene found on the Y chromosome (Andréasson and Allen 2003).

Currently, no commercial kits are available for X or Y chromosome quantification. A Y chromosome target is typically included in most human DNA quantification kits, such as the Quantifiler™ Trio or PowerQuant® System. The Plexor ® HY System includes two Y chromosome targets, allowing for the assessment of Y chromosome degradation, and the Investigator® Quantiplex® Pro RGQ Kit (Qiagen, Hilden, Germany) has also been reportedly used for Y chromosome degradation assessments due to the inclusion of large and small amplification fragments (Kupiec et al. 2018).

Mitochondrial DNA Quantification

Similar to the sex chromosomes, no commercial kit is currently available for the quantification of human mtDNA. Only two companies provide solutions, NovaQuant and ScienCell. However, these kits infer the quantities of mtDNA relative to the quantities of simultaneously quantified nuclear DNA.

The first qPCR assay performed for the quantification of mtDNA used a 142-bp target against the tRNA lysine and ATP synthase gene, which could detect even a single molecule of mtDNA, although testing revealed that a minimum of 80 mtDNA copies was necessary to perform successful sequencing (Andréasson et al. 2002). Other assays used two different amplicon sizes; for example, 113-bp and 127-bp amplicons against the *HV1* region were used to assess mtDNA quantity and the degradation state (Alonso et al. 2004). Many different mtDNA targets have been used during qPCR, including coding regions in *ND1* (Gallimore et al. 2018), *ND4* (Phillips et al. 2014), *ND5* (Kavlick 2019), *RB1* (Niederstätter et al. 2007), *16S* (Fregel et al. 2011), *chrM* (Refinetti et al. 2017), or tRNA lysine and ATP synthase 8 (Alonso et al. 2003) and non-coding regions, such as *HV1* (Alonso et al. 2003), *HV2* (Andréasson et al. 2002), or *HV3* (Phillips et al. 2014). Some mtDNA regions, such as *16S*, are interesting for forensic purposes, because they have been described as being species-specific.

A quadruplex PCR assay for the assessment of the nuclear autosomal DNA using a 95-bp target from *RB1*, the Y chromosome using a 126-bp target from *DAZ*, mtDNA using a 96-bp amplicon from ATPase8, and an IPC has been developed. Autosomal DNA, Y chromosome DNA, and mtDNA DNA can be quantified in one assay that can also determine whether the sample contains inhibitors (Whittle and Sumita 2008). The only missing components are long fragments of autosomal and mtDNA to determine degradation.

Two recent multiplex assays allow the inference of mtDNA degradation (Kavlick 2019). The first assay includes two mtDNA targets within the *16S* rRNA gene and an IPC (Kavlick 2019). The second assay combines one nuclear DNA multicopy target of 70 bp, two mtDNA targets of 69 and 143 bp, and an IPC. This combination of targets provides information regarding the degradation state of the mtDNA. This kit allows for the direct and absolute quantification of mtDNA through the use of a synthetic oligo that can be used to determine the amount of mitochondrial DNA copies in the genomic DNA (Xavier et al. 2019).

qPCR has also been used to study heteroplasmies to assess the heteroplasmic proportions in a given sample or to predict the heteroplasmy degree among samples. A strong relationship has been identified between the heteroplasmic states of hair and buccal cells from the same donor (Gallimore et al. 2018).

Ancient DNA Studies

aDNA is known to be challenging for forensic genetics due to low quantity, inherent DNA degradation, the risk of contamination with exogenous DNA, and the presence of PCR inhibitors. However, several protocols, procedures, and recommendations can be used to mitigate or avoid analytical problems, allowing qPCR to serve as a useful step. qPCR can reduce costs and efforts by providing information regarding whether DNA is present in old samples. In addition, qPCR provides information regarding DNA quantity and quality, allowing analytical procedures to be optimized if quantity is moderated or degradation is detected.

Commercial kits, such as Quantifiler™ Trio, have been successfully used to assess the degradation of bones and human remains. This kit amplifies both short and long DNA fragments, allowing a degradation index value to be calculated based on the relative ratio of small DNA target amount/large DNA target amount, giving the results shown in Table 2.

The difficulty analyzing low-copy-number DNA, defined as less than 100 pg DNA, is often encountered when studying aDNA, often challenging analysts with stochastic variations that can appear as allelic dropouts; therefore, augmenting the cycle number has been proposed, even though this may cause stutter bands and the introduction of exogenous DNA contamination (Gill et al. 2000). RT-PCR assays

Table 2 STR analysis optimization by quantification results (Vernarecci et al. 2015)

		Quantity	
		<33.3 pg/µl >3 pg/µl	**>33.3 pg/µl**
Quality	**DI < 4**	Full or partial profile expected Max volume input DNA recommended	Full profile expected Regular DNA input recommended
	DI > 4	Partial or no profile expected Max volume input DNA recommended	Partial profile expected High DNA input recommended

can be very helpful in these situations because they can assist in the detection of low-copy-number DNA samples; an IPC is used to detect PCR inhibitors, and the use of X and Y chromosome targets allows for the calculation of possible mixture imbalances. The use of two different sizes of amplicons allows for the evaluation of both degradation and copy number (Alonso et al. 2003). Quantitative PCR assays have also been useful during other DNA analysis steps, such as DNA extraction, and can be used to address low-copy-number DNA issues, suggesting the use of a more efficient DNA extraction protocol (Barta et al. 2014).

Non-human DNA Analysis

Obviously, the primary focus of forensic genetics investigations is human DNA; however, qPCR can be helpful for forensic investigations that involve non-human DNA analyses. Common samples in this field include animal species, plants, and microorganisms, primarily due to the wide variety of genomic markers and technologies that can be included in genomic analyses (Arenas et al. 2017). Through qPCR analyses, specific genetic regions, such as cytochrome b (*CYTB*), cytochrome c oxidase I (*COI*), and ribosomal RNA (*rRNA*) genes (Arenas et al. 2017), can be evaluated.

Mitochondrial DNA has the same characteristics in both humans and animals: uniparental inheritance, non-recombinant, rapid-evolution, high copy numbers, and simple structures. With a standard pair of primers, certain species-specific coding regions can be analyzed, and these sequences can be aligned with a DNA database to identify the species, such as BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Cytochrome B represents the most commonly amplified region (Kocher et al. 1989). Quantitative PCR has been used with animal DNA testing, for example, to distinguish domestic animals from human DNA samples in a multiplex PCR assay, capable of detecting even 0.4 pg of human DNA within 4.0 pg of animal DNA (Kanthaswamy et al. 2012). Other studies use qPCR for food authentication by applying a multiplex qPCR to the amplification of different simultaneous animal targets (Köppel et al. 2009).

The DNA testing of plants offers two interesting fields for forensics: one is the analysis of pollen to link a suspect or a victim to a specific area, and the other can be used to link drug cultivations to other plants or individuals (Butler 2011). Studies have been able to identify *Cannabis sativa* species through the use of a nuclear and chloroplast duplex qPCR assay (Johnson et al. 2013).

Microbiome DNA studies have also been developed in different fields. For example, RT-PCR has been suggested for three different oral bacteria DNA, *Streptococcus salivarius*, *Streptococcus sanguinis*, and *Neisseria subflava*, for the identification of saliva in forensic samples, which can indicate that human DNA analysis can also be performed (Jung et al. 2018). A similar principle can be applied to the characterization of soil, particularly as soil evidence is quite common in forensic contexts, and the characterization of microbial DNA, combined with an exhaustive knowledge of biogeography, can place an unknown soil sample within a 25-meter area (Habtom et al. 2019).

Single-Nucleotide Polymorphisms (SNPs)

STRs are a core feature used for human identification, and DNA databases focused on STR analyses continue to grow rapidly. However, SNPs are being increasingly investigated for forensic purposes. Furthermore, SNPs are not only useful for human identification but also for ancestry studies and phenotyping (Butler 2011); SNPs can facilitate the identification of sex, ancestry, eye color, or hair type associated with an unknown sample.

Several SNP sequencing methods have been developed for forensic use, including the SNaPshot® Multiplex System (Thermo Fisher Scientific, Massachusetts, USA), which is a primer extension-based method that can analyze up to 10 SNPs in a single reaction and is widely used for mtDNA analysis. Again, RT-PCR may be useful for SNP analyses because qPCR is well suited for the performance of quality checks in mtDNA examined by SNP analysis; a recent study of mitochondrial DNA analysis performed in hair reported no negative SNP results when the qPCR assay was positive, using a total amount of just 6 µl DNA extract, and simultaneously while reducing laboratory costs and working time (Köhnemann et al. 2010).

Issues Associated with qPCR

No perfect techniques exist, and although qPCR is a useful tool for forensic scientists, it is also associated with a variety of issues and limitations. Understanding these limitations will allow scientists to provide more accurate results.

During this chapter, several RT-PCR advantages have been described; however, each positive aspect is associated with a respective limitation (Table 3). Quantitative PCR allows a large range of quantification; however, because PCR products grow exponentially, a linear model must be constructed to assess DNA quantities.

Table 3 Advantages and limitations of qPCR (Klein 2002)

	Advantages	Limitations
PCR product	A large range of quantification	PCR product growths exponentially
Sensitivity	High sensitivity	The more the cycles, the more the variation introduced
Precision	High precision	Linear values while constructing the calibration curve increase variation
Post-PCR steps	No operations are needed after qPCR assay	Emission spectra may overlap if dye chemistry is not designed accurately
Contamination	Minimized risk of cross-contamination	Sample handling
Efficiency	High-throughput	Increased risk of false-negative and false-positive results
Multiplex	Multiplex approach is available	Only a certain number of simultaneous reactions are available

Although qPCR has a high sensitivity, an increasing number of cycles is also associated with the increased introduction of variations. Although qPCR is associated with high precision, the linear values that are introduced during the construction of the calibration curve, experimental variations may increase variations in this transformation.

Post-PCR steps are generally not necessary; however, the emission spectra of the various dyes may overlap if primers and probes are not designed accurately; similarly, although a multiplex approach is available, this approach is limited to a specific number of simultaneous reactions and depends on the instrument being used. The ability to perform a high throughput of samples is a strong advantage; however, increased risks of obtaining false-negative and false-positive results may result from the application of the same PCR conditions to all gene targets. Finally, the minimization of cross-contamination risk strongly depends on correct sample handling.

Trouble-shooting-related issues must also be considered, based on the following primary topics: primers, probes, and calibration curves. A few recommendations may improve any RT-PCR assay and even eliminate interpretation problems:

1. *Primers and probes* (ThermoFisher Scientific 2007): The use of suboptimal melting temperatures, incorrect concentrations, designs against low-complexity sequences, wrong fluorophores, incorrect target region, overly long amplicons, wrong species, non-target-specific, and fluorophores that are not supported by the instrument are the most commonly encountered issues.
   - Melting temperatures should be maintained between 58 °C and 60 °C for primers and 10 °C higher for probes.
   - The concentration of primers should be 10–100 μM and 2–10 μM in probes. When reconstituting, the proper concentrations can be assessed by spectrophotometric absorbance at 260 nm.
   - In regions with low-complexity, designed primers and probes may be problematic, and an alternative region should be selected, if possible; if not, choose longer primers and probes with higher melting temperatures.
   - Verify that the correct quencher and reporter are being used, as the use of the wrong probe can change the melting temperature and affect PCR efficiency.
   - Low-integrity sequences can result in failed assays, and the presence of SNP sites should be verified by using various databases, including the National Center for Biotechnology Information (NCBI). Increasing the annealing temperature and primer length may be necessary to avoid NABI primer-template mismatch.
   - Amplicon lengths should be 50–150 bp, as longer amplicons may cause poor PCR efficiency.
   - Primers and probes should be verified as being targeted against optimal gene regions by BLASTing (Basic Local Alignment Search Tool) the sequence against those found in public databases.
   - BLASTing is also necessary to verify the species.
   - Probes should be labeled with dyes that have been calibrated or are supported on by the qPCR instrument being used.

2. *Standard curve*: A standard curve can be used to assess PCR efficiency based on a clean matrix and the absence of interference. This curve may be different from the standard curve that is used to estimate target quantities in unknown samples (Svec et al. 2015). Some important issues should be considered:
   - The amplification efficiency curve should cover routine applications plus at least 20%. Extensive standard curve measurements at extremely high or low concentrations can deviate from linearity, and very-high target concentrations can cause problems related to the baseline.
   - Serial dilutions, which are typically used to generate standard curves, may be discouraged due to the carryover of errors. A large transfer volume, at least 5 µl is recommended, across dilutions may be useful for the avoidance of sampling and pipetting errors.
   - Five dilution steps, at six different concentrations, represent the minimum number of concentrations necessary to identify the linear range and to obtain reliable estimations; each concentration should be performed in at least two replicates.
   - When standard concentrations on the log scale are uniformly distributed, an even number of concentrations is recommended.
   - The collected calibration curve should be determined for a standard procedure, including the planned extraction kit or RT-PCR kit and the use of a specific instrument, a plate, and a sealing strategy. Any changes made to the protocol should start with the generation of a new calibration curve.
   - If different sample types are analyzed together, separate standard curves should be constructed for each type.
   - The performance of a single and precise estimation of PCR efficiency, under representative assays conditions, is preferable to the construction of minimal standard curves based on few concentrations in each run.

The availability of a precise nucleic acid quantification method is very important for molecular biologists in general and forensic scientists in particular. To achieve this aim, a forum should exist to promote the exchange of ideas, techniques, tools, and applications (Editorial 2015), which can improve forensic science and other fields.

Future Perspectives for qPCR in Forensic Genetics

Digital PCR (dPCR)

PCR technologies have evolved from end-point PCR, through RT-PCR, to the absolute quantitative digital PCR (dPCR). dPCR is capable of detecting a minimal trace amount of nucleic acids with reliable, quantitative results. Most dPCR instruments adopt the Poisson distribution which, when combined with the use of fluorescent probes and DNA dilutions, statistically results in one or zero molecules in each reaction chamber prior to amplification, which may explain the success of this method. This method is widely spreading due to its high sensitivity, precision, and accuracy for target quantification, requiring only a small amount of sample.

This recent technology provides new methods for the detection and quantification of nucleic acids and is viewed as an alternative method to qPCR, particularly for the detection of rare alleles. This methodology is used to divide DNA or cDNA samples into many individual PCR reactions, increasing the likelihood that a single molecule will be amplified a million-fold or more. Furthermore, the use of nanofluidic chips can provide convenient and straightforward mechanisms for performing thousands of PCR reactions in parallel (Scientific).

In brief, dPCR is a method for absolute nucleic acid quantification, which hinges on the detection of end-point fluorescent signals and the enumeration of binomial events [the absence (0) or presence (1) of fluorescence in a partition]. Using Poisson statistical approximations, dPCR identifies the parameters that constrain the performance metrics of this analytical method (Quan et al. 2018).

Most dPCR applications are in biomedical fields, such as the search for rare mutations and nucleic acid quantification. Moreover, dPCR can also play relevant roles in the detection of pathogens, chromosome abnormalities, tumor DNA in liquid biopsies from tumor patients, and prenatal diagnoses (Cao et al. 2017). In the forensic field, any minimal quantity of biological evidence is extremely valuable during the investigation of a crime or other forensic investigation. A primary challenge is the reduction of inhibitors, such as humic acid and hematin, which can impair DNA polymerase activity, or immunoglobulin G, which binds single-stranded genomic DNA, reducing template availability (Sidstedt et al. 2019).

However, few publications have examined the application of dPCR to forensic casework. dPCR has been shown to be very accurate in the determination of plasmid DNA levels and is particularly sensitive to DNA loss caused by degradation or adsorption. These determinations are relevant for obtaining better data analyses at forensic scenes (Wang et al. 2019). In cases of rare mtDNA deletions, dPCR tool could also be used as an accurate method for improving the detection range (Manoj 2016).

dPCR appears to be less affected by inhibitors than qPCR, due to the application of end-point measurements; therefore, there is no subordination to amplification kinetics. However, for RT-PCR, quantification cycle values are related to a standard curve; therefore, any inhibitory effects will affect quantification. In contrast, dPCR inhibitor tolerance cannot be explained by the use of the end-point strategy itself because complete inhibition has been observed for lower amounts during qPCR compared with dPCR. Sample partitioning may be another factor to consider, resulting in the reduced interaction between molecules during the polymerization process, including inhibitor molecules (Sidstedt et al. 2020). dPCR represents a promising technology that may surpass qPCR in clinical applications due to its strength and technical reproducibility (Quan et al. 2018).

Conclusions

DNA quantification represents a mandatory step in forensic DNA analysis, as stated by the Federal Bureau Investigation DNA Laboratories Quality Assurance guidelines and the Scientific Working Group on DNA Analysis Methods (SWGDAM)

interpretation guidelines. The importance of DNA quantification is due to the narrow range of acceptable DNA inputs when using commercial STR typing kits and the ability to save both time and money if the quantitation results are negative. However, quantification can be used for many other purposes.

Quantitative PCR assays have many applications within the field of forensic genetics, including the optimization of the DNA extraction, the detection of PCR inhibition, the assessment of DNA degradation, the detection of contamination, and the preparation of NGS libraries. Several commercial kits are available for autosomal DNA quantification, which often includes two target fragments, a long one and a short one, a Y-chromosome target, and an IPC that allows human autosomal DNA quantity, the degradation rate, the male human DNA quantity, the male-female proportions of the total human DNA detected, and the presence or absence of PCR inhibitors.

However, several RT-PCR assays have been used in the literature for many purposes, including the assessment of Y-chromosome degradation and quantification, X-chromosome quantification, fast sex determination, mtDNA multiplex qPCR assays to assess mtDNA quantity and quality, and even the presence of PCR inhibitors in a single assay. qPCR can also be applied to SNP typing, aDNA analysis, and emerging non-human DNA analysis, relative to animals, plants, and the microbiome.

DNA quantification assessments performed during daily laboratory work can be used to adjust the DNA inputs of STR typing, saving both work and money. Moreover, several commercial kits are available for assessing DNA quantity and quality, to allow the optimization of DNA analysis, by continuing with the planned process if the DNA quantity and degradation rate are sufficient and changing the amplification approach, choosing another extraction protocol, or using another sample if the quantity and quality are suboptimal. Subsequently, qPCR represents a valuable technique for forensic genetics, which is a challenging field in which any help is welcome.

References

Ahn SJ, Costa J, Emanuel JR (1996) PicoGreen quantitation of DNA: effective evaluation of samples pre-or post-PCR. Nucleic Acids Res 24:2623–2625

Alonso A (ed) (2012) DNA electrophoresis protocols for forensic genetics. Humana Press, London

Alonso A, García O (2007) Real-time quantitative PCR in forensic science. In: Rapley R, Whitehouse D (eds) Molecular forensics. Wiley, Chinchester

Alonso A, Martín P, Albarrán C et al (2003) Specific quantification of human genomes from low copy number DNA samples in forensic and ancient DNA studies. Croat Med J 44:273–280

Alonso A, Martín P, Albarrán C et al (2004) Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. Forensic Sci Int 139: 141–149

Andréasson H, Allen M (2003) Rapid quantification and sex determination of forensic evidence materials. J Forensic Sci 48:1–8

Andréasson H, Gyllensten U, Allen M (2002) Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis. BioTechniques 33:402–411

Arenas M, Pereira F, Oliveira M et al (2017) Forensic genetics and genomics: much more than just a human affair. PLoS Genet 13:e1006960

Baggesgaard Sterndorff E, Russel J, Jakobsen J et al (2020) The T-shirt microbiome is distinct between individuals and shaped by washing and fabric type. Environ Res 185:109449

Barta JL, Monroe C, Teisberg JE et al (2014) One of key characteristics of ancient DNA, low copy number, may be a product of its extraction. J Archeol Sci 46:281–289

BioCat Real Time PCR Dyes (2009) In: Genomics. https://www.biocat.com/genomics/real-time-pcr-dyes. Accessed 6 May 2016

Bunce M, Oskam CL, Alletoft ME (2012) Quantitative real-time PCR in aDNA research. In: Shapiro B, Hofreiter M (eds) Ancient DNA methods and protocols. Humana Press, New York, pp 121–133

Butler JM (2011) Advanced topics in forensic DNA typing: methodology. Academic Press, San Diego

Cao L, Cui X, Hu J et al (2017) Advances in digital polymerase chain reaction (dPCR) and its emerging biomedical applications. Biosens Bioelectron 90:459–474

Carracedo Á (ed) (2005) Forensic DNA typing. Humana Press, Totawa

Editorial (2015) qPCR, dPCR, NGS – a journey. Biomol Detect Quantif 3:A1–A5

Ewing MM, Thompson JM, McLaren RS et al (2016) Human DNA quantification and sample quality assessment: developmental validation of the PowerQuant® system. Forensic Sci Int Genet 23:166–177

Fregel R, Almeida M, Betancor E et al (2011) Reliable nuclear and mitochondrial DNA quantification for low copy number and degraded forensic samples. Forensic Sci Int Genet Suppl Series 3:e303–e304

Gallimore JM, McElhoe JA, Holland MM (2018) Assessing heteroplasmic variant drift in the mtDNA control region of human hairs using an MPS approach. Forensic Sci Int Genet 32:7–17

Giglio S, Monis PT, Saint CP (2003) Demonstration of preferential binding of SYBR green I to specific DNA fragments in real-time multiplex PCR. Nucleic Acids Res 31:e136

Gill P, Whitaker J, Flaxman C et al (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 112:17–40

Ginart S, Caputo M, Corach D, Sala A (2019) Human DNA degradation assessment and male DNA detection by quantitative-PCR followed by high-resolution melting analysis. Forensic Sci Int 295:1–7

Habtom H, Pasternak Z, Matan O et al (2019) Applying microbial biogeography in soil forensics. Forensic Sci Int Genet 38:195–203

Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. Genome Methods 6:986–994

Heß SA, Trapani S, Boronat MDM et al (2021) Ribosomal DNA as target for the assessment of DNA degradation of human and canine DNA. Legal Med 48:101819

Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5′→3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A 88:7276–7280

Holmes AS, Houston R, Elwick K et al (2018) Evaluation of four commercial quantitative real-time PCR kits with inhibited and degraded samples. Int J Legal Med 132:691–701

Holt A, Wootton SC, Mulero JJ et al (2015) Developmental validation of the QuantifilerTM HP and Trio kits for human DNA quantification in forensic samples. Forensic Sci Int Genet 21:145–157

Hopkins JF, Woodson SA (2005) Molecular beacons as probes of RNA unfolding under native conditions. Nucleic Acids Res 33:5763–5770

Johnson CE, Premasuthan A, Tarsk JS, Kanthaswamy S (2013) Species identification of Cannabis sativa using real-time quantitative PCR (qPCR). J Forensic Sci 58:486–490

Joseph LJ (2010) Setting up a laboratory. In: Weiss RE, Refetoff S (eds) Genetic diagnosis of endocrine disorders. Academic Press, MA, Burlington, pp 303–314

Jung JY, Yoon HK, An S et al (2018) Rapid oral bacteria detection based on real-time PCR for the forensic identification of saliva. Sci Rep 8:10852

Kanthaswamy S, Premasuthan A, Ng J et al (2012) Quantitative real-time PCR (qPCR) assay for human-dog-cat species identification and nuclear DNA quantification. Forensic Sci Int Genet 6: 290–295

Kavlick MF (2019) Development of a triplex mtDNA qPCR assay to assess quantification, degradation, inhibition, and amplification target copy numers. Mitochondrion 46:41–50

Klein D (2002) Quantification using real-time PCR technology: applications and limitations. Trends Mol Med 8:257–260

Kocher TD, Thomas WK, Meyer A et al (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86: 6196–6200

Köhnemann S, Pennekamp P, Schmidt PF, Pfeiffer H (2010) qPCR and mtDNA SNP analysis of experimentally degraded hair samples in forensic casework. Int J Legal Med 124:337–342

Köppel R, Zimmerly F, Breitenmoser A (2009) Heptaplex real-time PCR for the identification and quantification of DNA from beef, pork, chicken, Turkey, horse meat, sheep (mutton) and goat. Eur Food Res Technol 230:125–133

Krenke BE, Nassif N, Sprecher CJ et al (2008) Developmental validation of a real-time PCR assay for the simultaneous quantification of total human and male DNA. Forensic Sci Int Genet 3:14–21

Kupiec T, Janula M, Doniec A (2018) Assessment of Y chromosome degradation level using the investigator® Quantiplex® pro RGQ kit. Kraków, Hilden, Germany

Kutyavin IV, Afonina IA, Mills A et al (2000) 3′-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res 28:655–661

Lee SB, McCord B, Buel E (2015) Advances in forensic DNA quantification: a review. Electrophoresis 35:3044–3052

Lucy D (2005) Introduction to statistics for forensic scientists. Wiley, Chichester

Maggi RG, Richardson T, Breitschwerdt EB, Miller JC (2020) Development and validation of a droplet digital PCR assay for the detection and quantification of Bartonella species within human clinical samples. J Microbiol Methods 176:106022

Mandrekar MN, Erickson AM, Kopp K et al (2001) Development of a human DNA quantitation system. Croat Med J 42:336–339

Manoj P (2016) Droplet digital PCR technology promises new applications and research areas. Mitochondrial DNA A DNA Mapp Seq Anal 27:742–746

Nicklas JA, Buel E (2003a) Quantification of DNA in forensic samples. Anal Bioanal Chem 376: 1160–1167

Nicklas JA, Buel E (2003b) Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 48:1–9

Niederstätter H, Köchl S, Grubweiser P et al (2007) A modular real-time PCR concept for determining the quantity and quality of human nuclear and mitochondrial DNA. Forensic Sci Int Genet 1:29–34

Phillips NR, Sprouse ML, Roby RK (2014) Simultaneous quantification of mitochondrial DNA copy number and deletion ratio: a multiplex real-time PCR assay. Sci Rep 4:1–7

Pineda GM, Montgomery AH, Thompson R et al (2014) Development and validation of InnoQuantTM, a sensitive human DNA quantitation and degradation assessment method for forensic samples using high copy number mobile elements Alu and SVA. Forensic Sci Int Genet 13:224–235

Ponchel F (2006) Real-time PCR using SYBR® green. In: Dorak MT (ed) Real-time PCR. Taylor & Francis, Abingdon, pp 139–153

Prediger E (2013) How to design primers and probes for PCR and qPCR. Integrated DNA Technologies. https://eu.idtdna.com/pages/education/decoded/article/designing-pcr-primers-and-probes. Accessed 20 Aug 2018

Quan P-L, Sauzade M, Brouzes E (2018) dPCR: a technology review. Sensors (Basel) 18:1271

Refinetti P, Warren D, Morgenthaler S, Ekstrøm PO (2017) Quantifying mitochondrial DNA copy number using robust regression to interpret real time PCR results. BMC Res Notes 10:E1–E7

Richard ML, Frappier RH, Newman JC (2003) Developmental validation of a real-time quantitative PCR assay for automated quantification of human DNA. J Forensic Sci 48:1041–1046

Ruijter JM, Ruiz-Villalba A, van den Hoff AJJ et al (2019) Removal of artifact bias from qPCR results using DNA melting curve analysis. FASEB J 33:14542–14555
Thermo Fisher Scientific Digital PCR (2014) Polymerase Chain Reaction. https://www.thermofisher.com/es/es/home/life-science/pcr/digital-pcr.html
Shipley GL (2007) An introduction to real-time PCR. In: Dorak MT (ed) Real-time PCR. Taylor & Francis, Abingdon, pp 1–39
Sidstedt M, Steffen CR, Kiesler KM et al (2019) The impact of common PCR inhibitors on forensic MPS analysis. Forensic Sci Int Genet 40:182–191
Sidstedt M, Rådström P, Hedman J (2020) PCR inhibition in qPCR, dPCR and MPS—mechanisms and solutions. Anal Bioanal Chem 412:2009–2023
Sobrino B, Brión M, Carracedo A (2005) SNPs in forensic genetics: a review on SNPS typing methodologies. Forensic Sci Int 154:181–194
Svec D, Tichopad A, Novosadova V (2015) How good is a PCR efficiency estimate: recommendations for precise and robust qPCR efficiency assessments. Biomol Detect Quantif 3:9–16
Swango KL, Timken MD, Chong MD, Buoncristiani MR (2006) A quantitative PCR assay for the assessment of DNA degradation in forensic samples. Forensic Sci Int 158:14–26
ThermoFisher Scientific (2007) Top Ten Pitfalls in quantitative real-time PCR PRimer probe design and use. In: Sample Preparation for PCR, Real time PCR RT-PCR. https://www.thermofisher.com/es/es/home/references/ambion-tech-support/rtpcr-analysis/general-articles/top-ten-pitfalls-in-quantitative-real-time-pcr-primer.html
Tyagi S, Kramer FR (1996) Molecular beacons: probes that fluorescence upon hybridization. Nat Biotechnol 14:303–308
Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3 – new capabilities and interfaces. Nucleic Acids Res 40:e115
Vernarecci S, Ottaviani E, Agostino A et al (2015) Quantifiler® trio kit and forensic samples management: a matter of degradation. Forensic Sci Int Genet 16:77–85
Vraneš M, Scherer M, Elliott K (2017) Development and validation of the investigator® Quantiplex pro kit for qPCR-based examination of the quantity and quality of human DNA in forensic samples. Forensic Sci Int Genet Suppl Ser 6:e518–e519
Walsh PS, Varlaro J, Reynolds R (1992) A rapid chemiluminescent method for quantitation of human DNA. Nucleic Acids Res 20:5061–5065
Wang Y, Keith M, Leyme A et al (2019) Monitoring long-term DNA storage via absolute copy number quantification by ddPCR. Anal Biochem 583:113363
Whittle MR, Sumita DR (2008) Quadruplex real-time PCR for forensic DNA quantitation. Forensic Sci Int Genet Suppl Ser 1:86–88
Xavier C, Eduardoff M, Strobl C, Parsons W (2019) SD quants-sensitive detection tetraplex-system for nuclear and mitochondrial DNA quantification and degradation inference. Forensic Sci Int Genet 42:39–44
Ye J, Coulouris G, Zaretskaya I et al (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinfor 13:134
Young ST, Moore JR, Bishop CP (2018) A rapid, confirmatory test for body fluid identification. J Forensic Sci 63:511–516

Logical Errors and Fallacies in DNA Evidence Interpretation

36

Andrei Semikhodskii

Contents

Abstract

Logical errors and fallacies in assessing forensic evidence are a major issue recognized by both forensic and legal community. They have a propensity to affect interpretation of DNA and other types of evidence by the fact finder and can be a potential cause of miscarriage of justice. Even though they have been a subject of many criminal appeals and publications in forensic and legal journals, they keep on cropping up again and again. This chapter discusses how DNA evidence is assessed and evaluated by the forensic expert and the fact finder in the courtroom, reviews commonly encountered reasoning errors and logical fallacies, examines their causes, and considers the ways which need to be taken to avoid them.

A. Semikhodskii (✉)
Medical Genomics Ltd., Tver, Russian Federation
e-mail: andrei@medicalgenomics.ru

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_40

Keywords

DNA evidence interpretation · Logical errors · Probative value of DNA evidence · Weigh of DNA evidence · Random Match Probability · Evidetial proof · Prosecutor's fallacy · Defendant's fallacy · Uniqueness error

Introduction

The strength of DNA evidence is reflected by the value of the probability of a random match between the DNA profile from the crime sample and the defendant. Any probabilistic approach introduces uncertainly into interpretation of scientific evidence which, if not properly understood, can give rise to a misinterpretation of the evidence and inaccurate conclusions as to what the evidence really means. To be able to deal with errors of evidence misinterpretation, it is important to know their origin and how they can affect court proceedings.

The errors of misinterpretation of DNA evidence refer to logical errors by forensic scientists or lawyers that happen in the courtroom. These errors have a propensity to affect interpretation of evidence by the fact finder and can be a potential cause of miscarriage of justice. Most of these errors are related in one way or another to a tendency to interpret DNA evidence as means of giving a comment on the truthfulness or otherwise of the prosecution's proposition (Aitken and Taroni 2004) as to the source of the crime stain or the guilt of the accused.

It is not in the domain of the forensic expert to assess the prosecution or defense propositions. It is a matter for the fact finder to determine whether DNA evidence, together with all the other evidence adduced, can prove that the crime stain was left by the defendant or is sufficient enough to warrant a finding of guilt. Not to fall into danger of committing a reasoning error the scientist has to understand the meaning of DNA evidence, the questions DNA evidence can and cannot answer, how evidence is used to prove a fact at issue, and the role of a forensic expert during a criminal trial. Nearly all of the errors or interpretation stem from misunderstanding one of these points.

In this chapter the logic of forensic DNA inference and the errors and fallacies commonly encounter in the court room will be examined. Interpretation of forensic evidence is typically performed within a framework appropriate for both scientific analysis and presentation in the courtroom. We start with how the propositions for evidence evaluation are formed and how, by comparing prosecution and defense theories of events, the weight of DNA evidence is obtained. Then, we will look at the random match probability and what it does and does not mean. We will examine the reasoning process employed by the fact finder to reach conclusion on the source of biological material and ultimately the issue of guilt and were the role of the forensic scientist and DNA evidence lays in this process. Thereafter, various logical fallacies and reasoning errors will be examined and elucidated. Last but not least we discuss why these errors happen and what can be done to minimize this.

Initial Assessment of DNA Evidence

Interpretation of DNA evidence happens within a framework of contrasting propositions. In order to evaluate the evidence correctly, the scientist must first formulate different propositions representing both prosecution and defense hypotheses. Here I follow Aitken and Taroni (2004) and Buckleton (2005) in describing different propositions available to the scientist for setting DNA evidence into the framework acceptable for scientific evaluation and presentation in court.

The framing of the propositions is based on three key principles (Evett and Weir 1998):

1. Evaluation of DNA evidence is only meaningful when at least one of the two (or in some cases more) competing propositions is addressed. Usually these are the prosecution hypothesis (H_p) and the defense hypothesis (H_d).
2. Evaluation of DNA evidence considers the probability of evidence under the two competing propositions.
3. Evaluation of evidence is carried out within a framework of circumstances and is conditioned not only by the competing propositions but also by the structure and content of the framework.

The propositions or hypotheses must be mutually exclusive (Robertson and Vignaux 1995) and are formulated so that they represent the prosecution and defense views (even though because of the presumption of innocence, the defense is not required to put forward any proposition whatsoever), for example:

H_p: The DNA comes from the suspect
H_d: The DNA comes from a male other than the suspect

Commonly, there are only two propositions which are analyzed in pairs, but because the hypotheses do not need to be exhaustive, there may be three or more propositions although, in most cases, they can be reduced to just two alternatives. For example, the following propositions:

H_p: The DNA comes from the suspect
H_{d1}: The DNA comes from the father of the suspect
H_{d2}: The DNA comes from the brother of the suspect

can be reduced to:

H_p: The DNA comes from the suspect
H_d: The DNA comes from a male related to the suspect

The choice of the propositions to be addressed depends on the circumstances of the case, the observations that have been made, and the available background information (Aitken and Taroni 2004). Setting appropriate alternative propositions is crucial for

correct evaluation of DNA evidence as the inappropriate choice of proposition can make the evidence irrelevant in the circumstances of a particular case, leading to its dismissal.

There will be various propositions considered by various parties involved in the case and ultimately the trial. Aitken and Taroni (2004) classify the hierarchy of the propositions into four levels:

(i) Offence (crime) Level Proposition
(ii) Activity Level Proposition
(iii) Source Level Proposition
(iv) Sublevel (sub-source level) Proposition

The Offence Level Proposition is at the top of the hierarchy and deals with the ultimate issue such as "the defendant is guilty of raping the complainant." This proposition is solely the domain of the fact finder and is above the level of forensic scientist. All evidence adduced at the trial (and DNA evidence usually is only one of many types of evidence) must be taken into account when assessing the guilt or innocence of the defendant. At this level, the fact finder evaluates whether all the necessary elements of crime (e.g., *mens rea* and *actus reus* for common-law jurisdictions) in relation to the defendant have been proved by the prosecution to the required standard.

The second level of hierarchy, the Activity Level, deals with the act in question, for example, "the defendant had a sexual intercourse with the complainant." This level of proposition can be addressed by the scientist, but, because many other types of evidence are not accessible to her, the Activity Level should also be left for the fact finder to consider. As an example, semen of the defendant can be found on a vaginal swab from the complainant, but it does not necessarily mean that the defendant had a sexual intercourse with her. Semen obtained from the defendant in a way other than vaginal intercourse (e.g., oral sex or masturbation) could have been planted inside the vagina by the complainant who had a grudge against him and then accused him of raping her. The evidence related to the possibility of oral sex or masturbation may not be available to the scientist, making her conclusion on the possibility of the intercourse between the defendant and the complainant not only groundless but a potential cause for miscarriage of justice.

The Source Level is the penultimate level of the hierarchy. This level is an exclusive prerogative of the forensic scientist who considers questions such as "did the semen from the vaginal swab originate from the suspect?" using available background information for the case.

At the very bottom of the hierarchy is the Sublevel (sub-source level) Proposition. Because of the sensitivity of DNA technology, there may be uncertainty regarding which body fluid or cell type the DNA could have come from and at this level is required to answer such a question. Sublevel propositions consider just the DNA in isolation, without trying to attribute it to a particular body fluid or cell type. The typical proposition considered at this level is "does the DNA come from the suspect?"

Framing the propositions is the first stage of the analysis. Next, the scientist has to calculate the probative value of evidence and provide it to the court.

The Probative Value of DNA Evidence

When the fact finder evaluates the most probable origin of DNA evidence (evidentiary fact), it is done by weighing the prosecution and the defense hypotheses as to the fact at issue and determining which of the alternative explanations is most probable. This is done by considering the DNA evidence and all the other evidence adduced. The scientist employs a similar process. Although she cannot express her opinion directly as to the source of DNA evidence, she can provide the court with her estimation of under which of the alternative propositions observation of DNA evidence is more likely. This is done by comparing the likelihood of the evidence if the prosecution theory is true with the likelihood of the evidence if the defense theory is true. The relative values of these two likelihoods provide a measure of the meaning and the probative value of the evidence. This estimation is usually given to court as a ratio known as the likelihood ratio.

Central to this approach (called logical or Bayesian approach) is Bayes' theorem, formulated by Reverend Thomas Bayes and published posthumously in 1763 in *An Essay towards solving a Problem in the Doctrine of Chances*. Applied to DNA evidence, Bayes' theorem deals with the likelihood of the fact at issue under two alternative propositions (the prosecution and defense hypotheses) before and after DNA evidence is adduced.

Let:

H_p be the hypothesis advanced by the prosecution
H_d be the hypothesis advanced by the defense
E be the event that defendant's DNA profile matches the DNA profile from the crime stain
I be all background (other) evidence in the case

Then we can formulate probability Pr of the hypotheses advanced by the prosecution H_p and the defense H_d, given DNA evidence E and other evidence I as $\Pr(H_p \mid E,I)$ and $\Pr(H_d \mid E,I)$, respectively (the sign "|' reads as "given').

Using laws of probability, it is possible to compare the two alternative probabilities as:

$$\frac{\Pr(H_p \mid E,I)}{\Pr(H_d \mid E,I)} = \frac{\Pr(E \mid H_p,I)}{\Pr(E \mid H_d,I)} \times \frac{\Pr(H_p \mid I)}{\Pr(H_d \mid I)}$$

The expression:

$$\frac{\Pr(H_p \mid I)}{\Pr(H_d \mid I)}$$

termed "prior odds" represents the view of the alternative hypotheses which is formed by the fact finder *before* DNA evidence was adduced. This is something which is not expressed numerically and does not need to be, as the opinion on the fact at issue is

formed in the minds of the fact finder based on the evidence adduced *prior* to DNA evidence. Being the exclusive realm of the fact finder, "prior odds' are beyond the scope of the forensic scientist who must not express her opinion on this matter.

The expression:

$$\frac{\Pr(H_p \mid E, I)}{\Pr(H_d \mid E, I)}$$

called "posterior odds" represents the view of the fact finder on the alternative hypotheses of events *after* DNA evidence was adduced.

The expression:

$$\frac{\Pr(E \mid H_p, I)}{\Pr(E \mid H_d, I)}$$

called "likelihood ratio" (LR) describes whether the DNA evidence is more likely under the prosecution (numerator) or defense (denominator) hypothesis and is the numerical measure of the weight of DNA evidence. When a forensic scientist assesses the evidence, she typically reports only the LR, which is then used by the fact finder to weigh the prosecution and defense hypothesis against each other. If the LR is greater than 1, the evidence supports the prosecution hypothesis on the origin of DNA in the crime sample. The higher is the LR, the more weight is put on the prosecution hypothesis. Alternatively, the smaller the LR, the more weight is put on H_d. When the likelihood ratio is close to 1, both prosecution and defense hypotheses are likely, and the probative value of DNA evidence is very small.

To illustrate how the likelihood ratio is calculated, let us look at the following simple example. Assume that a crime was committed. Let c be the stain recovered at the scene. Suppose DNA analysis revealed that the profile of the stain is G_c. Later a suspect s was arrested whose DNA profile G_s is identical to G_c. The prosecution alleges that DNA in the crime stain comes from the suspect s who is the perpetrator of the crime. The defense's case is that it was not the suspect s but the real criminal x who is the donor of the DNA found in the crime sample and that the suspect s and the real criminal x have identical DNA profiles due to chance. In this case, the prosecution and defense hypotheses will be:

H_p: The DNA in the crime stain c came from the suspect s.
H_d: The DNA in the crime stain c came from the real criminal x.

In order to estimate the weight of DNA evidence, let us calculate the probability Pr of DNA evidence under H_p and H_d. This probability for H_p will be $\Pr(Gc \mid G_s, H_p)$ which is read as "the probability of finding the DNA profile G_c in the crime stain c, given that the suspect s has the DNA profile G_s." As under H_p it was the suspect who had committed the crime and nobody else, $\Pr(Gc \mid G_s, H_p) = 1$.

According to H_d the suspect s and the real criminal x have identical DNA profiles. Let us write the probability of finding this profile as $\Pr(Gc \mid G_s, H_p)$. Then, we can write the likelihood ratio as follows:

$$\mathrm{LR} = \frac{\Pr\left(G_c \mid G_s, H_p\right)}{\Pr(G_c \mid G_s, H_d)} = \frac{1}{\Pr(G_c \mid G_s, H_d)}$$

The likelihood ratio is an objective estimation of the weight of DNA evidence because it is based on parameters and statistics that are obtained experimentally (e.g., allelic frequencies or the coefficient of co-ancestry).

The expression $\Pr(G_c \mid G_s, H_d)$ is the random match probability which is the probability that a randomly taken individual has the same DNA profile, given that the defendant has identical DNA profile. The smaller is random match probability, the more the LR would favor H_p; on the opposite, the higher it is, the more weight is given to *Hd*.

As the concept of the match involves two people, the match probability is not the same as the probability (frequency) of a DNA profile, and it is important to distinguish between the two. The probability of a DNA profile is a chance that a randomly taken individual will have a particular profile (e.g., a DNA profile of the crime stain), while the random match probability is the chance that a randomly taken individual will have a particular profile when it is already known that another individual (defendant) has this profile. Therefore, the random match probability explicitly requires statements about two profiles – the profile of the defendant and the profile of a randomly taken individual. It should also be noted that the random match probability is not the measure of (a) the probability that the accused has committed the crime in question, nor of (b) the probability that someone other than the accused has committed the crime, nor of (c) the probability that someone other than the accused is the source of DNA in the crime stain (Koehler 1993). Approaches to calculations of the random match probability for various types of DNA evidence can be found in Balding (2005) and Buckleton et al. (2016).

The Logic of Evidential Proof

When DNA evidence is examined in court, there is one issue that is crucial for determining how much weight should be given to it in the proceedings. The issue, though simple, is not trivial: is it possible to relate the DNA match between the defendant and the crime stain with the probability of him being the perpetrator of the crime (the ultimate probandum)? What needs to be understood and remembered is that only the fact finder can assess this relationship and evaluate the probability of guilt. It is in no way the job of the forensic scientist to comment on this issue. The same is true for the link between the match and the probability that the defendant is the source of DNA in the crime statin.

When evaluating the probative value of DNA evidence in the context of a particular case, the fact finder proceeds by building the following logical chain (given here in a very simplified manner) in reaching the decision on the ultimate probandum:

1. Evidence matches the accused; *following from this:*
2. The accused is the source of evidence; *following from this:*
3. The accused is the perpetrator of the crime.

Every following statement is independent from the previous ones, and it is exclusively the domain of the fact finder to decide whether Statement 2 could be inferred from Statement 1 or Statement 3 from Statement 2. DNA evidence provides information only about the first statement. Using information about the strength of DNA evidence in conjunction with other relevant evidence in the case (to which the DNA scientist is not privy), the fact finder makes a decision whether Statement 2 can be inferred from Statement 1 and Statement 3 from Statement 2. The confusion and misunderstanding of DNA evidence occurs when, by giving information about Statement 1, the scientist expresses an opinion on Statement 2 or even Statement 3. In other words, instead of commenting on the evidence, she comments on the probability of the prosecution (or the defense) hypothesis.

DNA is circumstantial evidence. Circumstantial evidence allows a fact at issue to be proven inferentially rather than directly by providing factual information from which an inference may be drawn on the probability of the fact at issue (Emson 2004). It is for the forensic scientist to provide the facts but for the fact finder to make this inference from them.

The instances of misunderstanding of DNA evidence in the courtroom are far too common and are committed not only by lawyers who in most cases do not have special statistical training but also by forensic scientists who should have been trained in probabilistic reasoning. For example, when using DNA evidence in arriving at the decision as to the guilt of the defendant, the probability that the defendant's DNA profile would match that of the criminal, given that he is innocent, may be confused with the probability that the defendant is innocent, given that his DNA profile matches that of the real criminal.

This particular error of evidence misinterpretation is called "the prosecutor's fallacy." It can be detrimental to the case of the defendant and has been the subject of many appeal cases in various jurisdictions. Other logical errors of interpretation include the defendant's fallacy, the uniqueness fallacy, and several other ones. All of them will be discussed below.

Logical Fallacies and Reasoning Errors in DNA Evidence Interpretation in Reports and the Courtroom

The Prosecutor's Fallacy

The prosecutor's fallacy is a specific case of a logical error of the transposed conditional (Thompson and Schumann 1987) which arises when the probability of DNA evidence under a particular hypothesis is confused with the probability of the hypothesis given DNA evidence. This fallacy tends to favor the prosecution and was a subject of numerous publications in legal literature. In spite of this, it still often happens, especially in cases when DNA evidence is involved.

The error of the transposed conditional is an elementary logical error in which the evidence interpreter assumes that if A implies B, then B implies A. David Balding (2005) gives a simple explanation of this error. Let A denote "a cow" and B denote

"has four legs," then a statement "a cow has four legs" is not the same as "if an animal has four legs it is a cow." A cow, of cause, is one of the possibilities but the animal can also be a sheep, a goat, or a dog.

The reasoning behind the prosecutor's fallacy is similar to this logical error, only in terms of probabilities. This can be illustrated by the following example. If there are 100 four-legged animals living on a farm, of which 10 are cows and 90 are sheep, then the probability of randomly picking a four-legged animal which will be a cow is 0.1 (i.e., every tenth animal is a cow). Assume that the probability of having four legs if you are a cow on the farm is 1 (i.e., every cow on the farm has four legs, disregarding here rare genetic anomalies and accidents leading to amputation of limbs). In such a case, the probability of the statement "a cow has four legs" is 1, while the probability of the statement "a four-legged animal on the farm is a cow" is only 0.1. The first statement is the probability of evidence ("four legs") under a particular hypothesis ("is a cow"); the second statement is a probability of a particular hypothesis ("is a cow") given evidence ("has four legs").

In relation to DNA evidence, the prosecutor's fallacy is to confuse the probability of a match between the accused and the crime stain (Statement 1 in the previous section) with the probability that the accused is the source of the crime stain (Statement 2) or with the probability of accused being guilty of the offence (Statement 3). These two types of prosecutor's fallacy are called the source probability error (the former) and the ultimate issue error (the latter).

The Source Probability Error

The source probability error is the most common type of the prosecutor's fallacy. This error happens when the probability of DNA evidence (e.g., the random match probability) is equated with the probability of the defendant being the source of the crime stain sample. Equating these two probabilities tends to exaggerate the strength of the prosecution hypothesis and is detrimental to the defendant. The information on the source is inferred from the factual information provided by the expert, and it is exclusively the prerogative of the fact finder to make such an inference from adduced evidence. Because of this, it is not for the expert to express her opinion on the source of evidence; this lies in the domain of the fact finder.

Examples of the source probability error are numerous. When the probability of a random match is, say, 1 in 1 billion, the fallacious statement may look like this:

> "It is 1 billion times more likely that Mr X is the source of the biological material found at the crime scene, than an unknown unrelated to him individual"

A correct version of such a statement should always involve a conditional statement of the form "if the defendant were not the source, then the probability . . ." (Balding 2005), for example:

> "If Mr X were not the source of the biological material found at the crime scene then the profiles must match by chance. I estimate that the chance of obtaining matching DNA profiles if the biological material comes from an unknown unrelated to him individual is 1 in 1 billion"

Forensic scientists are specially trained not to commit the errors of transposed conditional in their statements, but the above examples indicate these types of errors are likely to be committed when the scientist is asked about the meaning of evidence in court. A legitimate question arises about the extent to which experts are responsible for the mischaracterization of their scientific testimony by others. Under the pressure and stress of direct and cross-examinations, it may not be reasonable to expect an expert to catch and correct all subtle distortions and misunderstandings expressed within the courtroom (Koehler 1993). To avoid this type of error, one must always remember that the scientist provides the court with the weight of DNA evidence, expressed as the random match probability or the LR, and it is then for the fact finder to evaluate the hypotheses that the defendant is the source of DNA in the crime stain.

The Ultimate Issue Error

Sometimes the source probability error is extended to comment on the probability of the guilt of the defendant. In this case the ultimate issue error is committed.

When scientists report the probability of a random match between the accused and the crime samples as, say, 1 in one million, this may erroneously be interpreted by the fact finder or prosecutor as meaning that the probability of the defendant being innocent is also 1 in one million. This logic is fallacious, but it can still be encountered during criminal trials. An example of this type of erroneous reasoning is the following:

> "If the defendant has a DNA profile matching that obtained from the crime stain and the observed random match probability is 1 in 1 million then, considering the population of the United Kingdom being equal to 65 million people, he will be one of perhaps 33 men in the United Kingdom who share that characteristic. If no fact is known about the defendant, other than that he was in the United Kingdom at the time of the crime, the DNA evidence tells us that there is a statistical probability of 1 in 33 that defendant is the real perpetrator of the crime."

The ultimate issue error often arises when the prosecutor or judge asks the scientist who presents the evidence a question in the form of a transposed conditional. Koehler (1993) gives an example of a dialogue between the prosecutor and the DNA expert witness after he reported a random match probability between the DNA profile of the accused and the crime stain samples as 1 in 5 billion:

> Q.: So that in the event that the accused sitting in this chair would happen to be White, you're telling the members of this jury that there would [be] a one in 5 billion chance that anybody else could have committed the crime; is that correct?
> A.: One in 5 billion, correct.

As illustrated by this dialogue, it is very easy for an expert witness to get caught by such a question and give an erroneous testimony. As well as experts, it is thought that ultimate source errors are committed by the judges and jurors, even when the experts don't commit them (Koehler 1993). As this is often the case, the expert in her testimony has to emphasize to the fact finder that DNA evidence does not provide

any information as to the guilt or otherwise of the accused, but only indicates the likelihood of DNA originatig from another person with the DNA profile identical to that of the accused, should the accused be not the donor of the biological material in the crime stain.

The ultimate issue error also can be found in DNA parentage testing reports. There it equates the Combined Parentage Index (CPI) with the likelihood that the alleged parent is the true biological parent of the child in question (the probability of guilt). The fallacious statement often reads like this:

> "The Combined Parentage Index value of 107,463 indicates that Joe Bloggs is 107,463 times more likely to be the biological father of Jane Bloggs as compared to an untested, unrelated male of the same ethnicity."

The correct statement in this situation should be something like this:

> "The Combined Parentage Index value of 107,463 indicates that the genetic data obtained are 107,463 times more likely if Joe Bloggs and not an untested, unrelated male of the same ethnicity is the biological father of Jane Bloggs."

As with random match probability, CPI is not a measure of guilt but a parameter indicating the likelihood of obtaining the observed results under two contrasting hypotheses: H_p, the alleged parent is the true biological parent of the child, and H_d, an untested unrelated man of the same ethnicity is the true biological parent of the child.

The False-Positive Fallacy

A declared match between the defendant and the crime scene can be either true or false. The false match happens when a forensic scientist erroneously finds "a match" between defendant's DNA profile and the profile obtained from the crime scene when such a match does not actually exist. This type of error is called a "false-positive error." The probability of reporting such an erroneous match is called probability of a false-positive match.

Many consider that for correct evaluation of DNA evidence, it is only important to have information about the value of the random match probability. However, due to a false-positive error, a reported match between the defendant and the crime stain can be false. So, to evaluate the evidence correctly, it is also necessary to have data on the false-positive error rate in the forensic laboratory.

The false-positive fallacy happens when a scientist mistakenly assumes that if the false-positive probability (i.e., the prior probability of declaring a match falsely) is low, then the probability of a false match (i.e., the probability that a declared match is false) in the case in question must also be low (Aitken and Taroni 2004). A scientist may think that if the probability of a false-positive match in the laboratory is 0.001, then the probability of a true match reported in a particular case is 0.999. This fallacy, which is yet another version of the prosecutor's fallacy, arises from mistakenly equating the conditional probability of a match being reported when the samples do not match (the false-positive probability) with the probability that the samples do

not match when a match has been reported (Aitken and Taroni 2004). These two probabilities are not the same.

The false-positive probability depends only on the probability of genotyping (and/or human) error in the laboratory and is the probability that a match between the accused's sample and the crime sample will be reported when there is no match. The probability that the samples do not match when a match has been reported depends on both the probability of genotyping (and/or human) error and the prior odds that such a match will occur. The assumption that the false-positive probability is equal to the probability that the samples do not match when the match is reported is fallacious because it ignores the prior probability that the accused's profile matches the crime profile (Aitken and Taroni 2004). In a simple case, assuming there are no false negatives, the probability $\Pr(M \mid R)$ of a true match M between the accused and the crime sample, given that a true match R has been reported, can be calculated using the following formula (Aitken and Taroni 2004):

$$\Pr(M|R) = \frac{1}{1 + k\Pr(R|\bar{M})}$$

where k are the prior odds that the accused's profile matches that from the crime scene and $\Pr(R \mid \overline{M})$ is the false-positive probability. A more complex formula, which takes into account the probability of false negatives, can be found in the same work.

Aitken and Taroni (2004) provide a compelling example illustrating the false-positive fallacy. Let us assume that the prior odds of the accused matching the crime sample is 1 in 1000, because the accused was identified by a DNA search and appeared initially to be an unlikely perpetrator. The match between the accused and the crime stain was obtained. Further, let the probability of the false-positive match in the forensic laboratory which analyzed the samples be 0.01. The false-positive fallacy would suggest that because the false-positive error in the laboratory is 0.01, the probability of a match reported correctly in this particular case would be 0.99. In reality, this probability is:

$$\Pr(M \mid R) = \frac{1}{1 + 1000 \times 0.01} = 0.0909$$

which is almost 11 times smaller than that assumed under the false-positive fallacy.

How Not to Commit the Prosecution Fallacy

The error of transposed conditional is very common. It happens to inexperienced and trained scientists, especially under immense pressure of court proceedings. The error stems from misinterpretation of the meaning of random match probability and also misunderstanding of what the expert can and cannot comment on during criminal proceedings. Very often, an incorrectly asked question confuses even a well-trained testifying scientist to commit the error. To avoid the prosecution fallacy, the scientist has to remember that:

1. DNA evidence is circumstantial evidence and as such allows only *indirect inference* about the fact at issue (which can be, e.g., "the defendant is the source of biological material in the crime stain").
2. The weight of DNA evidence (LR) *only* indicates the *probability of obtaining evidence* under two contrasting propositions (e.g., H_p, "the defendant is the source of biological material in the crime stain," and *Hd*, "an untested unrelated man of the same ethnicity and not the defendant is the source of biological material in the crime stain") and *not the probability of the propositions*.
3. The scientist is not the fact finder and must never comment directly on the fact at issue (i.e., express her opinion as to whether the defendant is the source of biological material in the crime stain). Only the fact finder can directly draw conclusion as to the fact at issue. This is done by taking into account all available evidence, including DNA evidence. However, the scientist can provide an *indirect opinion* as to the fact at issue by giving estimation of the likelihood of observing DNA evidence when comparing two contrasting propositions.
4. The scientist must not comment on which of the evaluated hypotheses is more or less likely. She provides the probability of evidence, and the fact finder deals with interpretation.

The Defendant's Fallacy

Besides the prosecutor's fallacy, the other logical error commonly heard in the courtroom is the defendant's fallacy (Thompson and Schumann 1987) which, as the name states, usually favors the defendant. The fallacy stems from the (erroneous) assumption that in a given population, anyone with the profile matching that of the crime stain is as likely to have left the sample as the defendant. The fallacious reasoning goes like this:

> "The crime was committed in the UK by a male. Suppose the total UK male population is 33 million. The random match probability reported for the case is 1 in 1 million. As there are 33 people in the UK who have the DNA profile matching that from the crime scene, the probability that the defendant is the purpertrator of the crime is 1 in 33 or only 3%."

Some may go even further and argue that because the number of possible culprits is high, the probative value of DNA evidence is low (only 3% in the above case) and it should be ignored.

Strictly speaking, the defendant's fallacy is not a fallacy in the sense of the prosecutor's fallacy. The logic behind the defendant's fallacy's reasoning would be correct if everyone indeed was equally likely to be guilty of the crime in question (which of course contradicts the main principle of criminal law – innocent until proven guilty). In practice, though, this assumption is very unlikely. Even when there is no evidence directly incriminating the defendant of the crime, there is usually enough background information related to the location and nature of the crime to

exclude some or even most individuals from the number of potential perpetrators and make others more plausible suspects. In addition, the DNA evidence in the above example has reduced the number of possible perpetrators from 33 million down to 33 individuals, the accused being one of them, and increased the odds in favor of his guilt 1 million times. In other words, the evidence is 1 million times more likely if the accused is the perpetrator of the crime than if he is innocent.

The Uniqueness Fallacy

Another fallacy which often crops up in cases when DNA evidence is used is the uniqueness fallacy. This fallacy arises when the match probability for an unrelated perpetrator is small enough to imply that no other individual with the same profile could be found in the relevant population. For example:

> "The crime was committed in the UK by a male. The random match probability between the defendant and the crime stain is 1 in 1 billion (billion is one thousand million). Let assume the total male population of the UK to be 33million people. Considering the size of the population it can be safely assumed that defendant's DNA profile is unique in the UK, and he is the only source of DNA in the crime stain."

The uniqueness fallacy misinterprets the role of forensic scientist in criminal proceedings and the role of statistics in forensic inference. It stems from taking the value of random match probability literally. Because of relatedness, the actual number of people who might have the same DNA profile as the defendant may be higher. In the above case, it is wrong to assume that there are 33 million males in the UK who are unrelated to each other. Considering the meaning of the word "unrelated" as "second cousin," the number of unrelated males is significantly lower and is probably in the region of 3–5 million.

Instead of having a literal meaning, the value of the random match probability indicates how strong the evidence against the defendant is and how much weight the fact finder should assign to it. If all 7 billion people living on the planet were not related to each other, probably, a match probability which is several orders of magnitude smaller than the 1 in 7 billion would be sufficient to judge that a profile is unique and no identical one can be found. However, this is not the case. All people are related to each other to a certain degree, especially island populations like the one in the UK. Because of this, the chances of finding another individual with an identical DNA profile are not negligible, making the claim of uniqueness statistically unsubstantiated, even when calculations of the random match accounted for co-ancestry. When the match probability is 1 in 1 billion, it does not mean that we need to test 1 billion people to find the matching profile. There is a high probability that the number of people needed to be tested to find the matching profile is significantly smaller. This is discussed in more detail in the next section. But even in the hypothetical case when all people on the planet are not

related to each other, because of the non-DNA evidence, there seems to be no satisfactory way for an expert witness to address the question of uniqueness in court (Balding 1993).

The other problem with the uniqueness fallacy is related to misunderstanding of the law of evidence, the mechanics of proof in criminal proceedings and what is the domain of the fact finder during a criminal trial. In expressing uniqueness fallacy, the scientist takes on herself the role of the fact finder to draw conclusions as to the identity of the source. In making a statement on uniqueness of the DNA profile, the scientist has to account for the size of the relevant population of potential perpetrators. Because the size of this population depends not only on demographic characteristics of the total population but on other evidence not available to the forensic scientist, in expressing her opinion on uniqueness, she will always be using the population of the wrong size. In addition, it is not in the domain of the forensic scientist to regard some sections of the population as more capable of committing the crime than the other ones.

The Probability of another Match Error

The logical flaw of the probability of another match error is similar to the uniqueness fallacy. This error stems from a mistaken belief that the probability of a random match is the same as the probability of finding another person in the population with the same DNA profile. The logical error would go as follows:

> "The random match probability between the defendant and the crime stain is 1 in 1 million meaning the probability of finding a male with the DNA profile identical to that of the defendant is 1 in 1 million."

The logical error here is to think that the small value of the match probability between the defendant and the crime stain implies a similar small probability of finding another man with the same genetic characteristics. The mistake is to take the probability that a randomly selected person would have the profile identical to the defendant (the match probability of 1 in 1 million) for the probability that there exists someone else in the population who has the same profile as the defendant. The latter probability may be greater than the former.

Following Aitken and Taroni (2004), let N denote the size of the relevant population and F the match probability. Then the probability that a randomly selected person will not match the crime profile will be $1 - F$, while the probability that not a single member of the relevant population with the size N will match the crime profile will be $(1 - F)^N$. It follows from this that the complement probability θ, that at least one match will be found in the population, is $\theta = 1 - (1 - F)^N$.

If the population N is 1 million and the random match probability F as 1 in 1 million, the probability of finding at least one other person out of the million people in the population with matching DNA profile will be:

$$\theta = 1 - (1 - F)^N = 1 - (1 - 1/1,000,000)^{1,000,000} = 0.63212 \text{or} 63.21\%$$

When F is 1 in one million, it may be extremely unlikely to randomly pick a person whose DNA profile matches that of the defendant out of a million people, but the probability that at least one such man exists is 63.21%.

The probability of finding another person with matching DNA profile is a good argument when talking about uniqueness. In the UK, when a complete match is observed between the defendant and the crime stain, the match probability is conventionally reported to be "less than 1 in 1 billion (a billion is a thousand million)." By calculating θ we can get an estimation of how unique a DNA profile with such low match probability would be in the UK. Let us assume the crime was committed by a man; the total male population of the UK N is 33 million, and the match probability F is 1 in 1 billion. Then:

$$\theta = 1 - (1 - F)^N = 1 - (1 - 1/1,000,000,000)^{33,000,000} = 0.03246 \text{or} 3.25\%$$

Can we consider a DNA profile to be unique when there is 3.25% probability of finding another man in the UK with matching genetic characteristics?

The Numerical Conversion Error

Sometimes, the random match probability may be incorrectly interpreted as the number of people who need to be tested until the second identical profile is found. A conclusion that when the probability of a random match is, say, 1 in 33 million, the number of people who need to be tested until the matching profile is found is at least 33 million is fallacious and is known as the numeric conversion error. This error exaggerates the number of people who would need to be tested before a match may be expected and thus exaggerates the probative value of DNA evidence (Koehler 1993).

As in the previous example, following Aitken and Taroni (2004), let N denote the relevant population, F be the match probability between the defendant and a particular crime profile, and n be the minimum number of people who would have to be tested before another match was found. The numerical conversion error is to equate N with n.

As above, the probability that a randomly selected person will not match the crime profile will be $1 - F$, while the probability that, for n randomly selected individuals, none will match the crime profile will be $(1 - F)^n$. The probability Pr *(M)* that at least one match will be found is $1 - (1 - F)^n$. For a match to be more likely than not, this probability Pr*(M)* has to be greater than 0.5, and so:

$$(1 - F)^n < 0.5$$

Solving this inequality for n, we obtain the formula for calculating the number of people needed to be tested before we might expect the match to be more likely than not (i.e., with the probability greater than 0.5):

$$n = \frac{\log(1 - 0.5)}{\log(1 - F)},$$

and the generalized formula will be:

$$n = \frac{\log(1 - (1 - \Pr(M))}{\log(1 - F)}.$$

Taking F to be 1 in one million, we would expect to find a match more likely than not after testing 694,147 people. At the same time, the number of people needed to be tested to ensure that finding a match is guaranteed (say, with the probability of 0.99) is a staggering 4.6 million people!

Other Interpretation Difficulties

Besides the abovementioned errors and fallacies, forensic reports sometimes contain statements that have a potential for causing misinterpretation of DNA evidence. Some of these statements are discussed below.

Relative Frequency of Occurrence

Sometimes an expert statement contains, instead of the match probability, the reference to the frequency of occurrence of a particular DNA profile in the population. In such a case, the report includes a statement about the DNA evidence along the following lines:

> "The DNA profile of a particular type occurs in about one person in 100,000 of the population."

There are several major objections to this approach (Aitken and Taroni 2004):

1. When the relevant population is greater than the inversed frequency of occurrence, then it is possible by this approach to estimate the number of potential perpetrators and use DNA evidence as prior odds when evaluating the remaining evidence in this case. For example, when the size of the relevant population is one million and the frequency of the DNA profile is 1 in 100,000, then the number of potential perpetrators will be 1000,000 × (1/100,000) = 10. However, if the inversed frequency of occurrence is bigger than the relevant population (the frequency is, say, 1 in 1 billion, while the relevant population is one million), we would not expect to see another profile like this, which makes it impossible to combine DNA evidence with other evidence that might provide the support or an alibi for the defense case.
2. When a relative of the accused is considered as a potential perpetrator, a statement like:

> "The DNA profile in question occurs in about one brother in 400"

does not make sense. The scientist has to find alternative ways of expressing the meaning of this statement, which may be even more confusing to the fact finder.

3. When assessing DNA evidence, there are at least two conflicting hypotheses – the prosecution hypothesis, which states that the accused is the donor of the crime stain, and the defense hypothesis, stating that the accused is not the donor of the crime stain but owing to chance his/her DNA profile matches that of the real perpetrator. This means that there is another person in the population with the profile matching that of the accused. In contrast to frequency, the concept of a match involves two profiles which permits evaluating both prosecution and defense theories. It is not the relative frequency of a particular DNA profile or the probability of finding a particular profile in the population that is essential for a criminal identification but rather the probability of finding a particular DNA profile in the population, given that the accused has this profile. It is important to emphasize that in human populations, because of relatedness and population subdivisions, the match probability exceeds the probability of finding a particular profile in the population (Balding 2005). Thus, the use of DNA profile frequency instead of the match probability favors prosecution and should be avoided in criminal trials.
4. When a mixed profile is obtained from the crime stain or in cases of paternity testing or identification of victims of mass disasters, it is impossible to use the profile frequency to state the value of DNA evidence as these types of profiles cannot be expressed in the form of "one in. . .."

The "Could be" or "Could Have Come/Originated From" Approach

Often, forensic scientists use statements regarding the source of DNA evidence of the form

> ". . . it is likely that the accused could be the donor of the DNA recovered from the left pocket of the trousers. . ."

or

> ". . . in my opinion the blood stains on the track suit bottoms could have come from the victim. . ."

and so on.

This type of statement may be treated as the source probability error as they express a view on the probability of a hypothesis (the defendant is the donor of the crime stain) and can be understood as establishing a proven association between the crime and the accused. It should be noted, however, that "Could have come/originated from" does not preclude other possible sources, and there may well be other, even more plausible, explanations as to the origin of biological material in the crime stain that have not been provided by the scientist. In addition, this approach does not give any indication of the likelihood that the defendant is the actual source of the evidence (Aitken et al. 2010).

The "Could Not Be Excluded" Approach

Another type of statement commonly found in forensic reports is:

> ". . .. the accused cannot be excluded from being a donor of the crime stain. . .."

In some cases, especially when dealing with mixtures, the exclusion probabilities approach is used for evaluating DNA evidence. Under this approach, the probability that a random man or a contributor to the mix will be excluded as the donor of the DNA from the crime scene is calculated. However, such a probability only states the proportion of the population which would be excluded; it is thus the measure of efficacy of a particular test (Robertson and Vignaux 1992) as it answers the question "how likely is the individual to be the donor of the crime stain?" However, the other question is much more important in criminal identification, namely, "how much more likely is the evidence if the accused is the donor of the crime stain than a randomly taken individual?" (Aitken and Taroni 2004). The "Could not be excluded" approach does not give an answer to this question, and the probability of exclusion is not relevant for evaluating DNA evidence in such a case.

"Consistent with" Approach

Sometimes a scientist during her testimony may state that the evidence is "consistent with" a particular proposition, for example:

> "Finding DNA matching that of the defendant on a vaginal swab is consistent with the theory that the defendant had a vaginal intercourse with the complainant."

Again, like with "could be" approach above, this type of statement may be treated as the transposed conditional error as it could be expressing a view on the probability of a hypothesis (the defendant had a vaginal intercourse with the complainant) and can be understood as establishing a proven association between the crime and the accused. Also, to say that evidence is "consistent with" a particular proposition means only that the stated proposition cannot be excluded by the evidence and says nothing about how likely the proposition is to be true (Aitken et al. 2010).

The Association Fallacy

When evaluating Source Level Propositions, the scientist often considers questions whether a particular bodily fluid or cellular type can be the source of DNA. Association of a DNA profile with the tissue source could make the evidence seemingly more probative than just conformation of a DNA profile without providing the evidence as to the bodily fluid origin. The "association fallacy" is the assumption that the observed DNA profile has come from a particular bodily fluid (e.g., semen) on the strength of presumptive (or RNA) tests, verified by "expert opinion." The fallacy here is that the observation of a body fluid and the detection of a DNA profile are two separate tests, and it is erroneous to assume that there is a dependency between two observations or events. It cannot be implied with certainty

that the body fluid and the DNA have the same source. By extension, this fallacy also describes wrongful association of the presence of a bodily fluid (e.g., sperm), with an activity (e.g., vaginal intercourse) (Gill 2014).

The Use of the Verbal Scale to Express the Strength of DNA Evidence

Forensic scientists sometimes accompany the likelihood ratio with a qualitative verbal explanation of the support that should be given, in their view, to one of the competing hypotheses (usually the prosecution hypothesis). This is typically done by assigning a non-numerical weight to LR using one of several verbal scales proposed for this purpose (Buckleton 2005). For example, LR of 10,000 can be descried as showing "strong scientific support" and LR of 100,000 – "very strong scientific support" to a particular proposition. When using the verbal scale, the scientist usually provides an explanation like:

> "In expressing the evidential significance of my findings, I have used the following scale of scientific support: no support, weak, moderate, moderately strong, strong, very strong, extremely strong support."

The idea behind the verbal explanation is to assist the fact finder in understanding how much weight should be given to the hypothesis that the accused is the donor of the DNA in the crime stain. By admittance of forensic scientists, this scale is arbitrary and hence subjective (Buckleton 2005).

When presenting results using the verbal scale, the differences in meaning between "moderate support" and "moderately strong support" or "very strong support" and "extremely strong support" together with corresponding LR values have to be explained to the fact finder and the prosecution and defense. These meaning of the wording is not standardized and expresses either the subjective opinion of the testifying scientist or the subjective opinion of the scientist who developed this verbal scale.

Last but not least, by expressing the strength of the scientific support to one of the alternative hypotheses, the scientist appears to be commenting on the issue of whether or not the defendant is the source of the crime stain. This is the prerogative of the fact finder who is guided in his decision by the statistical information provided by the forensic scientist and should not be affected by her personal opinion as to the strength of scientific support that should be given to one of the competing propositions.

Statements about Identical Twins

Sometimes we come across statements which are now rarely made in court but often in the press, that

> ". . .only identical twins have identical DNA profiles. . ."

or

> ". . .no two people in the world have matching DNA profiles apart from identical twins. . .."

These statements are not correct. Identical twins indeed have matching DNA profiles; however, other people may have matching profiles owing to chance or because they are related. The chances of such a match happening are expressed by a random match probability and, however rare, a match between two unrelated individuals can happen.

Conclusions: Why Misinterpretation of Evidence Happens and What Can Be Done About It?

Most errors of DNA evidence interpretation tend to exaggerate the probative value of the evidence. Koehler (1993) in his review of interpretation errors addressed this issue. He investigated several potential factors, such as deliberate exaggeration of the probative value by scientists who wish to "puff up" the utility of their science, or by prosecutors, determined to win their case. He concluded that the most plausible explanation for this sort of error is ignorance in the legal profession of statistics and rules for evidence interpretation.

In a review of court cases conducted by Koehler, it transpired that DNA experts generally begin the interpretation part of their testimony with statements about population frequencies and comparison with a "random man." Although, as we have seen, there may be errors in these statements, they are not deliberate, and the very use of such statements indicate that the scientist tries to give an unbiased opinion. The errors in the testimony begin to appear in most cases only after the expert re-describes her findings in response to the questions from the prosecutor. It has to be kept in mind that while, for a prosecutor, a court hearing is part of the daily job, the expert witness is often under immense pressure both because of her role in the proceedings and also because of the adversarial nature of the criminal trial. In such circumstances, even highly experienced scientists can make errors in answering a logically erroneous question.

Because very few lawyers are trained in statistics and evidence interpretation, they may misunderstand the subtleties of probabilistic inference and have trouble differentiating probabilistic information from probabilistic hypotheses that the information suggests (Koehler 1993). All this indicates that the solution to the problem is specialized training of lawyers in matters dealing with evidence interpretation and probabilistic inference.

The issue of poor training of members of the legal profession has long being recognized by forensic statistical community. To somewhat remedy the situation in the UK, the Royal Statistical Society (RSS) has published four excellent guides (Aitken et al. 2010; Puch-Solis et al. 2012; Roberts and Aitken 2014; Jackson et al. 2015) which are freely available on the RSS's web site. The guides look at communicating and interpreting statistical evidence during court proceedings and are intended to assist judges, lawyers, and expert witnesses in coping with the demands of modern criminal litigation.

References

Aitken CGG, Taroni F (2004) Statistics and the evaluation of evidence for forensic scientists. Wiley, Chichester

Aitken C, Roberts P, Jackson G (2010) Practitioner guide no. 1. Fundamentals of probability and statistical evidence in criminal proceedings. Guidance for judges, lawyers, forensic scientists and expert witnesses. Royal Statistical Society's working group on statistics and the law. https://rss.org.uk/news-publication/publications/our-research/. Accessed 22 Nov 2020

Balding DJ (1993) When can a DNA profile be regarded as unique? Sci Justice 39:257–260

Balding DJ (2005) Weight-of-evidence for forensic DNA profiles. Wiley, Chichester

Buckleton J (2005) A framework for interpreting evidence. In: Buckleton J, Triggs CM, Walsh S (eds) Forensic DNA evidence interpretation. CRC Press, Boca Roca, pp 27–63

Buckleton JS, Bright J-A, Taylor D (eds) (2016) Forensic DNA evidence interpretation, 2nd edn. CRC Press, New York

Emson R (2004) Evidence, 2nd edn. Palgrave Macmillan, Basingstoke/Hampshire

Evett IW, Weir BS (1998) Interpreting DNA evidence. Sinauer Associates, Sunderland

Gill P (2014) Misleading DNA evidence: reasons for miscarriages of justice. Academic, London

Jackson G, Aitken C, Roberts P (2015) Practitioner guide no. 4. Case assessment and interpretation of expert evidence. Guidance for judges, lawyers, forensic scientists and expert witnesses. Royal Statistical Society's working group on statistics and the law. https://rss.org.uk/news-publication/publications/our-research/. Accessed 22 Nov 2020

Koehler JJ (1993) Error and exaggeration in the presentation of DNA evidence. Jurimetrics J 34:21–39

Puch-Solis R, Roberts P, Pope S, Aitken C (2012) Practitioner guide no. 2. Assessing the probative value of DNA evidence. Guidance for judges, lawyers, forensic scientists and expert witnesses. Royal Statistical Society's working group on statistics and the law. https://rss.org.uk/news-publication/publications/our-research/. Accessed 22 Nov 2020

Roberts P, Aitken C (2014) Practitioner guide no. 3. The logic of forensic proof: inferential reasoning in criminal evidence and forensic science. Guidance for judges, lawyers, forensic scientists and expert witnesses. Royal Statistical Society's working group on statistics and the law. https://rss.org.uk/news-publication/publications/our-research/. Accessed 22 Nov 2020

Robertson B, Vignaux GA (1992) Unhelpful evidence in paternity cases. New Zealand Law J 9:315–317

Robertson B, Vignaux GA (1995) Interpreting evidence. Wiley, Chichester

Thompson WC, Schumann EL (1987) Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy. Law Hum Behav 11:167–187

Part VI

Non-human Case Studies

Wildlife DNA Profiling and Its Forensic Relevance

37

Ulhas Gondhali and Aditi Mishra

Contents

Abstract

Wildlife DNA forensics is a relatively new field emerging to tackle challenges in wildlife crime investigation. It is an applied field which is a combination of conservation genetics and forensic genetics. DNA evidences are crucial and hold great evidential value in human identification. DNA evidence analysis techniques have been in focus since years. The present technological advantage is achieved through years of research and development in molecular techniques. The process of evidence analysis and reporting in DNA-based evidence in human identification is not much different than wildlife forensic analysis. There may be

U. Gondhali (✉)
Jindal Institute of Behavioural Sciences, O.P. Jindal Global University, Sonipat, India
e-mail: ugondhali@jgu.edu.in

A. Mishra
School of Forensic Science and Risk Management, National Security and Police University of India, Rashtriya Raksha University, Gandhinagar, Gujarat, India
e-mail: aditimishra@rsu.ac.in

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_41

difference in case assessment and evaluation of scientific data, but the fundamental principles of forensic investigation remain same.

The nature of wildlife crimes is different, and the laws controlling it are as well. The legality of any act is decided by a number of factors, such as the species involved, geographic location, source, and age. DNA evidence analysis helps wildlife law enforcement to answer important case-related questions, which are based on the above factors. Wildlife DNA profiling is essential to meet the increasing need of investigative tools in wildlife crime investigation.

Keywords

Wildlife DNA profiling · Wildlife crime investigation · Species identification · Geographic identification

Introduction

There is a global consensus on stopping the ever-increasing environmental crimes and overexploitation of the available natural resources. The international bodies such as Interpol, which is working to control such crimes on a larger scale, have also considered wildlife crimes as one of the prime issues. It has expressed its concerns on the increasing risk on the worlds of flora and fauna from criminal exploitation.

The CITES (the Convention on International Trade in Endangered Species of Wild Fauna and Flora), which is an international agreement that controls the trade of wildlife flora and fauna to ensure their survival, defines wildlife crimes as the taking, trading (supplying, selling, or trafficking), importing, exporting, processing, possessing, obtaining, and consumption of wild fauna and flora, including timber and other forest products, in contravention of national or international law.

Wildlife crime is unlike human-related crimes. Wildlife crimes such as poaching of rare species such as pangolins and subsequent trafficking of their body parts are complex issues. Such crimes often occur in remote places. Unlike human-related crime, there is no social taboo involved. The low risks and high rewards are also encouraging factors in such crimes. The financial rewards for such crimes can be highlighted in several examples. It includes poaching of musk deer (*Moschus* sp.). Male musk deer carry a secreting gland called "musk pod." It is used in certain traditional Chinese medicine products and a variety of cosmetics because of its scent. Few grams of musk pod could fetch hundreds of dollars in international markets. The international ivory market is also one of the biggest. Often African rhinos and elephants are poached as a source of illegal ivory. The high demand for ivory products in international markets attracts huge prices for such catches.

Often multiple actors are involved in wildlife crimes, and such crimes are executed at local and international levels. The involvement of local as well as international nexus is because of the intricate nature of source and consumer. Illegal wildlife trade is highly organized as sometimes it requires to bypass many

international borders without alerting authorities. For law enforcement agencies to track such trades requires well-operated teams, international cooperation, and scientific resources including forensics. Time and time again, a common issue faced by the authorities is related to the identification of such ceased products or articles generated from illegal wildlife trade. The confirmation of the source of origin from a protected wildlife species is important to demonstrate the illegal nature of an activity.

Forensic science is an applied field. The investigative tools of forensic science are end-user base. Usually, these tools are applied for human-related crimes. A specialized branch of forensic science that deals with nonhuman crimes or wildlife crime is called wildlife forensics. From an investigator's perspective in wildlife crimes, there are often no victims to give testimony. Additionally, species identification becomes an important task. DNA analysis is performed routinely in wildlife forensic investigations, and it is often compared with conservation genetics. Although the nature of work is similar, the complexity and challenges faced by the wildlife forensic investigators are different. The investigation process from a crime scene to courtroom trial is similar in comparison to routine human-related forensic analysis. However, most well-equipped forensic laboratories and associated scientists do not perform wildlife-related sample test. This is due to the type of samples involved and analysis techniques to perform, which require different expertise to those acquired by common forensic laboratories.

The type of samples analyzed by wildlife forensic scientists is vast. It ranges from a whole body of an animal in both live and dead condition, as well as in all stages of life (animals seized by customs authorities, wildlife protection cell): skin, shells, or exoskeleton (horns, skeletal remains, nails) and animal body parts (genital organs, heads, limbs, internal organs). Additionally, some pieces of evidence can also involve traces of DNA from body fluids or highly processed animal body parts. The challenges are faced when samples carry mixtures of processed animal parts, such as traditional Chinese medicine (TCM).

Analysis performed by labs catering to wildlife forensic analysis is directed by the investigative questions asked by the investigator concerning the case involved. The frequently asked questions can be put into the following five categories (Ogden and Jen 2015):

1. Species of origin
2. Geographic origin of a confiscated article/evidence
3. Wild or captive source
4. Individual origin of an article
5. Age of an article

Scientists answer these specific questions by performing standardized techniques. For example, in queries related to species of origin, a mitochondrial DNA analysis for a conserved region, such as region coding for cytochrome b and cytochrome oxidase subunit I in mammalian specimens, can be performed. If a query is for an individual origin, then tests for nuclear DNA are more suitable.

Identification of Species of Origin

One of the common most queries is related to the identification of species of origin of a confiscated article or specimen. Species identification can also be considered as a primary step before performing individual identification of an animal, because the techniques performed for individual identification, such as short tandem repeats (STR) and single nucleotide polymorphism (SNP), require prior information on the DNA sample being tested. Therefore, individual identification is performed after species identification using DNA techniques if the species of origin is not clear otherwise.

In some cases, individual identification is not required; only species identification is sufficient to verify the commission of a crime, for example, if a suspected consignment is tracked and confiscated by forest officials in a transporter carrying goods out of a forest area. If the suspected product contains monitor lizard parts, it does not matter which monitor lizard. It is protected under Indian Wildlife Protection Act, 1972, and internationally under CITES.

Similarly, if a meat product is labeled as 100% pork and found after testing it contains an amount of beef, it doesn't matter from which cow. It is a type of a food fraud, where a product is sold under a different misleading labeled content.

A species is a group of organisms having common characteristics and usually mate in a group to produce fertile offspring. A group of organisms belonging to a species shares a large amount of their DNA. On a phylogenetic tree, closely spaced species will share more DNA than the distant ones. Species testing gains more when dealing with cryptic species and phenotypic plasticity. Cryptic speciation is a process where different organisms belonging to different species share morphological similarities. For example, the skipper butterfly (*Astraptes fulgerator*), a widely distributed butterfly that was reported in 1975, is confused for a single species for a long time. It was tested for species identification with recently invented DNA barcoding technology. It was found to be ten different species sharing similar morphological characteristics (Hebert et al. 2004). On the contrary "phenotypic plasticity," which is also referred to as "polyphenism," is when a single species is showing different morphological characteristics influenced by a different environment. The European map butterfly (*Araschnia levana*) is a good example of phenotypic plasticity: the species displays the morphological difference in spring and summer season.

To apply DNA-based identification for species testing, the targeted locus of DNA should possess some qualities to qualify for species testing. A DNA test performed with such locus should be able to give constant results. That is, it should be able to identify species correctly and to be able to differentiate a species from another. To obtain such results, the interspecific variation of a targeted locus should be more for interspecific differentiation, and, at the same time, the intraspecific variation should be less to correctly identify a species. Species with a recent diversion on the phylogenetic tree would be challenging to differentiate with this method. Also, species identification can be problematic for animals that are produced on inbreeding among closely related species, especially when one species is protected and another

is not, for example, the problems related to differentiation of wild boar from domestic pigs: There have been constant appeals from farmers and their representative bodies in India to declare wild boar as vermin, to ease the strict laws on hunting of the species. As per India's Wildlife Protection Act (WPA), 1972, wild boars are protected under Schedule III, and anything related to its poaching is illegal without the forest department's permission. Wild boars have been causing problems to farmers of certain states of India. In such a situation, more complication is caused by inbreeding of wild boars and domestic pigs. There is evidence that the two species have been inbreeding again (Frantz et al. 2015). Legally it gets challenging to identify a species where inbreeding is common and one species is protected and another is not. It makes prosecution challenging under WPA, 1972.

Mitochondrial DNA (mtDNA) is preferred for species identification for its advantages over nuclear DNA. mtDNA is present in high copy number, and it is also robust because of its membrane protection. It also shows a greater divergence in closely related species as compared to nuclear DNA. This is because it lacks correction enzymes that are present in nuclear DNA. The absence of correction enzymes will lead to higher divergence in closely related species because of uncorrected mis-insertions. Also, the maternal inheritance of the mtDNA helps in linking members of the same species of one inheritance.

There are other practical reasons to pick mtDNA as well. mtDNA has been used for phylogenetic studies for a very long time, and because of that, there is a great amount of data available in the scientific literature. There are universal primers available in the market which ease the process of identification. The high copy number of mtDNA helps in the forensic identification process since often samples are in processed or degraded forms.

Species Testing

Samples seized by law enforcement personnel that require identification of source species are usually of two categories. A sample could be sourced from a single species such as ivory, hair, skin, and skeletonized remains. Any suitable gene for species identification can be selected. The method of identification depends on the type of sample.

Sequencing

There are many tests available for species testing. Out of which some have relied upon more as compared to others. DNA sequencing is one such technique that is performed routinely in labs. This method is also approved by the International Society for Forensic Genetics (ISFG) (Bär et al. 2000). If a sample is suspected to be sourced from a single source, then, using a universal primer, a small section of DNA is amplified. Usually, mitochondrial DNA is preferred. The amplified products are sequenced. The sequencing will result in a form of an electropherogram, in which sequence of bases is identified. An electropherogram will be interpreted and translated using the software. It's important to verify the sequence of a fragment; this can

be achieved by using a reverse primer and corresponding it with the forward read. Another option is multiple time sequencing of a fragment to verify the sequence.

Species Specific Primers

In cases where a mixture of DNA samples of different species is required to analyze for species identification, universal primers are not suitable. Universal primers that are applicable generally around species will fail to obtain separate profiles. At molecular level, two species are largely similar other than specific DNA sites. These minor sites or bases make all the difference. These minute differences need to be identified to separately identify species from a mixture of DNA. Separate identification of species in a mixture of DNA can be performed using species-specific primers. These primers are specifically designed to target unique sites of a species to identify it correctly. These primers provide an advantage over traditional sequencing methods, which include less time- and cost consuming; simultaneously identification of multiple species, species identification in mix sample, and degraded DNA samples may still provide results.

Species-specific primers are developed through known sequence information about the species for identification. A variation that is particular to a species is identified, and primers are designed on the basis of this variation. The basis of this identification method is by aiming variable bases for primers, and products will not be formed unless the targeted species is present. Usually, one primer of a pair is fluorescent labeled to be detected by a genetic analyzer. Species-specific primers can be used in several ways. A couple of species-specific primers can be used together. This can help to increase the specificity but can relatively increase the cost. This is because of the use of synthetic dyes, as it is difficult to synthesize labeled primers. A combination of one species-specific primer and a universal can be used, which gives an advantage of multiplexing several species-specific primer with one universal primer. By labeling only the universal primer, its relative cost can be reduced for detection of multiple species.

Species-specific primers also have disadvantages like other tests. It has the ability to detect only one species; if a mutation occurs on the binding site, it may give false-negative results due to the inability to bind to the targeted site. This can be overcome by using multiple primers. It is considered very unlikely that multiple mutations will occur, blocking both the binding sites. Multiple primers also give another advantage: if one primer accidentally binds to a non-targeted species, the other primer would not be expected to produce a product.

Common Locus Used for Species Identification

The common most loci in mitochondrial DNA-based species identification in mammals are the cytochrome b (Cyt b) gene and cytochrome oxidase I (COI). Other mitochondrial genes have been used, such as the two ribosomal RNA genes, the control region (or D-loop), and subunits of mitochondrially encoded NADH dehydrogenase.

Cytochrome b (Cyt b)

Cytochrome b gene is responsible for the synthesis of protein cytochrome b. It is a part of a protein complex called complex III. This protein complex has a function of mitochondrial oxidative phosphorylation. The energy required for cell functioning in the form of ATPs is generated through oxidative phosphorylation. In this process, oxygen and simple sugars are consumed to create the end product, which is adenosine triphosphate (ATP). The length of Cyt b is about 380 amino acids, and it varies with species. It starts with conserved methionine codon and lasts with AGA stop codon. The role of Cyt b as a protein is to be a part of an electron transport chain in the mitochondria. It sits within the inner mitochondrial region and lengths it up to eight times.

Within the transmembrane protein, several amino acid replacements are exchanged with diverse hydrophobic residues. Variation of amino acids, such as valine, leucine, or isoleucine, at a particular site can be observed among different species; they show similar properties. The position of the gene on the human mitochondrial genome is 14,747 to 15,887. Due to a change in the length of the HVI region of the mitochondrial genome, different species will show a change in the numbering of gene location. The presence of repetitive elements could also be the reason for a change in the gene location. The length of the gene is 1140 bp long, but only 400-bp-long initial stretch is employed (Irwin et al. 1991).

Cytochrome Oxidase I (COI)

Cytochrome oxidase is found in the inner part of the mitochondrial membrane. The protein complex is composed of a total of 13 protein subunits, out of which ten are encoded in the nuclear DNA and the rest three are part of mitochondrial DNA. Cytochrome oxidase I or subunit I is one of these three genes. It's located between 5904 and 7446 in the human mitochondrial genome. An initial 658-bp-long region was proposed as a universal region for species identification in 2003 by a team of scientists and termed this region as barcode (Hebert et al. 2003). A database is maintained of a specified cytochrome oxidase I region of species around the world. Various species have been identified using DNA barcoding that includes various species of fish (Ward et al. 2009) and avian (Hebert et al. 2004).

The Control Region (or D-Loop)

It's a noncoding region present on a mitochondrial genome. Its name is derived from the unique displacement loop structure created by a nascent short heavy strand that displaces the parental strand. It houses regulatory elements that are essential for replication and expression of the mitochondrial genome. The length of this region ranges from 880 to 1400 bp. The length among species varies based on the presence of the repetitive sequences. This region is sequenced for many species because of its availability its short size, and improved knowledge in replication and transcription mechanisms, it's preferred for phylogenetic studies. The intro order comparison of the control region is specifically useful in the case of closely related species (Sbisà et al. 1998).

DNA Repositories

DNA repositories are valuable tools for species identification. Such repositories hold millions of DNA sequence database, which is entered by the researchers around the world. Although it is serving as an important tool, it holds some disadvantages. The databases available freely are nonregulated, which means the data entered into the system has chances of misinformation or errors which can lead to misidentification. For many who performed DNA-based identification, this reference database is a primary source of comparison. To overcome misidentification, ISFG suggested using a voucher specimen for comparison of unknown specimen and comment should be made if it's not possible.

GenBank

It's a comprehensive open-source DNA database started in 1992. it is maintained by the National Center for Biotechnology Information (NCBI) in the United States. GenBank is a part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA Data Bank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. GenBank can be accessed through the NCBI nucleotide database. It further gives access to associated information such as taxonomy, protein structure and sequences, genomes, and biomedical journal literature in PubMed. It has a sequence search and alignment tool called the "Basic Local Alignment Search Tool" (BLAST).

EMBL

The European Molecular Biology Laboratories (EMBL) also started in 1982 and maintained at the European Bioinformatics Institute (EBI) in the UK. It started with a repository of 685 entries of individual authors and genome projects groups. Individual submitters preferred the Web-based submission system Webin, while big genome sequencing centers enter data through the automatic procedure. It has a network browser called Sequence Retrieval System (SRS), which allows access to nucleotide and protein databases plus other databases. For sequence, various tools are available such as Blitz and Fasta.

DDBJ

DNA Data Bank of Japan is the sole Japanese database that collects DNA sequences from researchers and issues the internationally recognized accession number to data submitters. It's mostly collected from Japanese researchers. It's started functioning in 1986, and it's the only DNA databank in Asia.

Identification of the Geographic Origin

Wildlife crime has been a global issue. Any criminal activity related to wildlife happening in a country can be traced to a country thousand miles far. For example, an orangutan baby captured in a remote forest of Sumatra to which they are native

could be transported to a buyer in the United States for private zoo collection (Foead et al. 2005). Wildlife trade which is rough counts for 20 billion USD per year (Alacs et al. 2009). It is leading to a rapid loss of biodiversity around the world. One of the challenges with the wildlife crime, which are precursors of illegal wildlife trade, is that it happens majorly in developing or underdeveloped countries, which are coincidently biodiversity hot spots. Such countries often lack the resources and ability to investigate and respond to such criminal activities.

Identification of the geographic origin of a wildlife article gets important in some instances. Some species that are legal to trade in one country may not be legal in another. In cases where illegally sourced timber is laundered, the geological origin becomes vital to investigate. In other cases where a species trade is completely prohibited such as ivory, the source or origin can lead investigators in narrowing down the county and possibly the region of illegal activity taking place. Such information is vital to uncover international syndicates dealing in illegal wildlife trade and their pattern of working.

Principle of Identification of the Geographic Origin

Determination of geographic origin is made simply through species identification. Species have distinct genetic makeup within their geographic range. Sometimes, the geographic isolation of a group of animals for several generations can lead to the creation of a new species altogether. For example, in Thailand, the trade of ivory is partially legal. The legal status of the ivory trade is only for domestic Asian elephants (*Elephas maximus*). However, ivory derived from African elephants (*Loxodonta africana*) is illegal. In a scenario where the law enforcement authorities have to differentiate one from another to test its legality, it can simply be achieved through testing for species of origin. However, establishing the geographic origin of a sample from a forensic genetics perspective requires narrowing it down to a reproductive population of origin. It is similar to the case of identifying a breed or variety of an animal or plant in food product identification (Woolfe and Primrose 2004). Identifying biological populations poses challenges due to the complexity caused by different levels of genetic variations, from stretched families to subspecies. Populations of species often have capabilities of breeding that result in exchanges of genetic materials. Mixing of genetic materials makes DNA markers less probable to differentiate among groups.

Geographic origin is based on recurrently linking a sample to a particular population while successfully differentiating it from other populations of the same species. This requires a population of different geographic origin to be genetically distinct. Based on such genetic variations, genetic databases are created by identifying distinct genetic markers. Determination of geographic origin relies on the population markers associated with a particular population. The task of geographic identification is further complicated with the difference in genetic population signals and legal boundaries pertinent to law enforcement. The ability at which the spread of a population over an area and its spatial resolution can be identified further

complicates the process. As compared to species identification where universally accepted tests and markers are available, geographic origin gets complicated with a focus on a single species location and defined investigative questions.

Methods of Determining Geographic Origin with DNA Markers

Lineage Markers

Traditionally mitochondrial DNA markers such as mitochondrial cytochrome oxidase I and cytochrome b genes are preferred for animals, while for plants rbcL and matK are preferred. These markers are ideal for species identification as discussed in the species identification part of this chapter. These are conserved regions of mitochondrial DNA. For identification of animals within species, it requires a hypervariable D-loop region or control region of mitochondrial DNA (Ishida et al. 2013). Similar to humans, D-loop hypervariable haplotypes in wildlife provide lineage information. Lineage information corresponds to specific geographic areas and ultimately enables to identify the geographic origin. There are several examples of the geographic origin of animals, such as African elephant ivory (Ishida et al. 2013) and cannabis (Gilmore et al. 2007). However, markers from mitochondrial and chloroplast DNA often fail to differentiate and identify samples among geographically distributed locations, which demand the use of nuclear DNA markers.

Microsatellites

The application of nuclear DNA tests for identification of geographic origin depends on the variation in the allelic marker frequency among populations. Microsatellites or STR markers show a high level of polymorphism. As the application of STR marker-based identification technology is being used and simultaneously developing in human identification, it is being used in the geographic identification of wildlife as well. It is used in several identifications, such as ivory (Wasser et al. 2008), tortoise (Schwartz and Karl 2008), bear (Andreassen et al. 2012), plants (Nazareno and dos Reis 2014), and fish (Glover et al. 2009). In a distinct population, a low number of microsatellite markers would give a sufficient degree of power for the identification of geographic origin. However, in a weakly structured population, it may take comparatively a greater number of markers to give confidence. In such a situation where due to the rise in the number of STR markers, it starts becoming excessive, an alternate maker technique may be employed, for example, the rhino indexing system. Rhino poaching for its ivory has been a threat to rhino conservation in India. For tackling such crimes, rhino index DNA system (RhODIS – India) has been developed by researchers of the Wildlife Institute of India in collaboration with WWF India. A DNA database is generated to build a DNA catalog of 749 individual Indian greater one-horned rhinoceros (*Rhinoceros unicornis*). The database is generated by analyzing microsatellite markers present in the nuclear DNA. Such 14 markers were employed to build a database that will create a baseline for linking a confiscated ivory item to the location of the crime (British Ecological Society 2020).

Single Nucleotide Polymorphisms (SNPs)

As a result of genome-wide analysis of biological samples in wildlife, several SNP markers are being identified and characterized in various wildlife species. This has given the chance to employ SNPs in wildlife forensic identification (Ogden 2011). Like STR markers, SNP markers not under selection display allelic frequency differences between different populations due to random genetic drift. Generating the same assignment power as of STR markers, a higher number of SNP is required because of their lower allelic variability. However, SNP markers hold an advantage over STR markers due to their accuracy, simple methods, and ease in interlaboratory data transfer. A big number of SNP markers can be analyzed simultaneously in a single sample. This technology is being utilized in human ancestorial investigation, which is leading its way to wildlife investigations (Helyar et al. 2007). SNP markers give the advantage of a selection of desired markers that are linked to genetic regions under selection for regional geographic conditions. These non-neutral markers provide better interpopulation divergence and relatively lower the number of markers needed for analysis.

Individual Identification

DNA profiling has played a significant role in the routine forensic analysis linked with human identification. Today, it has been a gold standard test for human identification. On the contrary, wildlife forensics does not offer importance to individual identification of animals as it is in human forensics. It's evident because individual identification of animals holds less importance in wildlife conservation barring some situations. In a certain situation, individual identification from animal processed products such as processed meat, animal parts of importance to poachers such as tusks, horns, fins could be performed. Individual identification could be performed for several reasons including curbing entry of illegally sourced meat into markets, identification of poached animals, and maintaining a DNA register for highly protected animals.

Individualization Testing

Individual identification of DNA is based on the selection of variable DNA markers intraspecies and show differences among individuals. The ability of a DNA profile to differentiate or match an individual is based on statistical significance. When two samples of DNA are being compared together, a set of DNA markers are profiled, and their results are compared to declare their relationship. STR or SNP markers are used to create a profile having several alleles. The more the number of markers, the greater the degree of accuracy. If two samples are showing dissimilarity in the profiles generated, the likelihood of two samples originating from a single source is excluded. Two samples showing similar profiles suggest they may originate from a single individual.

Short Tandem Repeats

In human genetic identification, the type of STR markers used is either 4 (tetra)- or 5 (penta)-base-pair repeats; however, in wildlife genetic testing for individualization, 2 and 6 base pairs are suitable. STR loci with greater repeats are less common than loci with fewer repeats. Because of their commonness, STR loci with fewer repeats are more preferred. Many studies on wildlife genetic population reported many dinucleotide repeats for identification, but these were never intended for use in forensic analysis. In the past, each STR locus was analyzed through an individual reaction. The reported primer sets were not designed to be conducted in a multiplex. Many such previously reported STR markers are being multiplexed and employed in forensic analysis in spite of these problems. Commercial kits are available for nonhuman dinucleotide testing. These are for domesticated animals, such as dogs and cats. These kits are for commercial use and not suitable for forensic analysis.

As per the recommendation of the International Society for Forensic Genetics (ISFG) commission on nonhuman genetic testing, tetra-repeats are to be used over dinucleotide repeats (Linacre et al. 2011). This recommendation is due to a problem with dinucleotide repeats. The problem is that the smaller repeats are unstable during DNA replication, which includes PCR as well. This results in the shortening of PCR products due to the loss of a repeat unit. This effect is called as "stutter" peak because now the PCR product is one repeat shorter. Such stutter peaks can be 30% or more of the size of the parent peak. If compared with tetra-repeats, it's reduced to 15% (Butler 2010). Interpretation of results becomes difficult due to an increase in stutter products. In spite of the availability of many tests, the STR test is developed for very few species. It is necessary to develop a test for the species which are being tested commonly for forensic analysis. The application of STR in forensic analysis has got greater importance. This is due to some advantages linked with STR. STR loci are inherited from both parents if they are present on the autosomal chromosome. Due to a low mutation rate, two alleged offspring can be tested to determine their hereditary lineage. STR loci are sufficient for providing discrimination to permit variation within a population. Careful selection of STR markers is necessary to gain enough discriminatory power. An added advantage of using STR markers is that there can be multiple STR assessments at the same time. When a single STR is analyzed, it's called "singleplex," while analysis of two STR markers together is a duplex. In the human genetic analysis, 21 STR markers are being analyzed together, and it's called "multiplex."

Unique genetic profiles can help to monitor the illegal trade of animal products in a commercial market. These products can be profiled even if they are processed. For example, the minke whale population is decreasing because of several reasons; one of the major reasons is climate change, followed by human-induced threats including poaching. A market survey was conducted for the sale of North Pacific minke whales (*Balaenoptera acutorostrata* spp.) in a total of 12 markets of South Korea from 1999 to 2003. The products collected from these markets were profiled using a 464 bp fragment of the mitochondrial DNA control region and eight STR markers. Profiles were compared together to identify a number of unique profiles. The number of unique profiles would suggest the number of individual animals killed to make

market-ready products. A total of 827 individual animals were recorded to be sold in the market over the duration of 5 years estimated from the information generated from DNA profiling. Compared to the official number of 458 bycatch, it's almost doubled, suggesting extensive killing of whales (Baker et al. 2007).

To distinguish a legally sourced meat product from an illegally sourced, a DNA register can be maintained. A DNA profile of legally sourced meat generated in a certified laboratory can be entered into a database. In Norway, a database of minke whales having DNA profiles of 2676 has been established. The genetic profiles are created using ten STR markers, two sex-determining markers, and a mitochondrial D-loop region. This Norwegian DNA register has been successful in distinguishing between legally and illegally sourced whale meat entering the commercial market (Palsboll et al. 2006).

Nonhuman Paternity Testing

Paternity testing in humans is a common practice. It's usually performed to determine the biological parent of an offspring. In wildlife, genetic analysis paternity testing also holds importance in cases where the investigators have a question on captive breeding of an animal. If an animal is not captive bred, then there is a possibility that an animal may have been sourced from the wild, which in some national legislation is prohibited.

The process of paternity testing in animals is similar to humans; it is based on STR analysis. The basis of this test is the likelihood ratio, where there are two opposing hypotheses: one hypothesis suggesting the tested sample is the offspring of the alleged parents and another hypothesis suggesting the sample is not an offspring of the alleged parents. Assuming the tested sample species is diploid and reproduces sexually, then an offspring will inherit one allele from each parent. For example, if the male genotype is M, N and the female genotype is O, P, then any offspring must be one of the four possible genotypes: M, O; M, P; N, O; or N, P. If only one parent is available for testing with an offspring and, the question is whether the animal being tested is a sibling of the already known offspring. And the question is whether the animal being tested is a sibling of the already known offspring. In between two offspring, the probability of a common allele is 0.5, while the probability of having both the alleles common is 0.25. There is also a probability of having no common allele despite being sibling, which is 0.25.

Paternity testing is also having application in wildlife conservation. Captive breeding of endangered animals is a practice followed to increase the population of a species. It also serves as a valuable genetic resource. These captive-bred animals are then reintroduced in a depleted habitat. Accurate paternity testing becomes important for the management of the captive population and selection of suitable individuals of known genetic characteristics for release in the wild. A study was conducted on the captive population of giant panda. A technique that uses two highly polymorphic microsatellites and five less polymorphic markers is found to be highly effective and faster (Li et al. 2011).

Conclusion

While emerging technologies are giving tools to law enforcement agencies to fight wildlife crime, existing DNA technologies are capable of tackling most forensic questions as discussed in this chapter. The selection of genetic markers will depend upon the requirement of the case, the species being tested, its ecology, and biology. Every present DNA technology and markers have their merits and demerits, which one has to evaluate carefully before choosing it for a case. Choosing complementary markers with appropriate resolution could be the best practice to avoid demerits associated with any marker.

An interdisciplinary approach is essential for the advancement of wildlife forensics. There are existing areas, such as conservation genetics and wildlife ecology, which have been dealing with knowledge of wildlife genetics, ecology, and biology traditionally. The existing knowledge of these fields could be a crucial factor in the advancements in wildlife forensics. Successful implementation of wildlife forensics is important for controlling wildlife crime, which eventually contributes to the conservation of endangered species. A collaboration between forensic scientists and conservation geneticists will benefit the overall protection of vulnerable wildlife to poaching and trade.

In wildlife investigation, species identification remains the top priority. Wildlife forensic labs that perform wildlife DNA techniques are often loaded with samples requesting for species identification. Any article or evidence seized by enforcement agencies has to be first tagged with a species it is sourced from. In some cases where seized articles' morphological characteristics are intact species, confirmation is relatively simple, for example, in the cases of tiger skin or deer antlers, where its morphology is important because its market value is depending on its looks. Some articles are deliberately processed to modify or change their function for reasons including its concealment, easy packaging, and transportation. Evidence that is a mixture of different animals is more challenging to identify. A sample containing multiple sources of DNA has to be analyzed differently. A multiplexing approach applies to such samples, as discussed in the chapter. However, the quality and quantity of the sample are crucial. Present technologies sometimes fail to successfully identify because of their limitation.

The most damaging threat is the illegal international trading of wildlife articles. It takes precise information about the location of the crime or sourced articles to diminish the trade. Wildlife articles from rare species fetch more price; such species are often located in the isolated locations. Identification of such locations to stop further exploitation of wildlife becomes important. DNA techniques for geographic identification become an essential tool for such tasks. The science of association of wildlife articles based on DNA analysis to a specific location is much more complex than other areas of wildlife DNA analysis. The challenge is to identify a group within the species, which are isolated due to geographic barriers. It requires development of a test for a specific locality for a particular species. It is combined with population

genetics and life history of the species being tested. Therefore, this type of testing requires detailed knowledge of the locality and a sizeable reference database for developing comparative standards.

Forensic genetics techniques are now being employed to address various wildlife forensic issues discussed in this chapter. Wildlife DNA technology will undoubtedly see steady progress. However, it requires persistent efforts of research and development like human forensic genetics. It is also dependent on research in conservation genetics and related areas. The reliability of any forensic application depends on valid laboratory techniques, evidential security, data analysis, and result interpretation. Such issues must be looked after for the development of successful forensic tools. The development of a reliable reference database is expanding the ability to develop and apply forensic DNA techniques in wildlife.

References

Alacs E, Georges A, FitzSimmons N, Robertson J (2009) DNA detective: a review of molecular approaches to wildlife forensics. Forensic Sci Med Pathol 3:180–194

Andreassen R, Schregel J, Kopatz A, Tobiassen C, Knappskog PM, Hagen SB et al (2012) A forensic DNA profiling system for Northern European brown bears (*Ursus arctos*). Forensic Sci Int Genet 6:798–809

Baker CS, Cooke JG, Lavery S, Dalebout ML, Yu M, Funashi N et al (2007) Estimating the number of whales entering trade using DNA profiling and capture-recapture analysis of market products. Mol Ecol 16:2617–2626

Bär W, Brinkmann B, Budowle B, Carracedo A et al (2000) Guidelines for mitochondrial DNA typing. Vox Sang 79:121–125. Birds through DNA barcodes. Plos Biol 2:1657–1663

British Ecological Society (2020) Researchers turn DNA detectives to aid rhino poaching prosecutions with forensic evidence. ScienceDaily. Retrieved December 21, 2020 from http://www.sciencedaily.com/releases/2020/12/201215164923.htm

Butler JM (2010) Fundamentals of forensic DNA typing. Elsevier, Sydney

Foead N, Elliott W, Ziegler S, Van Gool M, Nijman V (2005) WWF's species action plan for the conservation of orangutans in the wild. WWF's Species Action Plan, UK

Frantz L, Schraiber J, Madsen O et al (2015) Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat Genet 47:1141–1148

Gilmore S, Peakall R, Robertson J, Organelle (2007) DNA haplotypes reflect crop-use characteristics and geographic origins of *Cannabis sativa*. Forensic Sci Int 172:179–190

Glover KA, Hansen MM, Skaala O (2009) Identifying the source of farmed escaped Atlantic salmon (*Salmo salar*): Bayesian clustering analysis increases accuracy of assignment. Aquaculture 290:37–46

Hebert PDN, Cywinska A, Ball SL, Dewaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond Ser B Biol Sci 270:313–321

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2(10):e312

Helyar SJ, Hemmer J, Bekkevold D, Taylor MI, Ogden R, Limborg MT et al (2007) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Mol Ecol Resour 11:123–136

Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome-b gene of mammals. J Mol Evol 32:128–144

Ishida Y, Georgiadis NJ, Hondo T, Roca AL (2013) Triangulating the provenance of African elephants using mitochondrial DNA. Evol Appl 6:253–265

Li DS, Cui HM, Wang CD, Ling SS, Huang Z, Zhang HM (2011) A fast and effective method to perform paternity testing for Wolong giant pandas. Chin Sci Bull 56(24):2559–2564

Linacre A, Gusmão L, Hecht W et al (2011) ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. Forensic Sci Int Genet 5:501–505

Nazareno AG, dos Reis MS (2014) Where did they come from: genetic diversity and forensic investigation of the threatened palm species Butia eriospatha. Conserv Genet 15:441–452

Ogden R (2011) Unlocking the potential of genomic technologies for wildlife forensics. Mol Ecol Resour 11:109–116

Ogden R, Jen M (2015) A review of wildlife forensic science and laboratory capacity to support the implementation and enforcement of CITES. United Nations Office on Drugs and Crime (UNODC) CoP17 Doc. 25 Annex 4

Palsboll PJ, Berube M, Skaug HJ, Raymakers C (2006) DNA registers of legally obtained wildlife and derived products as means to identify illegal takes. Conserv Biol 20:1284–1293

Sbisà E, Tanzariello F, Reyes A, Pesole G (1998) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205:125–140. https://doi.org/10.1016/S0378-1119(97)00404-6

Schwartz TS, Karl SA (2008) Population genetic assignment of confiscated gopher tortoises. J Wildl Manag 72:254–259

Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol 74(2):329–356

Wasser SK, Clark WJ, Drori O, Kisamo ES, Mailand C, Mutayoba B et al (2008) Combating the illegal trade in African elephant ivory with DNA forensics. Conserv Biol 22:1065–1071

Woolfe M, Primrose S (2004) Food forensics: using DNA technology to combat misdescription and fraud. Trends Biotechnol 22:222–226

Barcoding of Plant DNA and Its Forensic Relevance

38

Gianmarco Ferri, Beatrice Corradini, Francesca Ferrari, and Enrico Silingardi

Contents

Abstract

Forensic botany refers to the study of plants and how they can relate to law and legal matters. Unfortunately, while widely known as a science, this discipline has few professionally trained botanists.

Law enforcement workers and forensic scientists are no more informed about the science of botany with the consequence that few individuals understand the importance of plants in criminal investigations. Therefore, important plant evidence is frequently overlooked. Plants or their parts can place a person or object at a crime scene, verify or refute an alibi, and help determine the cause and the time since death, the time of a crime, the place where a crime occurred, or the reason for an illness.

Plant DNA barcoding identifies species using recent advances in genetics, genomics, and bioinformatics using short DNA sequences, known as the DNA barcode. This concept was established almost 20 years ago, and it's based on the comparison of sequences obtained from unknown species against a reference database with the purpose of identification.

Gianmarco Ferri and Beatrice Corradini contributed equally with all other contributors.

G. Ferri (✉) · B. Corradini · F. Ferrari · E. Silingardi
Department of Biomedical, Metabolic and Neural Sciences, Institute of Legal Medicine, University of Modena and Reggio Emilia, Modena, Italy
e-mail: gianmarco.ferri@unimore.it; beatrice.corradini@unimore.it; francesca.ferrari@unimore.it; enrico.silingardi@unimore.it

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_42

This universal and highly standardized method has also proven its benefits in forensic investigations both for wildlife and plants, as several studies and real caseworks have proven over the years. However, especially for plants, some challenges in the implementation of this system still exist, mainly related to the current status of the global reference sequence databases.

Efforts must be encouraged to enrich data engendered by barcoding projects so that forensic botany can benefit from it. This is particularly critical in countries with extraordinary biodiversity.

Keywords

Forensic botany · DNA barcoding · Non-human DNA

Introduction

The use of nonhuman DNA analysis in forensic science has seen a very rapid growth in recent years.

Applications range from investigations of rape and murder of humans to cruelty and poaching toward domestic or wildlife animal species and analysis of plant evidence, bacteria, and viruses to aid the resolution of legal cases. Nonhuman biological specimens are often part of the physical evidence from crime scenes or involved in most forensic caseworks, and results have been used as evidence in court in a variety of cases (Miller Coyle et al. 2001).

Forensic botany, otherwise known as plant forensics, is the study of plants and how they can relate to law and legal matters, and it's a relatively recent young application of an old science. The kidnapping and death of Charles Lindbergh's young son in 1932 was the first modern era case to use such botanical evidence in court (Bock and Norris 1997).

This discipline regards the analysis of plant and their parts, such as leaves, flowers, pollen, seeds, wood, fruit, and spores, plus plant environments and ecology. Unfortunately, it remains a little considered discipline in the forensic context today.

Despite its poor implementation, the correct identification of plant species can be very relevant in forensic investigations, and the importance of the botanical evidence is amply demonstrated (Brown 2006; Stambuk et al. 2007; Margiotta et al. 2015). Traces of plants can be associated with crimes and could represent a valid aid in the process of environmental reconstruction as plants could act as indicators of a geographical place or an ecological environment. Also, plants can help experts in the temporal reconstruction giving information on the postmortem interval, the exposure time of an evidence, and the date of an hidden burial.

Moreover, as highlighted in a recent works (Aquila et al. 2014, 2019), botanical traces can help the forensic pathologist to answer fundamental questions about a case, including the dynamics of the event, and the cause of death (suicide, homicide, or accident) or when autopsy findings are not sufficient to ascertain the exact circumstances in which death occurred (Aquila et al. 2014).

Plants can be used in food traceability and quality control, illegal logging and trade, and investigations of poisoning with products derived from plants, among others (Coyle 2005).

Another important field is the study of diatoms in drowning cases. The peculiar silica wall that resists to the acidic environment of the human body and survive for many years makes these golden brown algae as an important tool in forensic investigations especially when it's necessary to determine if it's a case of antemortem or postmortem drowning (Liu et al. 2020a; Zhou et al. 2020).

Moreover another area of interest is the particular threat to some taxa that comes from the overexploitation for commercial trade in plants and their products, which has dramatically increased in recent decades. International conventions like CITES (Convention on International Trade in Endangered Species) are designed to combat illegal trades for endangered and threatened species, but the effectiveness of their governing rules and measures is highly dependent upon the rapid and accurate identification of the species of interest.

However, the reasons why botanical evidence is still underused and mostly ignored in the forensic context are different and linked to a variety of reasons.

A relevant reason concerns the first part of an investigation, that is, the recognition of pertinent plant evidences at crime scene and the subsequent collection, transportation, and preservation. This is a complex matter, requiring trained personnel, but unfortunately, very few professionals involved with law enforcement have an adequate background or training in botany.

Even if properly collected and stored, a botanical evidence can be fragmentary or could be a portion of the organism, preventing the accurate species identification based upon traditional morphological and anatomical characters, and often requires the intervention of experienced taxonomists or specialized forensic botanists, which are very rare.

Furthermore, the forensic reality is very different from that in which taxonomists are used to working as a complete intact and fresh specimen is collected very rarely by investigators.

Typically, legal investigators should seek a botanist with well-rounded experience who possesses knowledge of the various specialties within the botanical field in question. Some of the various botanical specialties involve systematics (plant names), anatomy (plant cells), morphology (plant structures), ecology (relationship of organisms within the environment), and physiology, chemistry, and genetics.

However, in common forensic scenarios where the specimen is incomplete, traditional morphological methods appear to be ineffective in the taxonomic identification the organism of origin.

Forensic botany tried to implement molecular biology techniques developed firstly for phylogenetics and botanical taxonomic studies which are fast, more accessible, and affordable.

The first criminal case that used plant DNA typing to gain legal acceptance was a homicide that occurred in 1992 in Arizona's Maricopa County using a PCR-based fingerprinting technique of randomly amplified polymorphic DNA (RAPD) (Mestel 1993).

In this case, a woman's strangled body was found near some Palo Verde (*Cercidium floridum*) trees. During the investigation, in the back of the suspect's truck, detectives found a few seed pods from a Palo Verde tree, which were sent for DNA analysis. RAPD analysis was carried out, and it was found that the banding pattern seen in these seed pods was identical to that obtained from the tree under which the woman's body was found and had shown signs of abrasion thought to have resulted from the suspect's truck. Moreover, each tree in the area was found to show unique banding patterns. This was the first instance for the admission of plant DNA evidence in court, and the jury convicted the suspect of first-degree murder. However, since RAPDs are not considered sufficiently reliable, their use in forensic science has since diminished.

In another homicide case, three species of bryophytes (mosses) were found on the suspect and identified to species. DNA fingerprinting analyses of traces bryophytes demonstrated that they likely originated from the crime scene (Korpelainen and Virtanen 2003).

Bryophyte material found on the suspect's car tires, shoes, and clothes was successfully linked with bryophyte patches located at crime scene. This study further demonstrated the great importance of botanical evidence by virtue of the ubiquitous presence of plants and their parts in the surrounding environment. Plant fragments, e.g., bryophytes, can easily become attached to shoes and clothes, and they can be analyzed quite long after the plant has been fragmented.

Another homicide was investigated using leaves of sand live oak, *Quercus geminata*, collected from three trees at the site where the victim was found and those found in the trunk of a suspect's car (Craft et al. 2007).

Four STR loci were analyzed, and the results showed that alleles from the plant material on the suspect's car did not match those obtained from the three trees near the burial site across four loci. Although in this specific case it did not provide the required physical evidence to link the suspect's car to the crime scene, the potential offered by the use of plant DNA evidence in forensic science was demonstrated.

AFLP is another technique used to create a DNA profile for plant varieties and has been applied to marijuana samples by some authors to link growers and distributors of clonal material and the subsequent individualization of a sample (Miller Coyle et al. 2003).

Those whole-genome DNA fingerprinting methods were used until a few years ago because they were the only ones available by the scientific community. They are quite adequate as a means of rapid screening, but the problem of reproducibility especially from degraded specimens and the need for high amounts of quality DNA and the difficulty of interpreting electrophoretic profiles in mixture prevent the standardization of results across laboratories, and subsequently their use in routine forensic investigations was abandoned after a short time. The fact of not having a standardized and reproducible molecular methods for many years, especially having to consider compromised forensic samples, has contributed to the poor consideration of forensic botany over time and to the fact that it remained an elitist subject accessible only to a restricted circle of few experts in the world.

Forensic Botany and DNA Barcoding

Almost 20 years ago, Hebert and colleagues of the University of Guelph (Ontario, Canada) launched the ambitious idea of creating a universal molecular system for the identification of every living species, reliant on the analysis of sequence diversity in small segments of DNA, called DNA barcoding (Hebert et al. 2003a).

The ability of DNA barcoding to distinguish species from a range of taxa and also to reveal cryptic species has really revolutionized the taxonomic world, facilitating species identification by using short, standardized portions of the genome.

As stated by the authors, in order for a region of DNA to be effective as a barcode, it must simultaneously contain enough variability to be informative for unique species identification; work easily on all group of animals, land plants, and fungi; be short enough to sequence in a single reaction; and contain conservative regions which can be used to develop universal primers.

This ambitious project has been supported from the beginning by an international consortium (Consortium for the Barcode of Life, CBOL) of major natural history museums, herbaria, and other organizations (https://ibol.org/).

Arguably, the greatest beneficiaries from this system will be the many professionals whose work involves solving real-world problems with broad impacts on all areas in which society interacts with biodiversity but whose job is not to carry out taxonomy, among which there are forensic scientists.

A DNA sequence of nearly 648 bp of the mitochondrial gene cytochrome c oxidase 1 (COI) is commonly accepted as the universal barcode for the animal kingdom because it meets all the features of universality, reproducibility, high level of standardization, and high discrimination to the species level (Hebert et al. 2003b).

The barcode approach demonstrated to have great potential for the identification of animal species, providing rapid and accurate recognition of unknown samples including partial fragments and compromised or old tissues, whose DNA barcodes have already been registered in a DNA sequence library. In agreement with forensic quality procedure and standardization requirements, the barcode identification system for animals has been widely applied in many forensic cases involving animals as the abundant scientific literature on the subject demonstrates (Dawnay et al. 2007; Wilson-Wilde et al. 2010; Johnson et al. 2014).

In casework situations, DNA degradation and recurrent sample inhibition may complicate the recovery of a full-length 648 bp barcode. Although longer sequences give greater resolution, also shorter sequences unexpectedly provide excellent resolution at the species level (Dawnay et al. 2007; Hajibabaei et al. 2007).

Therefore, for highly compromised material, shorter COI fragments can be sequenced with an alternative set of primer pairs developed to enhance the robustness of the test (Ferri et al. 2009).

As is well known, the species discrimination capacity strictly depends on the quantity and the quality of data available in a reference DNA library. Then, one of the primary goals of DNA barcoding focuses on the assembly of a reference library of barcode sequences for known species recovered from multiple voucher specimens using standardized protocols.

CBOL developed the Barcode of Life Data Systems (BOLD), an online freely available workbench that aids the collection, management, analysis, and publications of DNA barcode records together with morphological and distributional data.

Furthermore, the BOLD platform provides an identification system, namely, BOLD Identification System (IDS), a dedicated species diagnosis tool, which accepts sequence queries for animal, plant, or fungal species identification to be compared with specimens already registered in the repository.

While DNA barcoding enjoyed a remarkable success for animal identification, attempts to identify the barcode loci for plants have been more tortuous.

DNA barcoding in plants presents challenges that are not encountered in animals. In plants, as a consequence of their evolutionary history, species boundaries are less defined. Added to this, there are logistic difficulties of undertaking identification of something like more than 400,000 land plant species known to date.

Moreover, the mtDNA of plants is not suitable for species identification procedures since it is usually slowly evolving, resulting in the absence of interspecific variation, and has high intramolecular recombination and pseudogenes. Therefore, the standard barcode markers for animals could not be applied to the plant kingdom. The search for the corresponding DNA barcode focused on the plant chloroplast (cp) genome, which is an alternative to the animal mitochondrial genome. The cp genome could contain suitable barcoding markers because of its presence in each plant cell in a high number of copies and consists of conserved gene sequences. The downside of the chloroplast genome is its relatively low rate of evolution. Focus has been placed upon identifying those regions that evolve quite rapidly but still slowly enough to be present in all land plants and that are good candidates for robust, universal primers.

After years of research, in 2009 the Plant Working Group of CBOL promoted a multilocus solution comprising portions of the two plastid coding genes rbcL and matK as the core system for land plant identification, remarking that the discriminatory power is to be expected lower than for the animal kingdom, with a discrimination value of 70–80% to the species level (CBOL Plant Working Group 2009).

Even if the percentage of success doesn't equal that of animals, this system could be useful for specific applications where the resolution to the species level is not always required, as in forensic casework investigation.

Following the DNA Barcoding Consortium final decision on core barcoding markers for plants, our group tested the selected regions on some samples of local flora in order to evaluate the possible application to forensic botany (Ferri et al. 2015).

Our efforts have focused on developing a DNA-based identification system that provides criteria to progressively identify an unknown plant sample to a given taxonomic rank also by any nonspecialist botanist.

Our results, based on the official CBOL land plant barcode (matK and rbcL), as well as on two alternative intergenic plastid loci (trnH-psbA and trnL-F) tested individually and in combinations, showed as expected that a two-loci synergic

approach best addresses the needs for species testing of plant material also for forensics in the context of local flora, and it must also be explored in other contexts to extend and further evaluate its usefulness.

The selection of rbcL+trnH-psbA was driven by the availability of well-tested and robust universal primers for both loci which facilitate the reproducibility of results and the implementation of the method in forensics together with a large volume of sequence data available in public sequence databases.

In the implementation of this multistep model, a less variable coding locus (rbcL) complements the resolution given by a highly variable and therefore potentially more discriminating noncoding marker such as trnH-psbA. The use of trnH-psbA had been well documented in the literature, and also the Consortium proposed it as an optional supplementary marker, together with the advantage of possessing greater sequence variability than the proposed core barcode loci.

The results obtained confirmed the effectiveness of this assay based on barcoding principles in forensic investigations also for the plant kingdom with a higher success in identifying the correct species if the two loci were both analyzed, equal to nearly 70% of the samples of local flora analyzed.

Depending on the matter at hand, this core barcode system could be supplemented with additional loci such as the noncoding region trnL-trnF, for example, in cases involving highly degraded tissue.

Moreover, given the complexity and the great variety of the plant kingdom, it is to be expected that for some specific taxon groups particularly difficult to distinguish, the use of alternative markers may be required.

As an example of this, Liu et al. reported a work in the peculiar forensic context of biosurveillance, conservation management, and policing illegal trades (Liu et al. 2018).

They particularly focused on the genus *Taxus* whose species are part of the protected flora in China but critically endangered due to deforestation and land-use change, as well as illegal exploitation.

This is a notoriously taxonomically difficult genus with ongoing uncertainty and disagreement about its classification as well as with a broad distribution across temperate of the northern hemisphere, covering North America, Europe, North Africa, and Asia.

As they stated, policing this illegal trade requires an accurate species identification system.

However, morphological characters tend to vary greatly within species of this genus and often with overlap among species, leading to ongoing taxonomic controversy which in turn causes uncertainty about the distribution range.

In the mentioned study, three data sets, with a total of 4,151 individuals representing all the 15 currently known *Taxus* species worldwide, were used to determine the ideal DNA barcode and construct a structured species identification system for the genus considered. Five data analysis methods were tested for species discrimination power. Based on the performance of single barcodes and their combinations, they recommend trnL-trnF as the best single DNA barcode for *Taxus* and trnLtrnF + ITS as the best combined barcode to reach identification at species level.

For plants, more than for animals, a correct identification at species level strictly depends on the state of the available databases (Meiklejohn et al. 2019; Ekrema et al. 2007).

For trnH-psbA, the successful use of this marker depends on the number and the quality of sequences only in the GenBank repository, as the BOLD database does not accept queries from this marker but only from the standard accepted barcodes, rbcL and matK (http://www.barcodinglife.org/index.php/IDS_OpenIdEngine). Therefore the scientific community is kindly invited to submit the highest number of reference sequences of proven quality in order to minimize the risk of misidentifications or ambiguous and incorrect results and aid the implementation of this marker in the species identification process in the forensic field.

Therefore the growth of the reference barcode library BOLD and the improvement of the bioinformatic platform to better support a multimarker barcoding in plants are undoubtedly basic requirements to encourage the implementation of the barcoding approach even in the forensic botany routine.

Furthermore, the scientific community should be encouraged to produce more studies on the levels of inter- versus intraspecific variation in different taxon groups for the chosen barcode fragments, in order to enhance the confidence of results, for example, in such cases where there is not a 100% homology with the best closest match in the database search, a situation that is quite common for plants.

After choosing the best markers for forensics, in order to facilitate the implementation and simplify its use by forensic laboratories, we developed an operating protocol based on a specific DNA extraction, amplification with universal primers, Sanger sequencing analysis of selected markers, and a description of the main steps for database search considering the latest suggestions of the Plant DNA Barcoding Project (Ferri et al. 2012).

Few years ago, Meiklejohn and colleagues developed a particular and promising DNA barcode protocol suitable for processing complex forensic-type biological fragments recovered from surface soil samples. Its utility and applicability was broadly tested in the mentioned study with a wide variety of different types of fragments (e.g., seeds, leaves, bark, head, legs) (Meiklejohn et al. 2018).

Actually, the Barcode of Life Data Systems (BOLD) and GenBank are the main public repositories of DNA barcode sequences (Fig. 1).

A broad taxonomic barcode coverage has been achieved for certain groups in recent years, but this success has so far always been limited mainly to animal groups and restricted to specific geographical regions.

A comprehensive barcode library could be defined as one that captures 95% of genetic variation, and this has been estimated to require a minimum of individuals per species (Bergsten et al. 2012); this is also affected by the geographical scale of sampling as well as the population structure of the species sampled. However, it is often difficult or impossible to obtain material from the full distribution range of a species, especially for plants. Therefore, many existing libraries are incomplete, introducing bias and possible misidentifications. These issues could cause serious problems for conservation and especially law enforcement regarding endangered species or species protected by international conventions (e.g., the Convention on the International Trade of Endangered Species of Wild Fauna and Flora, CITES).

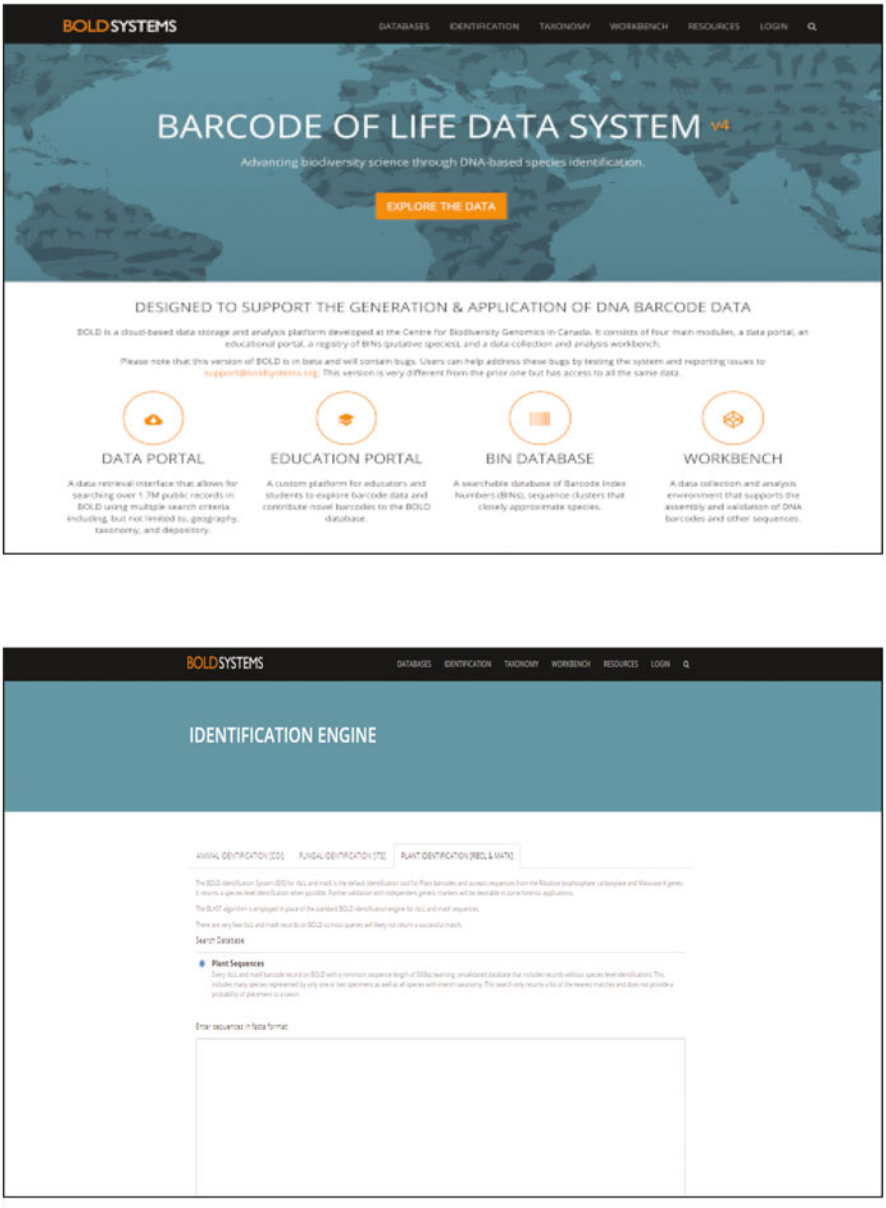


Fig. 1 The BOLD web platform interface and the specific identification engine page for the plant kingdom

Ideally, all barcode sequences contained in either database should have been derived from a vouchered specimen, initially identified by a taxonomic expert.

However, given the inherent nature of any public database, it is inevitable that some erroneous data will be present. The generation and submission of incorrect

sequences likely occurs due to misidentification of the original material, poor isolation techniques, contamination of cultures, endoparasites in insects and plants, and PCR-based errors. It is also known that sequence availability in GenBank varied greatly among taxa and markers (Kolter and Gemeinholzer 2021).

Thanks to the barcoding project, a carefully validated library of reference sequences from voucher specimens of all eukaryotic life is available. BOLD already contains barcode sequences for 319,715 formally described species covering animals, plants, fungi, and protists (with ~11.9 million specimens), and it is still in continuous expansion (http://barcodinglife.org/index.php/TaxBrowser_Home, last access to data 4th January 2021).

The BOLD Identification System (IDS) is the dedicated species diagnosis tool and accepts queries also from the plant kingdom for the selected barcoding genes rbcL and matk.

Supported by an integrated bioinformatics platform, the use of BOLD-IDS is a very important source and could allow for a rapid, easy, and reliable identification even by nonspecialists who paste their sequence record into the web window.

Actually, a forensic expert can rely on the availability of two reference databases to attempt to assign the unknown botanical evidence to the correct species to which it belongs (Liu et al. 2018; Pentinsaari et al. 2020).

While GenBank aims for comprehensive coverage of genomic diversity, the Barcode database would aim for comprehensive taxonomic coverage of just a single gene or at most few genes per kingdom, but it ensures tighter controls over the sequences deposited. In fact, to be able to deposit a sequence data in BOLD, gene sequences must derive from a designated gene region, and they must meet quality standards and derive from a specimen whose taxonomic assignment can be reviewed, ordinarily through linkage to a specimen that is held in a major collection (Table 1).

Integration of search with different databases and implementation of a multilocus barcode approach may improve the chance of reliability of results and come to aid in the decision process when making forensic species identifications.

In a recent study, Meiklejohn et al. demonstrated that both databases perform comparably for plants and macro-fungi (~81% and ~57%, respectively). Moreover, their results illustrated that using a multilocus barcode approach increased identification success (Meiklejohn et al. 2019).

However, beyond the theory, some challenges in the implementation of DNA barcoding in the forensic laboratory still last. As pointed out by Liu et al. referring to a group of organism of particular forensic relevance as the diatoms are (Liu et al. 2020b), they noted that a high percentage of incorrect sequences are deposited in GenBank, compromising the database searching. Reportedly, the original materials, from which the sequences were extracted, were not correctly identified. Thus, even if the barcode sequence was successfully obtained from the diatom evidence, it's very difficult to verify all these sequences without correct reference sequences in the repository.

Table 1 Scenarios that may be encountered in species identification searching in sequence databases (GenBank and BOLD)

Situation	Description	Outcome
The query sequence doesn't already exist in the database	The genomic region for a given barcode is not yet available for comparison	Ambiguous or incorrect (false-positive) ID, as comparison, is only achievable at higher taxonomic ranks (genus, family, order)
The query sequence exists in the database:		
a) Single best match with the exact sequence	Query sequence is returned as the unique best hit with the highest statistical values	Correct ID at *species* level
b) Multiple best matches with the exact sequence	Query sequence plus one or more identical sequences is returned as best hit with highest statistical values	Correct ID only at *genus* or *family* level (based on the taxonomic category giving the best match)
c) Best match with different unrelated sequence(s)	The true species is present in the reference alignment but the method failed to assign as the best hit	Incorrect ID

ID Identification

Another important field of forensic botany that can benefit from the implementation of the barcode project and its high degree of standardization are the complex discipline of forensic palynology.

Forensic palynology refers to the study of pollen and spores to prove or disprove a connection in criminal cases.

Pollens have been utilized by experienced researchers for at least the last three decades to provide forensic evidence and knowledge in certain legal circumstances (Chen and Shi 2020). Pollen grains are utilized in forensic applications because they are exceptionally impervious to chemical attack (Schield et al. 2016). They can remain at the crime scene for long time after that the event under investigation happened. Forensic palynology has been particularly useful in cases where there is a suspected movement of evidence or where a crime has occurred in a location with distinct plant species. Analysts could tie individual criminals' travel histories together based on finding a similar pollen species composition on seized evidence, possibly linking their crimes and providing direction for further investigation. Officials could determine illegal imports' country of origin. One of the most famous cases, following the Bosnian war, regards the uncovering of mass graves where bodies had been moved from different locations. Pollen was one of the lines of evidence used to trace bodies to their original burial sites (Brown 2006). Traditionally, forensic palynology is done by examining pollen grains under a microscope and comparing them to known pollen morphology. This is a highly specialized skill, and

there are few experts able to identify plant species based on the size, shape, and color of the pollen grains. After all, researchers estimate more than 400,000 species of plants living on our planet today. Forensic palynology is further limited by the labor intensiveness of morphological identification. Frequently it's impossible to determine the exact species present; identification is typically to a genus or family of plants, a group of species, in other words. This limits the technique's utility, because while many plant species occur in a small geographic range, the genus or family to which they belong may cover a much broader area.

In a work of Bell et al. (2016a), authors revealed how identifying pollen through DNA barcoding could be a practical alternative although not without challenges. As underlined by authors, DNA barcoding has the potential to greatly increase the efficiency and accuracy of pollen identifications, allowing forensics for the first time to unlock the potential of pollen as a key geographic and temporal marker. However, compared to the classic barcode workflow, in order to better distinguish the mixed-species pollen samples, they underlined the need to implement a "next-generation" or "high-throughput" sequencing (HTS) method that generates multiple reads per sample of the set of DNA barcoding markers, allowing the simultaneous identification of several species from a single mixed-species forensic sample.

High-throughput DNA sequencing (also known as DNA metabarcoding) is the methodological advance that could make pollen DNA barcoding more feasible. This method allows researchers to sequence multiple pieces of DNA at the same time, without separating them first. It's a key innovation because forensic pollen samples typically contain a mixture of species. Without high-throughput sequencing, these species would first need to be painstakingly separated, and the same efficiency problems of traditional morphological analysis come back. With high-throughput sequencing, the whole mixture of pollen grains can be ground up in one sample, and the DNA is isolated and sequenced and matched to a database.

To date, five markers have been employed for pollen DNA metabarcoding, including rbcL, matK, ITS2, trnL, and trnH-psbA with relative success (Bell et al. 2016b). The relative advantages and limits of different barcodes are mostly the same for pollen DNA barcoding as for general plant DNA barcoding. As underlined by Bell et al. (2016a), there are two main benefits of using a DNA barcoding method in forensic palynology field. First, this method is able to identify multiple taxonomy groups, and secondly, it more efficiently identifies parts of the organism that do not appear in the morphology.

Even unlinked, as in a DNA metabarcoding context, multiple markers can generate better discriminatory power than one. Additionally, amplicon fragment size is important in HTS DNA metabarcoding, as the read lengths in many platforms are limited. In this regards, ITS2 and trnL have been successfully sequenced via HTS with sufficient overlap for paired-end reads. The long amplicon length generated by standard rbcL and matK primers, as well as the length hypervariability of trnH-psbA, poses a technical limitation, necessitating a redesign of primer pairs for shorter amplicon lengths. Future improvements in sequencing technology are likely to increase read lengths of HTS technologies and may alleviate this issue in the future. However, even after more than 10 years of the identification of the core barcoding

regions for plants and the impressive work done by the Barcoding Consortium to collect sequences from as many species as possible, there are some challenges in the implementation of a validated species identification method for plants in forensic science.

Difficulties in training qualified personnel primary involved in the collection of crime scene evidences are still an obstacle to overcome in order for forensic botany to become familiar to the forensic experts.

Moreover, when analyzing plant evidence, forensics should be aware of the intrinsic limitations inherent in the identification of plant species with respect to the animal one.

Among these, we remember the larger numbers of plant species globally, the complexity of molecular evolution of plant genomes that makes resolution at the species level more difficult to achieve, the poorer coverage of species in the sequence databases, and the need to use a multilocus system including at least two markers instead of being able to use only one as for vertebrates.

In these years, forensic geneticists are more and more realizing the benefits deriving from the introduction of the DNA barcoding technique in assigning taxonomic names to unknown trace botanical specimens from a casework, even for the untrained user. The features of this easy but reliable DNA-based identification system might be able to bridge the current lack of a highly standardized identification process in forensic botany.

Conclusion

In a short time, forensic science has seen an exponential growth of nonhuman genetics, and the contributions of new research findings to investigations in this emerging field are enormous.

As widely demonstrated, forensic botany is a powerful tool for crime scene investigation. Even a small piece of plant may reveal important information for the court, and botanic materials may connect a suspect with a crime scene or reveal the manner and the time (PMI) of death (Hall and Byrd 2012).

But only if an investigator is aware of the potential existence of that evidence any efforts will be made to search for it. It must be highlighted that a botanical evidence can be present in a crime scene at the microscopic level (such as grains of pollen) and that careful collection, documentation, and preservation of botanical evidence are critical to the evaluation of plant evidence.

Unfortunately, it appears that despite the tremendous potential offered by plant DNA found as trace evidence in solving crimes, there has been little progress to date. Until recently, plant identification has been largely dependent upon morphology-based approaches, which in turn depended upon taxonomical specialists, who are generally the only experts on some specific groups of plant.

Furthermore, traditional taxonomic approaches can rarely be scaled up for high throughput, making it inconvenient for routine forensic applications in species identification.

DNA based-approaches have come to meet the limits related to botany classical and forensic use of identifying morphology. The introduction of DNA barcoding firstly in the taxonomic and systematic field with the aim of creating a practical, cost-effective tool to assign any unidentified specimen to the correct species based on sequence diversity has really revolutionized the field. The current popularity of DNA barcoding relates to its potential power to allow quick identification of organisms for the widest group of practitioners and potential societal benefits to various disciplines of the methodology which must be accessible and easily carried out by multiple users.

In recommending rbcL+trnH-psbA as a forensic two-loci barcode system for plants when species identification of an unknown specimen is necessary, some challenges must be kept in mind by users. Species discrimination with plant barcodes is typically lower than with COI in animals, being anyway limited by a discriminatory power near 70–80% even if adding more markers **[23].**

Some particular plant species belonging to complex taxonomic groups may require the use of additional and specific barcode markers to identify plant specimens or parts of them (Liu et al. 2018).

However, the lack of solid, easy, and reproducible protocols prevents the routine use of these analytical techniques in forensic laboratories, especially for plant species identification even if in forensic casework it is highly relevant to be able to deduce the species origin of an unknown biological sample.

The attractive features rely in the high standardization of the identification process which is rapid, practical, accurate, simple, and easily carried out by multiple users, especially nonspecialists of wildlife/botany.

In addition, it is applicable to a vast range of animal and plant species in a universal manner and must, therefore, prove a valuable tool for establishing species identity for forensic application.

Even if identification to the species level is not always possible, identification to higher taxonomic ranks can be helpful in forensic investigations as in many of the situations in which this method would be applied; the application of a broad species concept is accepted.

While past works have validated the ability of COI sequences to diagnose species in most taxonomic groups of animals, plant species are harder to discriminate, and it is straightforward that a multilocus approach will be necessary.

We emphasize that actually some technical difficulties in the sequence similarity database search, such as the influence of the registered sequence length and quality, as well as the number of ambiguities, affect the results and the success of species-species identification (Pentinsaari et al. 2020).

This is a fundamental point on which we must commit to work in the near future for the quality and reliability of the data of the identification process, as admissibility of any evidence in court clearly requires a certain set of quality standards.

However, when a complete global reference library of DNA based on international guidelines will be available in the public domain, and universal data standards are applied to DNA barcode records, the scientific community will be able to obtain reliable information concerning the barcode sequence for unknown nonhuman

specimens, and this will modify in a relevant manner the approach on non-human forensic genetics.

The expansion of the reference databases is an urgent priority of the scientific community, to include more species that might be of interest to forensics specialists. Moreover, it would beneficial if researchers submit to the literature more cases implementing the analysis of plant specimens to add evidence in solving the cases.

In conclusion, sequence analysis of the DNA markers (COI gene for animals and the multilocus regions for plants) with a barcoding strategy proved to be a very sensitive and powerful technique for specimen detection and, together with its unambiguous and direct identification capacity, becomes a useful tool in routine forensic issues.

It shows powerful universality and versatility at the species level and can sometimes provide insights beyond those obtained through morphological analysis alone. In the presence of a well-established reference library, an unknown sample can theoretically be identified to species using its DNA barcode sequences.

Some more complex disciplines of forensic botany such as forensic palynology will greatly benefit from further development of advanced molecular technologies of next-generation methods of sequencing as they are the only ones that can guarantee a better resolution in cases of mixed environmental samples in which the single species component is difficult to separate. However, to realize the full potential of DNA barcoding for plants, and particularly its application to metabarcoding for samples consisting of mixtures of different species, systematic sequencing of reference collections in the future is required using an augmented set of DNA barcode loci, applied according to agreed data generation and analysis standards (Alotaibi et al. 2020).

References

Alotaibi SS, Sayed SM, Alosaimi M, Alharthi R, Banjar A, Abdulqader N, Alhamed R (2020) Pollen molecular biology: applications in the forensic palynology and future prospects: a review. Saudi J Biol Sci 27(5):1185–1190

Aquila I, Ausania F, Di Nunzio C, Serra A, Boca S, Capelli A, Magni P, Ricci P (2014) The role of forensic botany in crime scene investigation: case report and review of literature. J Forensic Sci 59(3):820–824

Aquila I, Sacco MA, Ricci P, Gratteri S (2019) The role of forensic botany in reconstructing the dynamics of trauma from high falls. J Forensic Sci 64(3):920–924

Bell KL, Burgess KS, Okamoto KC, Aranda R, Brosi BJ (2016a) Review and future prospects for DNA barcoding methods in forensic palynology. Forensic Sci Int Genet 21:110–116

Bell K, de Vere N, Keller A, Richardson R, Gous A, Burgess K, Brosi BJ (2016b) Pollen DNA barcoding: current applications and future prospects. Genome 59(9):629–640

Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, Hendrich L, Geijer J, Herrmann J, Foster GN et al (2012) The effect of geographical scale of sampling on DNA barcoding. Syst Biol 61(5):851–869

Bock J, Norris D (1997) Forensic botany: an under-utilized resource. J Forensic Sci 42(3):364–367

Brown AG (2006) The use of forensic botany and geology in war crimes investigations in NE Bosnia. Forensic Sci Int 163(3):204–210

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci U S A 106: 12794–12797
Chen YH, Shi HF (2020) Research Progress on forensic palynology and its application in forensic science. J Forensic Med 36(3):354–359
Coyle HM (2005) Forensic botany principles and applications to criminal casework. Boca Raton: CRC Press
Craft KJ, Owens JD, Ashley MV (2007) Application of plant DNA markers in forensic botany: genetic comparison of Quercus evidence leaves to crime scene trees using microsatellites. Forensic Sci Int 165:64–70
Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int 173:1–6
Ekrema T, Willassen E, Stura E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol Phylogenet Evol 43(2):530–542
Ferri G, Alù M, Corradini B, Licata M, Beduschi G (2009) Species identification through DNA "barcodes". Genet Test Mol Biomarkers 13(3):421–426
Ferri G, Corradini B, Alù M (2012) Capillary electrophoresis of multigene barcoding chloroplast markers for species identification of botanical trace evidence. Methods Mol Biol 830:253–263
Ferri G, Corradini B, Ferrari F, Santunione AL, Palazzoli F, Alu M (2015) Forensic botany II, DNA barcode for land plants: which markers after the international agreement? Forensic Sci Int Genet 15:131–136
Hajibabaei M, Singer GAG, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 23(4):167–172
Hall DW, Byrd J (2012) Forensic botany: a practical guide, 1st edn. UK: Wiley-Blackwell
Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003a) Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–321
Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc Biol Sci 270(Suppl 1):S96–S99
Johnson RN, Wilson-Wilde L, Linacre A (2014) Current and future directions of DNA in wildlife forensic science. Forensic Sci Int Genet 10:1–11
Kolter A, Gemeinholzer B (2021) Plant DNA barcoding necessitates marker-specific efforts to establish more comprehensive reference databases. Genome 64(3):265–298
Korpelainen H, Virtanen V (2003) DNA fingerprinting of mosses. J Forensic Sci 48(4):804–807
Liu J, Milne RI, Moller M, Zhu GF, Ye LJ, Luo LH et al (2018) Integrating a comprehensive DNA barcode reference library with the global map of yews (Taxus L.) for species identification. Mol Ecol Resour 18(5):1115–1131
Liu M, Zhao Y, Sun Y, Li Y, Wu P, Zhou S, Ren L (2020a) Comparative study on diatom morphology and molecular identification in drowning cases. Forensic Sci Int 317:110552
Liu M, Zhao Y, Sun Y, Wu P, Zhou S, Ren L (2020b) Diatom DNA barcodes for forensic discrimination of drowning incidents. FEMS Microbiol Lett 367(17):fnaa145
Margiotta G, Bacaro G, Carnevali E, Severini S, Bacci M, Gabbrielli M (2015) Forensic botany as a useful tool in the crime scene: report of a case. J Forensic Legal Med 34:24–28
Meiklejohn KA, Damaso N, Robertson JM (2019) Assessment of BOLD and GenBank – their accuracy and reliability for the identification of biological materials. PLoS One 14(6):e0217084
Meiklejohn KA, Jackson ML, Stern LA, Robertson JM (2018) A protocol for obtaining DNA barcodes from plant and insect fragments isolated from forensic-type soils. Int J Legal Med 132: 1–12
Mestel R (1993) Murder trial features trees genetic fingerprint. New Sci 138(1875):6
Miller Coyle H, Ladd C, Palmbach T, Lee HC (2001) The green revolution: botanical contributions to forensics and drug enforcement. Croat Med J 42(3):340–345
Miller Coyle H, Shutler G, Abrams S, Hanniman J, Neylon S, Ladd C, Palmbach T, Lee HC (2003) A simple DNA extraction method for marijuana samples used in amplified fragment length polymorphism (AFLP) analysis. J Forensic Sci 48:343–347

Pentinsaari M, Ratnasingham S, Miller SE, Hebert PDN (2020) BOLD and GenBank revisited – do identification errors arise in the lab or in the sequence libraries? PLoS One 15(4):e0231814

Schield C, Campelli C, Sycalik J, Christopher Randle C, Sheree Hughes-Stamm S, Gangitano D (2016) Identification and persistence of Pinus pollen DNA on cotton fabrics: a forensic application. Sci Justice 56(1):29–34

Stambuk S, Sutlović D, Bakarić P, Petricević S, Andelinović S (2007) Forensic botany: potential usefulness of microsatellite-based genotyping of Croatian olive (Olea europaea L.) in forensic casework. Croat Med J 48(4):556–562

Wilson-Wilde L, Norman J, Robertson J, Sarre S, Georges A (2010) Current issues in species identification for forensic science and the validity of using the cytochrome oxidase I (COI) gene. Forensic Sci Med Pathol 6:233–241

Zhou Y, Cao Y, Huang T, Chen L, Deng K, Ma K, Zhang T, Zhang J, Huang P (2020) Research advances in forensic diatom testing. Forensic Sci Res 5(2):98–105

DNA-Based Analysis of Plant Material in Forensic Investigations

39

James M. Robertson, Natalie Damaso, and Kelly A. Meiklejohn

Contents

J. M. Robertson
Research Support Unit, Federal Bureau of Investigation Laboratory Division, Quantico, VA, USA
e-mail: jmrobertson@fbi.gov

N. Damaso
Biological and Chemical Technologies, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, MA, USA

K. A. Meiklejohn (✉)
Department of Population Health and Pathobiology, North Carolina State University, Raleigh, NC, USA
e-mail: kameikle@ncsu.edu

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_59

Abstract

Plant material has been used in traditional forensics to reveal and support investigative leads and establish the geographic origin. Identification of plant poisons in a rapid timeframe may be crucial in the treatment of accidentally or intentionally exposed victims. The establishment of the time or season of death may be estimated from plant evidence. Genetic analysis of evidentiary plant material can aid in regulatory forensics in the elucidation of illegal trade of endangered species, to monitor the trade of illegal substances, as well as provide information on the adulteration of foods, medicinal and herbal products. Yet, despite the several advantages of having plant identification capability, most forensic laboratories do not offer this service. There are numerous reasons for this deficiency in operational offerings. Case examiners who are experts in human DNA identification generally have no experience with non-human DNA analysis. External experts would have to be used to provide botanical knowledge for an investigation. Examiners would have to be trained for provision of plant DNA evidence in court. Another issue is the perception of a lack of standard procedures to perform botanic investigations that include field collection of evidence and reference samples, DNA isolation techniques, and genetic analysis methods. The aim of this chapter is to provide information on the genetic analysis of plant material that may alleviate concerns to induce a more proactive approach to forensic botany.

Keywords

Taxonomic identification · Individualization · Geographic origin · DNA typing · Forensics · Case studies · Plants

Introduction

Traditionally, forensic laboratories rely on morphological examination of pollen, plant parts (e.g., leaves and flowers), and physical characteristics of materials (e.g., sawdust and plant fragments). However, this approach fails with powders, seed mash, medicinal and herbal products, and plant fragments. It may be difficult to visually identify timber, which makes up a substantial proportion of the 35,000 plant species categorized as endangered by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Likewise, there may be no morphological characteristics of plant material in the stomach contents of a deceased individual. Plant fragments in questioned soil usually cannot be identified using morphology with confidence. However, in all these situations where morphological examination fails, it may be possible that genetic analysis of the plant material will provide information on (1) taxonomy (e.g., genus, species, etc.), (2) individual matching, and (3) geographic origin. Forensic botany can be divided into two categories: traditional (e.g., human criminal cases) and regulatory (e.g., illegal trade, mislabeling). Table 1 lists several genetic methods that have been applied to plant material of forensic interest. A brief description of each of the

Table 1 Summary of the progression in the use of DNA-based methodologies to analyze plant material associated with forensic evidence. Exemplar references are provided for each method and are not listed in chronological order. Abbreviations are as follows: RAPD, random amplified polymorphic DNA; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; AFLP, amplified fragment length polymorphism; SSR, simple sequence repeats; STR, short tandem repeats; EST, expressed sequence tags; MPS, massively parallel sequencing; SNP, single nucleotide polymorphisms; Indels, insertions and deletions; BAR-HRM, barcode DNA high-resolution melting; eDNA, environmental DNA

Method	Forensic use	*Application*			Reference
		Taxonomic identification	Individualization	Geographic origin	
RAPD	Homicide case		√		Carita 2005
RAPD and PCR-RFLP	Adulteration			√	Um et al. 2001
AFLP	Illegal substance		√	√	Carita 2005
AFLP	Homicide case		√		Koopman et al. 2012
Microsatellites (SSRs)	Homicide case		√		Craft et al. 2007
Microsatellites (SSRs)	Conservation			√	Jolivet and Degen 2012
Microsatellites (SSRs)	Adulteration	√			Bosmali et al. 2012
Microsatellites (SSRs)	Conservation			√	Vlam et al. 2018
Microsatellites (EST-SSRs)	Illegal substance	√	√		Vašek et al. 2020
STR Multiplex by length analysis	Illegal substance		√	√	Young et al. 2020
STR Multiplex by MPS	Illegal substance		√		Houston et al. 2018
SNPs	Conservation			√	Finch et al. 2020
SNPs and Indels	Trace evidence	√			Ward et al. 2009
SNPs and Indels	Conservation		√		Dormontt et al. 2020
STRs, SNPs, Indels	Illegal substance	√		√	Roman and Houston 2020
Barcodes	Poisonous plants	√			Bruni et al. 2010
Barcodes	Adulteration	√			Stoeckle et al. 2011
Barcodes	Illegal substance	√			Carrier et al. 2013

(continued)

Table 1 (continued)

Method	Forensic use	*Application*			Reference
		Taxonomic identification	Individualization	Geographic origin	
Barcodes, MPS	Medicines and conservation	√			Coghlan et al. 2012
Barcodes	Medicinal herbs	√			Cui et al. 2020
STRs, Barcodes	Conservation	√		√	Ng et al. 2020
BAR-HRM	Adulteration	√			Osathanunkul and Madesis 2019
BAR-HRM	Illegal substance	√	√	√	Solano et al. 2020
Metabarcoding	Trace evidence			√	Giampaoli et al. 2014
Metabarcoding; eDNA	Trace evidence			√	Fløjgaard et al. 2019
Metabarcoding; eDNA	Trace evidence	√		√	Boggs et al. 2019
Metabarcoding	Trace evidence	√			Timpano et al. 2020

methods used in forensic botany investigations are listed from top to bottom to depict chronological introduction, starting with random amplified polymorphic DNA (RAPD) in the late 1980s, continuing to metabarcoding in the 2010s. References to articles are provided that present examples of the use of the techniques in addressing forensic questions.

History of DNA-Based Analysis of Plant Evidence

The first DNA typing methods did not require knowledge of the DNA sequence. RAPD involves the use of primers of random sequence that bind to analogous binding sites dispersed throughout the genome. After amplification, a group of fragments is produced that can be separated by gel electrophoresis, producing distinct patterns. RAPD analysis was the first technique to be applied in a forensic case for a homicide in 1993, in an attempt to link plant evidence in a suspect's vehicle to trees at the location where the victim was found (discussed by Carita 2005). In this case, there was insufficient collection of reference material to rule out a match by chance. However, RAPD did provide sufficient reproducibility to distinguish ginseng roots originating from Korea and China by using long primers (20-mers) (Um et al. 2001).

Another technique, called amplified fragment length polymorphism (AFLP) analysis, also provides characteristic fragment patterns on gel electrophoresis. It leverages species-specific variations in restriction enzyme recognition sites to produce restriction fragments that can be amplified after ligating adapters to the sticky ends of the fragments (Carita 2005). AFLPs were used to fingerprint a clonally propagated, common weed known as knotgrass, relevant to a homicide. Since knotgrass is ubiquitous, it was expected to be in the vicinity of many areas surrounding and including the crime scenes. In this case, unique genotypes could be reported due to careful reference material collection and analysis of genotypes of the knotgrass plants in the area around the crime scene (Koopman et al. 2012). The major issues involving the use of RAPD and AFLP methods are the requirement for intact non-degraded DNA and the often poor reproducibility between laboratories and analysts (Carita 2005).

A gel analysis technique using simple sequence repeats (SSRs), randomly repeated 1–6 base pair units also referred to as microsatellites, has been used to distinguish plant materials such as timber, food adulterants, pollen, illegal substances (Table 1), and leaves in a homicide (Craft et al. 2007). Several articles involving microsatellite loci have been reported for successful timber tracing, and each uses multiple loci to provide statistical significance for discrimination. For example, five nuclear SSRs were used to discriminate timber and reportedly could distinguish the species only 14 km nearby as well as 836 km distant from the origin of the investigation (Vlam et al. 2018). In a modification of the technique, high-resolution melting of microsatellites and a DNA barcode were used to identify admixtures of 50% in lentils (Bosmali et al. 2012).

Another type of molecular analysis that produces length-based patterns is the amplification of short tandem repeat (STR) sequences, which can be discriminatory among individual plants. For example, a panel of 19 STRs was found useful not only for individualization of the opium poppy but also for determining the geographic origin between eight countries (Young et al. 2020). In a modification of this technique, another group reported using massively parallel DNA sequencing (MPS) (also known as next-generation sequencing (NGS) and high-throughput sequencing (HTS)) of STR fragments to provide more information on allelic variants than a length-based approach provides (Houston et al. 2018).

To accommodate challenging samples, such as those with degraded DNA, single nucleotide polymorphisms (SNPs) and short insertions or deletions of DNA bases (known as indels) have been targeted. These assays have smaller amplicons and are thus optimal for plant samples with low molecular weight DNA such as from processed, milled, or powder materials as well as from old or degraded samples. Additionally, a large panel of SNPs may be used to provide sufficient discrimination. For example, a panel of 17 indels and 68 SNPs was formulated and shown to work with degraded and trace amounts of DNA for molecular identification of grass, often an element of forensic evidence (Ward et al. 2009) and possibly difficult to discriminate morphologically. In the first report that describes a developmentally validated protocol for forensic botany, Dormontt et al. (2020) used an array of SNPs and indels for individualization of timber samples. A combination of STRs and SNPs has also been reported for individualization and geographic tracing of timber species (Ng et al. 2016, 2020).

In the early 2000s, a supplemental molecular analysis invoked sequencing informative regions of the genome for species taxonomic identification. DNA barcoding involves sequencing a small segment of either plastid (i.e., chloroplast), mitochondrial, or nuclear DNA. The resulting unknown sequence is searched against known sequences in a public or private DNA sequence database, to determine the taxonomic identity based on the most similar sequence (Bell et al. 2019). DNA barcodes must have sequence conservation to provide primer binding sites and enough variability within the amplified barcode sequence to allow discrimination between species. In 2009, a few well-researched barcodes were selected as being optimal for plant identification (reviewed by Hollingsworth et al. 2011). A useful application of barcoding for the identification of poisonous plant species used two plastid barcodes (Bruni et al. 2010). Another application of barcoding is food authentication and adulteration testing, whereby the presence of unlabeled ingredients may be revealed (Stoeckle et al. 2011). One must be careful in designing a multiple barcode system, because bias can be introduced by preferential amplification of one barcode over others (Timpano et al. 2020). While barcodes may be capable of distinguishing different species and sample populations, they might not be capable of classification of crop type or geographic origin (Roman and Houston 2020). Barcoding is currently performed using Sanger sequencing and MPS, whereas metabarcoding, a spin-off of barcoding, is performed exclusively on MPS instruments. Metabarcoding is used to simultaneously characterize multiple taxa present in a bulk sample. For example, a questioned soil sample will contain DNA from a community of species (bacteria, plants, and other eukaryotes) that may be diagnostic for the habitat and thus provide

information on its geographical origin. Associating reads generated during metabarcoding to both a target taxon and sample is achieved by adding short tags to the 5′-ends of generated amplicons (known as indexing or barcoding) and downstream bioinformatic processing. This technique allowed the separation of soil samples from 11 different habitats in Denmark (Fløjgaard et al. 2019). In contrast to identifying a species definitively with barcoding, it is sufficient in some scenarios to classify the plant components of the sample to the family or order level with metabarcoding. Additional discussion on the use of plant barcodes and metabarcoding, with several case descriptions for botanical forensics, will be presented later in "Taxonomic Identification" and "Applications of New Sequencing Technologies to Plant Evidence" sections of this chapter.

Traditional barcoding has also been combined with high-resolution melting analysis of the amplicons (identified as BAR-HRM), to simplify taxonomic identification. The principle is to leverage differences in sequence length, GC content, and nucleotide variation that result in different melting curves and indicative melting temperature (T_m) values for the amplicons. The decrease in fluorescence as the double strand amplicon melts is measured by real-time instruments commonly found in many forensic laboratories. The technique was successfully used for certification of timber in a theft and for discriminating drugs (Solano et al. 2016, 2020). BAR-HRM was described as another strategy to analyze low molecular weight DNA, especially when DNA degradation is suspected as shorter amplicons (i.e., mini-barcodes) can be used instead of the larger barcode amplicons (Osathanunkul and Madesis 2019).

There are a number of books and review articles that provide comprehensive general knowledge on the DNA-based analysis of plant material that cannot be included in this short chapter. *Forensic Botany, A Practical Guide* is a collection of ten chapters that include basic botany, plants as evidence, and retrieval and analysis guidelines. *Plant DNA Fingerprinting and Barcoding* is part of the Springer Methods in Molecular Biology Series (Sucher et al. 2012) and provides protocols for the techniques and the necessary equipment for analysis (listed in Table 1). Kress (2017) provided a review covering the history of plant DNA barcoding, phylogenetics, taxonomy, and the applications of barcoding to forensics. Knowledge required to select the appropriate barcode for a specific botanical question is available (Hollingsworth et al. 2011). Issues in the application of DNA barcoding to forensic botany questions, such as commercial product authentication, are reviewed by Mishra et al. (2016). The use of barcoding in forensic palynology, in which pollen can provide information on the time of death, crime scene location, and suspect association with the crime scene, is discussed by Bell et al. (2016). Three reviews are available that focus on DNA metabarcoding. One presents metabarcoding issues to distinguish CITES-listed endangered species (Staats et al. 2016); another one describes bioinformatics challenges, amplification, and sequencing errors (Coissac et al. 2012); and a third specializes in analysis of soil evidentiary samples (Young et al. 2017). Referral to these references on forensic botany will provide answers to questions of the novice law enforcement officer, as well as of the experienced biologist.

Sample Types of Botanical Evidence, Collection, and DNA Isolation

Questioned botanical samples often include the familiar parts of a plant such as leaves, stems, roots, bark, seeds, and pollen; however, when encountered in a forensic investigation, these common parts may be in modified forms such as timber, powders, and herbal and medicinal derivatives (Fig. 1). Considering the different physical matrices and textures, metabolites, chemical inhibitors of PCR, and possible differences in the freshness and age of evidentiary plant material, no standard protocol has been adopted for all questioned samples. Instead, analysts have reported specific protocols tailored to the different forms this evidence can take. The information provided in this section is meant to familiarize the reader with botanical evidence types, best practices for their collection, DNA isolation techniques, and preservation and storage practices.

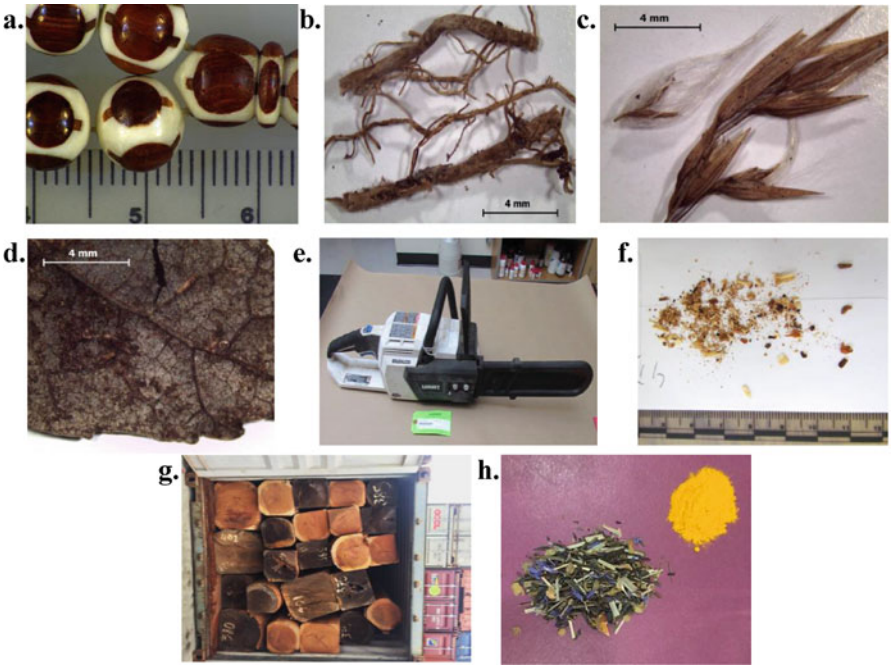


Fig. 1 Examples of plant material associated with forensic investigations: (a) wooden beads; (b) plant roots associated with soil; (c) grass seeds associated with soil; (d) leaf fragment associated with soil; (e), chainsaw; (f), wood shavings; (g), shipping container of timber; and (h), tea and spices. (Pictures kindly provided by Dr. Ed Espinoza (US Fish and Wildlife Service, National Fish and Wildlife Forensics Laboratory) and Microtrace LLC (Elgin, IL, USA))

Sample Types

Botanical evidence samples can be classified in two groups: macro, or visibly identifiable, and trace, which are identified via microscopy or other molecular and chemical methods. The macro sample group primarily includes CITES-protected, endangered whole plants and trees, plant fragments and wood, roots, fruits-seeds, dried and fresh mature leaves, flowers, and algae. This could include leaves taken from a suspect's vehicle (Craft et al. 2007) and blades of grass or seeds collected both near to and distant from a crime scene (e.g., roadway shoulder) (Ward et al. 2009; Koopman et al. 2012). Wood samples can include processed and fresh timber (Rachmayanti et al. 2009; Ng et al. 2016; Dormontt et al. 2020), saw dust, branches, pieces from tree trunks, cross-sectional disks (Paranaiba et al. 2020), growth rings (Solano et al. 2016), bark, and dry shavings. There are multiple descriptions of chemically processed and heated plant material, sometimes presented as a mash from toxic seeds or moassel (hookah) from tobacco (Carrier et al. 2013) or hemp (Houston et al. 2018), along with visible amounts of teas (Stoeckle et al. 2011) and herbal, medicinal, and culinary powders (Mishra et al. 2016; Staats et al. 2016; Kress 2017; Osathanunkul and Madesis 2019; Cui et al. 2020). Articles describe analyzing various forms of opium including a) dried sap and poppy seeds (Young et al. 2020); b) uncooked, cooked (with gummy texture), and chemically processed (Marciano et al. 2018); and c) culinary poppy and pharmaceutical opium (Vašek et al. 2020), along with the derivative heroin (white and brown powder, along with black tar) (Marciano et al. 2018). Also included in the macro group are commercial food grains (Bosmali et al. 2012), fraudulent or misleading products (https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-covid-19-products), and residues from stomach contents obtained during autopsy. Reference samples are available for this group in the form of vouchers containing whole leaves and seeds that are obtained from local herbaria and xylaria for woods.

Trace evidentiary botanical sample materials include small plant and wood fragments, flowers, pollen, powders, and seeds trapped in clothing, shoes, hair, dust, textiles, ropes, and tape used to bind a victim but also can be adhered to car undercarriages and tires (Hardy and Martin 2012; Hall and Byrd 2012; Schield et al. 2016). Evidence may include samples from tape lifts, dust (Craine et al. 2017) collected from unexploded IEDs (Wilks et al. 2017), and personal articles. These items may contain distinctive pollen grains that can be used to limit where the explosive was assembled or places a suspect has visited. Small wood fragments from a door frame on a knife or crowbar may be presented as evidence for a break-in. Soil samples collected from shoes, shovels, tools, rugs, and bulk soil from specific habitats may have traces of rare or geographically isolated plant material (Giampaoli et al. 2014; Boggs et al. 2019; Fløjgaard et al. 2019). Trace plant materials can be used to link the presence of the suspect to a location or habitat, or to eliminate a location as a possible source of the material. However, caution is warranted in making this assumption because microscopically small plant material can also be transferred by secondary contact with another person or sharing transportation used by others, for example.

Collection of Plant Materials and Soil

Collecting visibly large evidence such as intact leaves, flowers, pinecones, woody tissues, timber, and grass at a crime scene may seem like a simple operational task; however, an evidence collection team is advised to contact a local botanist or palynologist to guide them in what items to collect (Hardy and Martin 2012; Hall and Byrd 2012; Wiltshire 2016). Consultation is especially important in planning collection of appropriate and sufficient comparator materials, so that subsequent population genetic analysis of the questioned sample is statistically sound. To avoid clonality, the local specialist will be able to advise appropriate areas associated with the crime scene for collection (Craft et al. 2007). In addition, fragmentary botanical material associated with a crime scene may go unnoticed without a botanist present to identify it.

Botanists use specialized equipment for field collection of large samples, and these items are commercially available or can be assembled from parts obtained from a home improvement retailer. For morphological examinations, leaves are not put in ziplocked bags or plastic food containers, but rather laid flat between two paper sheets or field-pressing cardboard or pressed flat in a special box designed for transport and storage to inhibit morphological changes such as curling. The plant parts either are allowed to dry naturally at ambient temperature during transport or are placed on ice (Craft et al. 2007). An archival article provides practical suggestions to collect plant species in the field based on convenience, speed, and simplicity (Nickrent 1994). For DNA analysis, leaves, cambria, and bark are protected from bacterial and fungal degradation by placing them in ziplocked bags with a drying agent. Recent guidance on field collections that are suitable for DNA analysis has been provided by the Global Timber Tracking Network (https://globaltimbertrackingnetwork.org/2019/02/07/new-gttn-sampling-guide/). Collection of fresh wood samples includes not only the obvious bulk materials, but also small samples such as 1 cm patches of bark and sapwood that are stored with silica gel (Vlam et al. 2018). Freshly collected wood tissue may be placed in 2 mL tubes with silica gel for transport (Jolivet and Degen 2012). The sources listed in section "Guidelines and Standards for DNA Analysis of Plant Evidence" provide in-depth information on the methods of botanical material collection and references of accepted guidelines.

Soil samples may be collected at the origin of a crime scene to provide information on plant material of the habitat for comparison to other habitats. The taxonomy determined for the plant material in the sample may be useful for predicting geographic origin, especially if rare taxa and rich biodiversity are revealed. The depth of soil collection varies in reports of several authors, depending on the questions being asked. For example, samples have been collected from the surface to a depth of 3 cm for mock crime scene evidence (Meiklejohn et al. 2018; Boggs et al. 2019) and from cylindrical cores 5 cm in diameter by 15 cm long to capture the complete biodiversity of meadow and heath habitats (Yoccoz et al. 2012). To link soil to ecological habitats, one report described using a large volume of bulk soil, about 10 L, collected at a depth of 0–15 cm and from cores made in a 40 by 40 m site (Fløjgaard et al. 2019). Soil samples typically are sieved to remove

debris, roots, insects, and stones before transport in a collection container. Prior to storage, samples should have been either air-dried or heat-dried at 60 °C (Giampaoli et al. 2014).

For trace evidence sample collection, several techniques are used. Pollen and small plant fragments can be tape-lifted from the surface of fabrics, vehicles, and tools. The tape-lift method has the benefit of minimizing sample loss during transportation. The samples can be removed from the adhesive using organic solvents, such as hexane (Itamiya et al. 2020), or recovered using sterile forceps. Surface dust samples may be collected with a vacuum filter sock attached to a portable vacuum system available from several manufacturers. A portable vacuum device was described to collect pollen (Schield et al. 2016) as well as filter cassettes (Wiltshire 2016). To minimize contamination during trace evidence collection, it is common practice to use gloves, facemasks, and DNA-free devices.

Another component of evidence collection is the use of a validated method to preserve the samples during transport to the laboratory. However, preservation of the collected plant material does not necessarily follow a standard method. The reviewed reports described methods that were convenient for field collection. Depending on its type, samples have been air-dried, placed in containers with desiccating agents, or stored wet over ice. The use of alcohol, DMSO, or formaldehyde to prevent fungal growth may be acceptable for preserving morphology of the tissue, but their use may decrease the PCR efficiency following DNA isolation. While no report was found that recommended immediate freezing to preserve the sample for genetic analysis, freezing is used for long-term storage once the sample is in the laboratory.

Isolation of DNA from Plant Materials

The steps involved in obtaining DNA from plant material generally include (1) isolating and cleaning of the tissues, (2) grinding, (3) combining the sample with appropriate reagents and extracting, (4) measuring the DNA yield and purity, and (5) performing an additional purification if warranted, to remove remaining PCR inhibitors. In cases involving macro samples, one already has sufficient material to process. However, with trace evidence, the plant material may need to be screened under a microscope first to separate and isolate each putative plant sample one-by-one (Meiklejohn et al. 2018). Plant fragments isolated from soil typically are cleaned with 5% bleach and rinsed with water before beginning the extraction procedure to remove possible fungal contaminants; this step is also recommended for voucher specimens (Meiklejohn et al. 2018). If the plant material had been preserved in alcohol and glycerol, it should be washed off before processing, as any residual preservative could decrease the extraction efficiency. Next, the DNA is released from the plant tissue via grinding. Given plant samples can be dense, fibrous and have thick and non-pliable cell walls, grinding to a powder in liquid nitrogen is often used to efficiently lyse cells and subsequently release the DNA. Samples can be ground manually with a mortar and pestle, but mechanical grinding is rapid and may produce a more homogenized material (Carita 2005). Seeds can be ground in a

commercial blender. Hard tissues may require mechanical shearing. In the literature reviewed, the amount of sample to grind varied with the tissue type: fresh or dried leaves (5 mg–0.5 g), seeds (one to ten whole seeds depending on size or 0.1 g ground powder), fresh sawdust (5 g), and gummy material (e.g., 100 mg of tobacco). This variation in input amount is most likely due to the variety of matrices and various amounts of inhibitors and metabolites specific to each plant species. Thus, it is recommended to consult the relevant published literature to choose the appropriate ratio of plant material to volume of extraction buffer, to ensure successful DNA isolation.

The first popular DNA extraction protocol for plant material used cetyltrimethylammonium bromide (CTAB). This protocol removes polysaccharide and tannin containments in the extract that inhibit restriction nucleases and other DNA processing enzymes (Murray and Thompson 1980). Commercial plant extraction kits became available in the 1990s, some of which were based on the CTAB procedure. Their popularity is due to the provision of an extract free of metabolites, convenience, and reasonable cost. Out of the papers we have referenced in this chapter that described extraction procedures, 42% used the same commercial kit, 23% used either the original or modified CTAB method, and remaining papers used a variety of kits or methods (Fig. 2). Following extraction, DNA quantity is typically measured either by UV measurement or fluorescence, with purity established using the A_{260}/A_{280} ratio (for protein and phenolic contaminants of plant origin) and A_{260}/A_{230} ratio (for protein contaminants). If the quantity is considered too low, it can be increased by whole-genome amplification using a commercial kit (Finch et al. 2020). If inhibitors such as phenolic compounds are co-isolated, the extract will be insufficiently pure and will reduce the success of downstream DNA analysis (Meiklejohn et al. 2018). These inhibitors can be removed by adding 1% (w/v) polyvinylpyrrolidone (PVP) to the extraction buffer (Jhang and Shasany 2012) or by

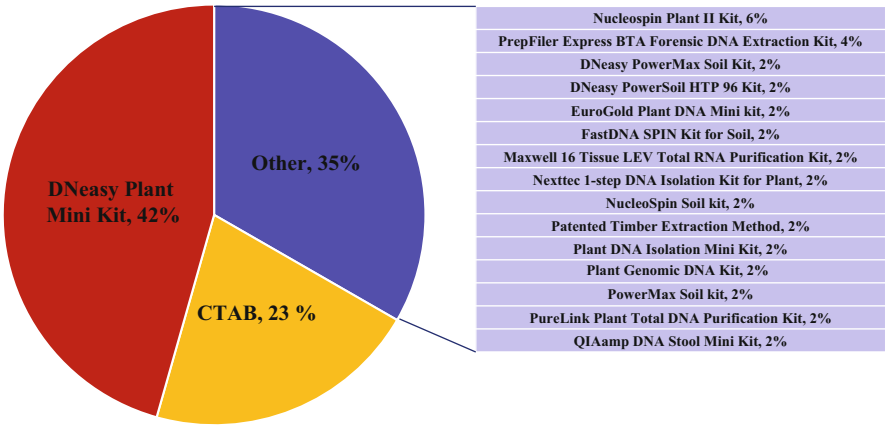


Fig. 2 Summary of plant DNA isolation methods employed in chapter references. Pie chart illustrates the relative proportions of the two popular plant DNA isolation methods with the remaining kits listed in the table

using a polymerase designed to work effectively in the presence of such inhibitors (Meiklejohn et al. 2018). However, for wood samples, PVP is added to the lysis buffer due to their relatively high tannin content. When further purification was required, two reports described the use of two different template cleanup kits. A filtration method was also used to concentrate DNA in cases of low yield (Marciano et al. 2018). In summary, one can generally conclude that more than half of the scientific community prefers to use a variety of kits and procedures for DNA isolation, thought to be tuned to their particular plant samples.

There is no standardized isolation protocol for all plant tissues because of the encountered biodiversity but also because the extraction efficiencies of different tissues may vary considerably. An all-encompassing protocol may bias taxa detection and genotyping, especially where there is insufficient DNA from one taxa or uninhibited extracts (Young et al. 2017). For example, wood is very dense, making extractions from a dry sample challenging (Ng et al. 2016). A step-by-step extraction protocol for DNA from heartwood is available (Vlam et al. 2018). It has been demonstrated that chloroplast DNA quantity decreases as samples are taken from the outer sapwood region toward the inner heartwood region (Rachmayanti et al. 2009). The quantity of samples available to isolate DNA is also a consideration against a standardized procedure, since it varies with the type of sample under investigation. To reduce the hands-on time and improve consistency, an automated protocol was developed for wood in which multiple sawdust samples were pretreated via cryogenic maceration and ground to a powder while wet (Paranaiba et al. 2020). Protocol selection may be dependent on the ultimate use of the DNA. For example, extraction without grinding was recommended for obtaining high molecular weight template suitable for whole-genome sequencing (Jagielski et al. 2017). Another issue is extraction from preprocessed samples, because the treatment may reduce the DNA quality and lower the yield. One example is the extraction of DNA from some heroin samples, including those that have been preprocessed. To obtain inhibitor-free DNA from such samples, commercial extraction kits including those that use magnetic bead or silica column technologies were evaluated for heroin samples. The resulting extract was subjected to membrane filtration to concentrate the template (Marciano et al. 2018). A nondestructive extraction method may also be needed in instances where sample morphology needs to be preserved. One such protocol was developed for pollen grains, where ethanol in a sealed capillary tube held at 60 °C permitted DNA isolation without grain lysis (Kelley et al. 2020). Finally, researchers developed a protocol to reduce DNA yield variability from soil samples and recommended it for soil metabarcoding studies (Minich et al. 2018).

When isolating DNA from plants, the same working conditions implemented to reduce contamination when processing human DNA samples should be followed. Surfaces and pipettes should be cleaned with bleach followed by alcohol to remove any existing contaminants in the workspace (e.g., residual biological materials). Basic hygiene principles must be deployed, such as wearing gloves and lab coat. Researchers performing highly sensitive procedures need to take special precautions to prevent contamination, particularly with low biomass samples. Tubes should be handled without touching the lid and opened one at a time, touching the small tip end

and not the underbelly of the cap (Bell et al. 2019). One report recommended extracting in single tubes instead of plates to avoid cross contamination (Dormontt et al. 2020). The same authors warned about producing aerosols from bead beating and transfer of powders. Working in an uncluttered laminar flow hood and organizing both pre- and post-amplification space are necessary. Hood filters should be replaced regularly because they could trap seasonal pollen in the laboratory air and become a source of contamination or become ineffective after reaching filtration capacity. The same is true for laboratory coats, which can also trap plant debris and pollen and thus should be replaced or laundered regularly.

Applications of Analyzing Plant Material Associated with Forensic Evidence

The loci and methods employed to analyze plant evidence are highly dependent on the questions being asked. Generally speaking, analysis falls into one of three key areas: taxonomic identification, geographic origin, or individualization (Fig. 3). In this section we will explore the commonly used loci and typing methods implemented in each scenario, along with examples of their use in forensic practice.

Taxonomic Identification

Taxonomic identification to a particular species or higher taxonomic level is a crucial first step in analyzing botanical evidence to aid in forensic investigations. Traditionally, identifications have been conducted using physical characteristics that require largely intact material and expert taxonomists. However, in typical forensic samples, diagnostic characteristics are often absent or small, limiting classification only to higher taxonomic levels. DNA-based methods have been demonstrated to be a fast, easily implementable, affordable, and reproducible approach for taxonomic identification of forensic-type samples that are often fragmented, degraded, and compromised (e.g., food processing conditions). Plant identification can aid in forensic investigations by providing linkages between crime scenes and individuals, as well as determining cause of death. For instance, identification of vegetable matter in a victim's stomach contents can be used to verify suspect's alibi or constrain time since death (Bruni et al. 2010; Lee et al. 2020). Further, the identification of small plant fragments associated with soil such as grass seeds and pollen can provide useful investigative information (e.g., location of the crime scene and time of the year) due to their ubiquity in the environment, spatial and temporal variability, and transferability (Ward et al. 2009; Bell et al. 2016, 2019; Meiklejohn et al. 2018). Taxonomic identification using DNA-based analysis has also demonstrated to be a powerful tool to aid the investigation of illegal drugs (Carrier et al. 2013; Paranaiba et al. 2019), authentication of food and herbal medicines (Stoeckle et al. 2011; Coghlan et al. 2012; Osathanunkul and Madesis

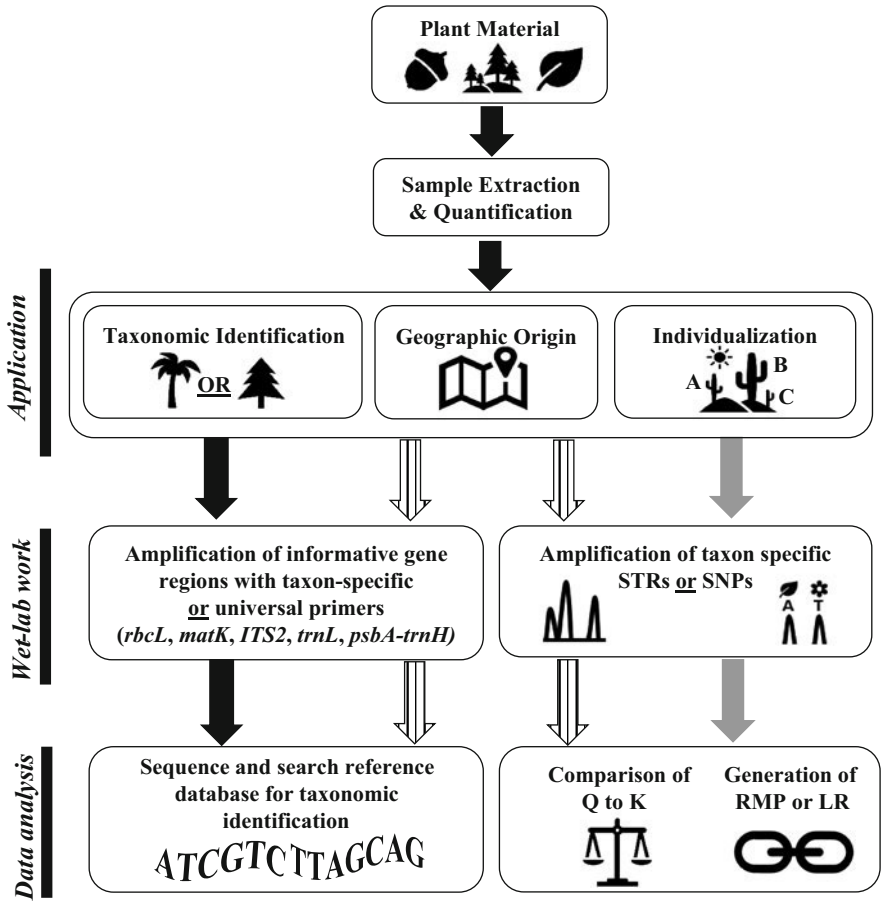


Fig. 3 Simplified flowchart for plant DNA analysis for taxonomic identification, geographic origin, and individualization. Abbreviations are as follows: STRs, short tandem repeats; SNPs, single nucleotide polymorphisms; Q, questioned sample; K, known sample; RMP, random match probability; LR, likelihood ratio

2019; Uncu and Uncu 2020), and enforcement of illegal trade of endangered or protected species (Ng et al. 2016, 2020).

DNA barcoding is the most common method for taxonomic identification by targeting short yet informative regions of DNA that permit species discrimination. DNA barcoding ideally targets one or more standard loci that are usually highly conserved and easily amplifiable, produce high-quality sequence data (i.e., limited homopolymer stretches), and have high discriminatory power (Hollingsworth et al. 2011). To date, the search for a unique barcode for universal plant identification has been challenging as no single region meets the requirements in all taxonomic groups. Lower success of species discrimination for plants using DNA barcoding has been attributed to multiple factors such as hybridization, polyploidy, life history, breeding

system, species history, level of taxonomic splitting, and seed dispersal mechanism (Hollingsworth et al. 2011). In 2009, the Consortium for the Barcode of Life (CBOL) Plant Working Group recommended the two-loci combination, ribulose 1,5-bisphosphate carboxylase (*rbcL*) and maturase K (*matK*), as the plant barcode given these loci together permitted 70–75% species level discrimination (Hollingsworth et al. 2009). Additional supplemental markers (*trnL*, *psbA-trnH*, and ITS2) were recommended in instances where species resolution is not possible with the two-loci approach and have been widely used as supplementary loci (Cui et al. 2020; European Network of Forensic Science Institutes 2015). Consideration of the appropriate markers based on question of interest and application is vital. For instance, *rbcL* mini-barcodes and *psbA-trnH* (~200 bp) may be sufficient for identification of highly degraded DNA and have been used for higher-level identifications (e.g., genus, family, etc.) for CITES-listed plant species (Dormontt et al. 2015; Staats et al. 2016). Even though *matK* can permit species-level discrimination, amplification and sequencing are less consistent than with *rbcL* considering a) DNA degradation can limit the recovery of the ~850 bp region typically targeted, and b) universal primer design is difficult and therefore class-specific primer pairs are needed at a minimum. With a forensic unknown, DNA degradation can impact data recovery, and selection of an appropriate primer pair for PCR is difficult. Understanding the limitations of the barcode markers is equally important. For instance, ITS2 primers are highly similar to the fungal ITS sequences, leading to off-target amplification and possible misidentification (Mishra et al. 2016). Implementing a tiered approach to barcoding can be useful. As an example, amplification and sequencing of *rbcL* is simpler and can still permit taxonomic identification to a higher level (e.g., family or order). The information gleaned from *rbcL* can subsequently assist in the selection of most appropriate DNA region(s) and associated primer pairs to verify the presumptive identification. Such a multi-locus approach for plants has been demonstrated to greatly improve the identification capability and reliability (Meiklejohn et al. 2019).

A DNA barcoding database is a crucial component for accurate and reliable taxonomic identification. Currently, Barcode of Life DataSystems (BOLD) and GenBank are the two main public databases of DNA barcode data. Ideally sequence databases used for taxonomic identifications should consist of sequences generated from high-quality vouchered specimens housed in museum collections and herbaria, which were quality checked (e.g., vector contamination, proper translation for coding regions, correct bibliographic citation, and correct taxonomy) before inclusion. In public sequence databases, incorrectly labeled taxonomic sequences have been reported, which are primarily the result of poor-quality DNA sequences, contamination of the DNA sample (e.g., fungal endophytes), PCR-based errors (e.g., unintentional sequencing of pseudogenes), and obsolete or dated nomenclature (de Boer et al. 2014). Searching algorithms, such as GenBank's basic local alignment search tool (BLAST), provide distance measures between two sequences and reports which sequence in the database are most similar to the unknown. These fast distance-based algorithms are limited to the database entries and therefore can provide misleading or ambiguous identifications. For instance, a false positive can

result if a sequence for the true source species is not present in the database. Moreover, ambiguous identifications can result when there is overlap in the intra- (within) and interspecific (between) genetic divergence, such that multiple species will share the same barcode sequence. Bell et al. (2019) found a small number of false positives due to contamination or misidentification, many of which could be eliminated with more stringent laboratory and data analysis methodologies, along with a more comprehensive reference database. Meiklejohn et al. (2019) observed ambiguous correct matches (i.e., multiple records with different taxonomic names that have the same top match statistics) when searching against both GenBank and BOLD and reported that multi-locus barcoding approaches provide greater precision and reliability. Overall, careful interpretation and reporting of search results is vital and will be discussed in section "Guidelines and Standards for DNA Analysis of Plant Evidence."

Geographic Origin

Somewhat akin to determining human ancestry, pinpointing the origin of plant evidence at either a population or regional scale is a question often posed in forensic casework. Geographic origin is primarily used to identify whether a violation or crime occurred (e.g., determining whether the species was sourced from a trade-restricted area) and for investigative leads for illegal activity (e.g., the DEA's Heroin Signature Program focused on determining where the poppy was grown). A modified application of geographic origin is for plant material associated with geologic materials (soil, dust), which can be used to constrain the circumstances of a crime. Identification of the plant community present in such a sample could serve to limit the search area in a homicide investigation. For example, Craine et al. (2017) characterized plants associated with outdoor dust and reported that samples from urban lower latitudes with hotter climates had a higher proportion of moss and grass, whereas those from rural northern latitudes with cooler climates had a higher proportion of pine and conifer. In a real forensic scenario, interpretation is likely not as straightforward given that a murder victim could have been transported or dragged over various soil types. A plant community recovered for origin prediction in such a scenario would likely be a mixture. Morphological examination of the trapped pollen grains on the victim's clothing or hair may indicate a certain pathway was taken to a possible burial site. Machine learning models of the genetic analysis results could assist with such interpretations (Boggs et al. 2019), enabling the use of plants to predict sample origin even for highly mixed samples.

For forensic purposes, a population is defined as a "group of individuals of the same species that live and breed together in a given geographic region, which gives rise to allele or haplotype frequency differences among populations" (Moore and Frazier 2019). Given this, assigning an individual to either a population or region depends on specific regions of the genome that exhibit spatial genetic structure. Given that plants are immobile, a "breeding" population will largely be defined by the mechanism in which a species achieves fertilization. For example, species that

use insects, mammals, or birds for pollen dispersal can successfully breed with individuals several miles apart, whereas species that rely on wind or water for dispersal typically are restricted to individuals within a 500 m radius (Shivanna and Tandon 2014). Additionally, both natural and man-made barriers can restrict the ability for individuals to successfully interbreed, thus creating population structure that can be exploited for determining source origin.

Population or region assignment of an individual plant is commonly achieved using nuclear STRs or SNPs, or SNPs from either the chloroplast or mitochondrial genomes (Table 1). To ensure assignments have statistical confidence, a comprehensive database comprised of genotypes from individuals across a species known distributional range is necessary, with a minimum of 500 samples per population recommended (Carracedo et al. 2013). Thus, while there are published examples of plant origin assignment for forensic applications with fewer samples in the genotype database, this analysis is typically only implemented in specific cases given the substantial investment needed to create a large database. Origin assignments also rely heavily on specialized statistical software (most commonly STRUCTURE) to assign an individual to a specific population and provide statistical weight to the assignment (Moore and Frazier 2019).

Individualization

As discussed in previous chapters, forensic laboratories globally have used DNA-based approaches to permit individual matching of human biological materials, such as blood, saliva, and sexual fluid, associated with forensic evidence since the 1990s. Similar to human casework, individualization of plant evidence has been pivotal for prosecution and investigative leads associated with human and wildlife forensic casework. For instance, plant individualization can be used to link a suspect to a crime scene (Craft et al. 2007; Koopman et al. 2012), link seized wood material back to the stumps of illegally felled trees, and verify chain of custody (Dormontt et al. 2015).

Completing individualization comparisons with plant evidence comes with a range of challenges not encountered when dealing with human biological evidence. First and foremost, the reproduction strategy of the plant will determine whether individualization is possible (Koopman et al. 2012). For plants that self-breed (10–15% of seed plants (Wright et al. 2013); e.g., orchids, rice) or are clonally propagated, the resulting progeny will be very similar or identical (except for rare mutations), respectively. Plants that reproduce via crossbreeding will produce genetically variable offspring (Koopman et al. 2012). The life span of the plant, annual or perennial, can also impact the success of individualization. If the evidence in question is from an annual plant, timely collection of reference samples to permit questioned-to-known comparisons or the generation of a population database is paramount before die off (Hall and Byrd 2012). An additional challenge when collecting known samples from the crime scene for comparison is the size and density of the plant species (Hall and Byrd 2012). For instance, if the size of the

questioned species is small and there are hundreds of individuals at the crime scene (e.g., grasses), sufficient sampling to identify the source individual with certainty may not be feasible. Alternatively, if the plant is large (such as a tree) and sparsely distributed (i.e., acres to hectares), adequate sampling within proximity to the crime scene might be logistically difficult.

Another key challenge is the vast number of different plant species that could be encountered in casework. Unlike human biological evidence which only deals with a single species, there are approximately 374,000 plants globally which could in theory be associated with forensic evidence (Christenhusz and Byng 2016). Highly polymorphic loci targeted for individualization typically are species-specific, meaning that typing will only be successful in closely related taxa (Koopman et al. 2012). Thus, initially determining taxonomic identity is imperative to ensure appropriate loci/markers are typed for individualization. For species more commonly encountered in casework, such as those used to make drugs (e.g., cannabis and poppy) and CITES listed timbers, informative loci that permit individualization have been identified (Ward et al. 2009; Houston et al. 2018; Roman and Houston 2020). However, the associated genotype databases are typically small (<1000), because sampling individuals from natural populations requires expansive, labor-intensive, and costly field work. In investigations where the plant evidence is pivotal, but loci are unrecognized to permit individualization, rapid screening of the genome and development of comprehensive genotype databases are necessary (Tnah et al. 2015). This type of development is typically outside the scope of most forensic laboratories, such that partnering with academic institutions or industry is recommended.

Studies listed in Table 1 highlight the diversity in markers that have been used in forensic casework or developed for use in casework for plant individualization. These markers primarily include AFLPs, STRs, SNPs, and INDELs from the nuclear genome. Statistical approaches commonly used in human individualization (i.e., random match probability (RMP) and likelihood ratio (LR)) are also implemented when analyzing plant evidence. Depending on the marker used, phylogenetic trees and distance matrices can also be used to visualize the differences between questioned and known samples (e.g., Korpelainen and Virtanen 2003). The paper by Dormontt et al. (2020) can serve as a guide for creating a developmentally validated protocol for examination of plant material evidence (specifically timber) for individualization.

Case Studies

Case Study One: It is in his shoes! *Cynodon dactylon* DNA connects the suspect with the crime scene

Alejandra Figueroa
Policía de Investigaciones de Chile

- An elderly woman was attacked and strangled in her bedroom. The police found a shoe print next to the corpse. Based on other evidence, a suspect was arrested and

his shoes recovered. The shoe print from the crime scene matched the shoes recovered from the suspect.

- A fresh plant fragment was discovered associated with one of the shoes. The victim lived in a neighborhood surrounded by many different types of vegetation, such as trees, ferns, lawn, wild grass, and various medical herbs. To determine the species of the botanical evidence, DNA was isolated from the fragment and sequenced by high-resolution melting analysis (HRM) based on published internal transcribed spacer (ITS) primers.
- The ITS sequence was aligned using MEGA 5 software and searched against GenBank. The plant fragment was identified as a type of grass, *Cynodon dactylon*, commonly known as decorative garden lawn. Based on this information, the area surrounding the crime scene was searched again, finding the same type of grass only on the neighbor's property.
- It was concluded that the attacker ran away after committing the crime, crossing the neighbor's garden where the shoe collected the aforementioned botanical evidence.
- These findings were presented in the trial, and the court sentenced the offender to 13 years in prison. This case illustrates the role of forensic botany to connect a suspect with a crime scene.

***Case Study Two*: Botanical evidence links suspects to a dumping site: a successful combination of poplar and wild lettuce**

Irene Kuiper, Monique Wesselink, Stefan Uitdehaag
Netherlands Forensic Institute (NFI), The Hague, The Netherlands

- On a waste land near Rotterdam (the Netherlands), a partly burned body was found by a pedestrian walking his dog. At the site, a tire track was found and lifted. During the subsequent investigation, camera images were collected from the routes leading to the site. Using these images and the tire track, a Volkswagen Transporter van was identified. The apprehended suspects admitted driving the van but denied involvement with the crime.
- A 50-cm-long withered stem and some leaves were secured from the underbody of the van, together with some leaves from inside the van. Due to the condition, length, and number, it was believed this plant material could not have persisted underneath the van for a long time.
- Morphological examination of the stem only permitted identification to the family level, due to a lack of distinguishing features. Using green tissue from the stem, DNA analysis was performed. When comparing the generated nucleotide sequences of the barcodes ITS and *trnLF* in GenBank, the species was tentatively identified as wild lettuce (*Lactuca* sp.).This is a pioneer weed common to disturbed environments, but not to places were vehicles are usually driven. Morphological examination of the leaf permitted identification as poplar (*Populus* sp.).
- Upon reexamination of the site, wild lettuce and leaves from the surrounding poplar trees were found to be in abundance.

- The municipal landscaping department kept a detailed overview of all planted (poplar) trees in the area, including species name, age, arborist, and maintenance reports. Based on recorded information, the following were collected as reference samples for analysis: a) poplar leaves from the site, b) material from all poplar trees surrounding the site and on the route toward it, and c) poplar trees near the location of the van seizure.
- Samples collected from the poplar trees and leaves were typed with a validated STR marker system [ENFSI BPM APS 01] and resulted in a match between the leaves from the van to those collected both at and immediately surrounding the site. Other sampled trees had different genotypes.
- To establish the value of the botanical evidence in this case, the laboratory evaluated the condition of the plant material secured from the underbody of the van, together with the ecology of wild lettuce and the occurrence of the poplar genotypes. Taken together, the chance of finding this combination of botanical traces was considered more likely (out of a range with equally, some more, substantially more, far more) under the assumption the van had been at the crime site than under the assumption the van had been at other sites within the Netherlands.
- The suspects were convicted for dumping the body at the site.

***Case Study Three*: Individualization of bigleaf maple (*Acer macrophyllum* Pursch) supports illegal logging prosecution**

Andrew Lowe, Eleanor Dormontt,
Advanced DNA Identification and Forensics Facility, University of Adelaide

- Illegal logging is predominantly a tropical forest problem, but it also happens in the USA with an estimated value of US$1 billion annually.
- Bigleaf maple, *Acer macrophyllum*, is a large deciduous tree that grows along the west coast of the USA from California in the south to Washington in the north. The timber is often "figured" with a unique patterning highly sought after for decorative wood work. In response to a spate of bigleaf maple poaching in the State of Washington, the USDA Forest Service sought the help of Double Helix Tracking Technologies to help bring the perpetrators to justice.
- Offcuts of wood were seized from a suspect's timber mill which were thought to be remnants of illegally felled bigleaf maple trees. Samples were also collected from the remaining stumps of stolen trees in the National Forest to enable comparison.
- Double Helix Tracking Technologies together with the University of Adelaide funded by the World Resources Institute, sought to develop an individualization test for bigleaf maple that could determine whether the seized samples originated from the illegally felled trees.
- A genetic reference database of 394 individuals taken from 43 sites across the distribution of the species was created with the greatest number of samples taken from where the thefts occurred (Gifford Pinchot National Forest). Using a combination of MPS and MassArray approaches, a set of 128 SNP loci and

3 INDEL loci were genotyped in every individual (Dormontt et al. 2020). The markers were forensically validated to ensure suitability for legal application, in accordance with the SWGDAM Validation Guidelines for DNA Analysis Methods.

- The case samples were analyzed under forensic conditions, in accordance with the SWFS Standards and Guidelines for Forensic Botany Identification. A genotype match was found between the mill offcuts and the stump of an illegally felled bigleaf maple in Gifford Pinchot National Forest. Based on the reference database, the chance of this match being obtained from two samples which did not originate from the same tree (or clone) was less than one in five septillion (5×10^{24}).
- In 2015, a case was brought against the mill owner and three other defendants purported to have been responsible for the illegal felling. All defendants eventually pleaded guilty and received jail time of between 30 days and 15 months, along with substantial fines for the mill owner. The case represented the first domestic prosecution to apply forensically validated DNA methods for timber individualization.

***Case Study Four*: DNA analysis verifies timber supply chains**

Eleanor Dormontt, Arif Malik, Andrew Lowe
Advanced DNA Identification and Forensics Facility, University of Adelaide

- The *Lacey Act* (2008) is a fact-based statute, specifying that the trade of any illegal products is a violation. For timber, compliance with the *Lacey Act* generally requires "due care" on the part of traders; all reasonable steps must be taken to minimize the risk of illegal timber entering supply chains. Penalties for illegal timber vary according to the "due care" taken, and hence traders have the responsibility and incentive to police their own supply chains in order to demonstrate compliance.
- Between 2016 and 2018, Double Helix Tracking Technologies worked with their industry clients to utilize DNA barcoding as an additional tool to support "due care" efforts on the importation of timber to the USA with the trade name "meranti" (sourced from South East Asia). To assist the client with correctly declaring the species at import and determining the reliability of species claims made by suppliers, a selection of timber purported to be *Shorea* spp. (Dipterocarpaceae family) was sent to the University of Adelaide for identification via DNA barcoding.
- Fourteen of the 26 provided samples (65%) were successfully Sanger sequenced for *trnL*, and the results were compared to the NCBI databases using the nucleotide BLAST function.
- Of the 14 successfully sequenced samples, a) 9 (64%) returned species identifications consistent with the claim of *Shorea* spp., b) 1 sample returned a result of *Hopea* (another genus within the Dipterocarpaceae family), and c) the remaining 4 samples (29%) were from different families all together. Potential species names were provided for all samples where possible, with the caveat that for closely

related species (such as many of those in the *Shorea* genus) barcode sequences may be shared.
- The results were used as part of the clients' "due care" procedures to correctly declare species for import and to assess the reliability of their suppliers' claims regarding the species being sold. This example demonstrates how DNA analysis of plant material can be used proactively by industry to support compliance claims, as well retrospectively by law enforcement to detect illegal activity.

Applications of New Sequencing Technologies to Plant Evidence

The advent of MPS nearly a decade ago has revolutionized molecular biology, making sequencing faster and more affordable per base. MPS offers several benefits to forensic biology over traditional typing methods (i.e., Sanger sequencing and STR analysis). Indexing permits multiple samples to be processed simultaneously, and the high sensitivity of the sequencing platforms facilitates recovery even from low quantity DNA samples. Using MPS, hundreds of loci can be multiplexed for typing in a single reaction as opposed to the approximate two dozen simultaneous STR targets currently capable with capillary electrophoresis (CE). As commercial companies have developed human-focused forensic MPS assays, crime laboratories globally are evaluating and validating these for casework. Given some crime laboratories have acquired MPS instruments, they may be receptive to the analysis of non-human biological materials associated with forensic evidence. Several research groups have demonstrated that MPS can be used to streamline the identification of plant species present in bulk or mixed sample types, such as surface soils (Young et al. 2017; Boggs et al. 2019), dust (Craine et al. 2017), food products (Bruno et al. 2019), herbal medicines, and pollen (Bell et al. 2016), using metabarcoding. Compatible with the Ion GeneStudio S5 system, a commercial assay and bioinformatics workflow was developed (ThermoFisher Scientific; A38456) to streamline taxonomic identification of plants in food mixtures for authentication purposes. To simplify SNP genotyping of Spanish cedar samples, Finch et al. (2020) used hybridization capture (a method in which biotinylated RNA "baits" complementary to the DNA of interest are used to isolate target DNA prior to MPS) for origin assignment. Finally, to efficiently get data from a large number of loci, shotgun sequencing-based "genome skimming" can be a useful alternative for taxonomic identification (Bohmann et al. 2020).

New concerns have appeared following the introduction of MPS. For example, low-level alleles may not be detected when using MPS, resulting in a false negative. This often is the result of pooling too many samples for sequencing on a single run, which should be adjusted to ensure low-level alleles in a mixture have higher read counts and are subsequently detected. Obtaining adequate read lengths of barcode sequences is also important to ensure sufficient variation is captured to determine taxonomic identity with confidence. If the bioinformatics analysis pipeline filters out reads below a certain length or targets with a set number of reads, then species not meeting these thresholds will not be characterized. Species may be missed in

environmental samples with degraded DNA, if long barcode regions are used for taxonomic identification. To address this, researchers are shifting toward using shorter barcode regions to ensure the entire community is adequately characterized. Further, if the analysis of samples from two highly similar habitats is required, one can select a protocol that might provide a sufficiently high number of reads to capture the low-level variation necessary to differentiate them. It is worth noting that as the technology continues to improve sensitivity, the scientific community might develop standard protocols to address many of these issues (as discussed in section "Guidelines and Standards for DNA Analysis of Plant Evidence").

Guidelines and Standards for DNA Analysis of Plant Evidence

Best practice guidelines on the collection of botanical material are available. The Smithsonian Institution (Washington, DC) published a best practice guideline for collecting botanical vouchers and tissues that can be referred to when preparing a reference sample collection (Funk et al. 2017). The European Network of Forensic Science Institutes (ENFSI) has published a *Best Practice Manual for Forensic Comparison of Soil Traces* (ENFSI-BPM-APS-02, 2019). The manual includes suggestions on how close and far to collect comparator samples from the scene, how to avoid contamination and label samples, along with procedures for sample preservation and packaging (http://enfsi.eu/documents). The US Department of Agriculture (USDA) has prepared a plant materials collection guide (St. John et al. 2010). Other guidelines from international government agencies (e.g., Australia and British Columbia, Canada) are also available to the public online. The book, *Forensic Botany, A Practical Guide*, has a chapter on evidence collection and analysis (Hall and Byrd 2012). There are many photographs illustrating mock crime scenes and actual clandestine grave sites, evidence collection devices, and examples of proper documentation. A guide for procedures applicable to forensic botany in general and protocols for forensic palynology, based on decades of field work with law enforcement agencies in various regions and habitats of the UK, is recommended for both the practitioner and novice alike (Wiltshire 2016).

It is essential that forensic casework, regardless of discipline, be performed to recognize standards that meet criteria for analyzing evidence that is admissible in court. Standards and guidelines applicable to the DNA-based analysis of plant material exist and have been developed by bodies such as the Organization of Scientific Area Committees for Forensic Science (OSAC, US), the Scientific Working Group on DNA Analysis Methods (SWGDAM, US), and ENFSI (Europe). These bodies are composed of leading forensic practitioners from private and public forensic laboratories, industry, and academia. The Society for Wildlife Forensic Science (SWFS, US) has developed *Standards and Guidelines for Forensic Botany Identification* (www.wildlifeforensicscience.org), and ENFSI published the *Best Practice Manual for the Application of Molecular Methods for the Forensic Examination of Nonhuman Biological Traces* (ENFSI-BPM-APS-01, 2015). Both of these documents provide a framework for evidence collection, data analysis,

interpretation, and reporting. More generally, the OSAC has also developed standards and guidelines covering topics including minimum general molecular biology laboratory standards (evidence handling, training, and DNA procedures; ANSI/ASB 019 and 048), validating STRs (ANSI/ASB 046) and Sanger sequencing primers (ANSI/ASB 047), and report writing (ANSI/ASB 029). These documents are publicly available (www.asbstandardsboard.org and www.nist.gov). Further, SWGDAM has developed several quality assurance standards (www.swgdam.org), providing a foundational framework for training, validating, and testing that may be applicable for the DNA-based analysis of plants. Another source of standards is derived from researchers in the environmental DNA community, whose goal is to reduce sources of variation in metabarcoding results. A protocol was developed by the Earth Microbiome Consortium to standardize data generation for microbial communities in soil samples (www.earthmicrobiome.org/protocols-and-standards). While this standard protocol was designed for microbial community analysis from soil, it could be adapted for the analysis of soil flora as well. The developed standards and guidelines discussed here should be reviewed by laboratories considering implementing DNA-based analysis of plants into forensic casework.

Alternatives to DNA for Analyzing Plant Material Associated with Forensic Evidence

Genetic analysis of the plant material can aid the investigation by either providing support to the conclusions reached from other analytical techniques or disputing questionable conclusions. For example, genetic analysis can augment soil geological analysis by showing habitat differences in soil samples that are similar geologically or by complementing the conclusions reached from the geological analysis (Giampaoli et al. 2014). Taxonomic identification by DNA analysis can support conclusions reached by chemical profiling such as those determined by mass spectrometry and isotope ratio analyses, although sometimes the isotope ratios do not mirror the DNA results (Vlam et al. 2018). Spectroscopic methods have been used to classify plant materials, such as pollen grains that have been distinguished by FTIR spectroscopy (Diehn et al. 2020). Instead of increasing the number of DNA markers targeted, chemical analyses such as Raman spectroscopy can be used to improve the taxonomic resolution reached by genetic analysis (Schulte et al. 2008). High-performance thin layer chromatography coupled with mass spectrometry (MS) has been used for authentication of *Ginkgo*-containing herbal products (Avula et al. 2016). Near-infrared spectroscopy has been used to discriminate between macroscopically similar woods for identification of endangered species listed in CITES (Pastore et al. 2011; Braga et al. 2016). Over the past decade, ambient ionization methods such as direct analysis in real-time high-resolution mass spectrometry (DART-MS) have become a popular analytical tool for various forensic applications. In cases that require quick identification of the plant evidence, DART-MS may be a good choice because it does not require sample preparation prior to making measurements and it can acquire multiple spectra rapidly. For instance, DART-MS has

been used to characterize CITES-listed species (Lancaster and Espinoza 2012), authenticate olive oil samples (Vaclavik et al. 2009), and differentiate between toxic seed species (Lesiak et al. 2015) and psychoactive plant products (Musah et al. 2015). In difficult or complex cases, the analysis of plant evidence can benefit from a multidisciplinary approach, in which a combination of visual, chemical, and DNA-based methods is conducted to provide a synergistic, reliable, and comprehensive assessment of the botanical evidence. For instance, the Brazilian Federal Police used gas chromatography-coupled mass spectrometry and DNA barcoding methods to assess seized fragmented plant material that was suspected to contain illicit substances (Paranaiba et al. 2019). A report has discussed using different diagnostic methodologies synergistically for timber identification to support identification results from genetic technologies (Dormontt et al. 2015). Features considered by the authors included cost, speed, equipment, requirements for prior knowledge of the sample, and obstacles to implementation.

Conclusions

As novel procedures were developed in molecular biology, some were applied to analyze plant material associated with forensic investigations. While the earliest technologies demonstrated discrimination was possible, the analysts addressed several issues to bring them up to forensic standards. For example, when visualizing STRs run on electrophoretic gels, contamination resulted in miscellaneous, unexplained bands on gels. Furthermore, mixtures resulted in multiple bands in a single lane, shared bands, and allele dropout due to degradation, minute samples, primer bias, and preferential amplification. These issues were addressed by improving extraction procedures with reliable commercial kits that help reduce co-extraction of inhibitors, using better designed primers to increase sensitivity, and introducing standard working conditions to avoid contamination. Following the introduction of MPS, new concerns appeared such as allele/taxa dropout from DNA degradation or low template, adequate read lengths to permit taxonomic discrimination, and suitable bioinformatics that ensure all data is appropriately interpreted. Researchers and practitioners are developing standards and guidelines to address these concerns.

Differences in DNA isolation efficiencies from the various sample types have added to the technological issues in forensic botany. The method selected to isolate the DNA is of paramount importance and should only be selected after thorough consideration of the sample type and possible inhibitors. Low amounts of DNA and residual inhibitors in the extract will influence the genotyping success and library preparation for MPS. Considering this, one must be constantly on alert for recent papers on enhancements in DNA isolation methods, as exemplified earlier in the chapter for processed opium and heroin samples along with nondestructive treatment of pollen grains. If the purity of the template is questionable, one can perform a secondary cleanup procedure before using the DNA for genetic analysis.

A requirement for all the genotyping procedures presented in the previous sections is a representative population database. Unlike for human DNA analysis, a large database curated to forensic standards for plants does not exist. Ad hoc databases are developed by researchers for specific samples of forensic interest, such as for regulated timber. A large database of the subject plant species is necessary to indicate that a match between the questioned sample and the reference material is not random and does not result in false positives. When the database is of an insufficient size, a match cannot be made with statistical certainty surrounding the taxonomic identification. In regard to questioned plant samples involved in an investigation, the local agriculture agency may have samples, or even an herbarium/arboretum, from which vouchers of reference plant materials could be obtained for comparisons. Further, a local botanist or plant society may provide valuable metadata, such as information on the habitat and dispersal of plant species in the area. To address the need for information on the population of plant species, a central repository containing metadata would be useful. Such an effort has been started for plants that are involved in timber investigations. The Global Timber Tracking Network was formed to collect information pertinent to timber tracking in real time (https://globaltimbertrackingnetwork.org). Initiatives like this would be beneficial for other plants of forensic interest, like grasses and pollen. In addition, the European Commission has collated multiple metadatabases to provide information on instances of food fraud and authentication issues that could aid forensic investigations (https://knowledge4policy.ec.europa.eu/publication/food-fraud-data-bases_en).

It is apparent that DNA-based analysis of plant material associated with forensic evidence could be pivotal in forensic investigations. Current analysis techniques, such as DNA barcoding and metabarcoding, have been widely evaluated by the broader scientific community and thus provide a valuable framework for forensic science. While DNA analysis of plant evidence is not widely implemented by forensic laboratories at this time, principles discussed in this chapter could serve as a guide to develop operational capabilities in forensic botany.

Disclaimer This is publication no. 21-09 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for information only, and inclusion does not imply endorsement by the FBI or the US Government. The views expressed are those of the authors and do not necessarily reflect the official policy or position of the FBI or the US Government.

Natalie Damaso is currently an MIT Lincoln Laboratory employee. No Laboratory funding or resources were used to produce the result/findings reported in this publication.

References

Avula B, Sagi S, Gafner S, Upton R, Wang Y-H, Wang M, Khan I (2016) *Ginkgo Biloba* supplements adulteration identification using HPTLC and UHPLC-QTOF-MS. Anal Bioanal Chem 407:7733–7746. https://doi.org/10.1007/s00216-015-8938-1

Bell KL, de Vere N, Keller A, Richardson RT, Gous A, Burgess KS, Brosi BJ (2016) Pollen DNA barcoding: current applications and future prospects. Genome 59(9):629–640. https://doi.org/10.1139/gen-2015-0200

Bell KL, Burgess KS, Botsch JC, Dobbs EK, Read TD, Brosi BJ (2019) Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. Mol Ecol 28(2): 431–455. https://doi.org/10.1111/mec.14840

Boggs LM, Scheible MKR, Machado G, Meiklejohn KA (2019) Single fragment or bulk soil DNA metabarcoding: which is better for characterizing biological taxa found in surface soils for sample separation? Genes 10(6):431. https://doi.org/10.3390/genes10060431

Bohmann K, Mirarab S, Bafna V, Gilbert MTP (2020) Beyond DNA barcoding: the unrealized potential of genome skim data in sample identification. Mol Ecol 29(14):2521–2534. https://doi.org/10.1111/mec.15507

Bosmali I, Ganopoulos I, Madesis P, Tsaftaris A (2012) Microsatellite and DNA-barcode regions typing combined with High Resolution Melting (HRM) analysis for food forensic uses: a case study on lentils (*Lens culinaris*). Food Res Int 46(1):141–147. https://doi.org/10.1016/j.foodres.2011.12.013

Braga J, Pastore T, Coradin V, Bergo M, Wiedenhoeft A (2016) NIRS identification of *Swietenia macrophylla* is robust across specimens from 27 countries. IAWA J 37:420–430. https://doi.org/10.1163/22941932-20160144

Bruni I, De Mattia F, Galimberti A, Galasso G, Banfi E, Casiraghi M, Labra M (2010) Identification of poisonous plants by DNA barcoding approach. Int J Legal Med 124(6):595–603. https://doi.org/10.1007/s00414-010-0447-3

Bruno A, Sandionigi A, Agostinetto G, Bernabovi L, Frigerio J, Casiraghi M, Labra M (2019) Food tracking perspective: DNA metabarcoding to identify plant composition in complex and processed food products. Genes 10(3):248. https://doi.org/10.3390/genes10030248

Carita E (2005) Classical and future DNA typing technologies for plants. In: Coyle HM (ed) Forensic botany: principles and applications to criminal casework. CRC Press, Boca Raton, pp 253–274

Carracedo Á, Butler JM, Gusmão L, Linacre A, Parson W, Roewer L, Schneider PM (2013) New guidelines for the publication of genetic population data. Forensic Sci Int Genet 7(2):217–220. https://doi.org/10.1016/j.fsigen.2013.01.001

Carrier C, Cholette F, Quintero C, Fulcher C (2013) Potential use of DNA barcoding for the identification of tobacco seized from waterpipes. Forensic Sci Int Genet 7(1):194–197. https://doi.org/10.1016/j.fsigen.2012.04.005

Christenhusz M, Byng JW (2016) The number of known plants species in the world and its annual increase. Phytotaxa 261:201–217

Coghlan ML, Haile J, Houston J, Murray DC, White NE, Moolhuijzen P, Bellgard MI, Bunce M (2012) Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. PLoS Genet 8(4):e1002657. https://doi.org/10.1371/journal.pgen.1002657

Coissac É, Riaz T, Puillandre N (2012) Bioinformatic challenges for DNA metabarcoding of plants and animals. Mol Ecol 21:1834–1847. https://doi.org/10.1111/j.1365-294X.2012.05550.x

Craft KJ, Owens JD, Ashley MV (2007) Application of plant DNA markers in forensic botany: genetic comparison of *Quercus* evidence leaves to crime scene trees using microsatellites. Forensic Sci Int 165(1):64–70. https://doi.org/10.1016/j.forsciint.2006.03.002

Craine J, Barberán A, Lynch R, Menninger H, Dunn R, Fierer N (2017) Molecular analysis of environmental plant DNA in house dust across the United States. Aerobiologia 33(1):71–86. https://doi.org/10.1007/s10453-016-9451-5

Cui X, Li W, Wei J, Qi Y, Li R, Yang Y, Shi Y, Meng X, Mi Y, Huot T, Sun W, Zheng X (2020) Assessing the identity of commercial herbs from a Cambodian market using DNA barcoding. Front Pharmacol 11:244. https://doi.org/10.3389/fphar.2020.00244

de Boer HJ, Ouarghidi A, Martin G, Abbad A, Kool A (2014) DNA barcoding reveals limited accuracy of identifications based on folk taxonomy. PLoS One 9(1):e84291. https://doi.org/10.1371/journal.pone.0084291

Diehn S, Zimmermann B, Tafintseva V, Bağcıoğlu M, Kohler A, Ohlson M, Fjellheim S, Kneipp J (2020) Discrimination of grass pollen of different species by FTIR spectroscopy of individual pollen grains. Anal Bioanal Chem 412(24):6459–6474. https://doi.org/10.1007/s00216-020-02628-2

Dormontt EE, Boner M, Braun B, Breulmann G, Degen B, Espinoza E, Gardner S, Guillery P, Hermanson JC, Koch G, Lee SL, Kanashiro M, Rimbawanto A, Thomas D, Wiedenhoeft AC, Yin Y, Zahnen J, Lowe AJ (2015) Forensic timber identification: it's time to integrate disciplines to combat illegal logging. Biol Conserv 191:790–798. https://doi.org/10.1016/j.biocon.2015.06.038

Dormontt EE, Jardine DI, van Dijk K-J, Dunker BF, Dixon RRM, Hipkins VD, Tobe S, Linacre A, Lowe AJ (2020) Forensic validation of a SNP and INDEL panel for individualisation of timber from bigleaf maple (*Acer macrophyllum* Pursch). Forensic Sci Int Genet 46:102252. https://doi.org/10.1016/j.fsigen.2020.102252

Finch KN, Cronn RC, Ayala Richter MC, Blanc-Jolivet C, Correa Guerrero MC, De Stefano Beltrán L, García-Dávila CR, Honorio Coronado EN, Palacios-Ramos S, Paredes-Villanueva K, Jones FA (2020) Predicting the geographic origin of Spanish Cedar (*Cedrela odorata L.*) based on DNA variation. Conserv Genet 21(4):625–639. https://doi.org/10.1007/s10592-020-01282-6

Fløjgaard C, Frøslev TG, Brunbjerg AK, Bruun HH, Moeslund J, Hansen AJ, Ejrnæs R (2019) Predicting provenance of forensic soil samples: linking soil to ecological habitats by metabarcoding and supervised classification. PLoS One 14(7):e0202844. https://doi.org/10.1371/journal.pone.0202844

Funk VA, Gostel M, Devine A, Kelloff CL, Wurdack K, Tuccinardi C, Radosavljevic A, Peters M, Coddington J (2017) Guidelines for collecting vouchers and tissues intended for genomic work (Smithsonian Institution): botany best practices. Biodiversity Data J 5:e11625. https://doi.org/10.3897/BDJ.5.e11625

Giampaoli S, Berti A, Di Maggio RM, Pilli E, Valentini A, Valeriani F, Gianfranceschi G, Barni F, Ripani L, Romano Spica V (2014) The environmental biological signature: NGS profiling for forensic comparison of soils. Forensic Sci Int 240:41–47. https://doi.org/10.1016/j.forsciint.2014.02.028

Hall D, Byrd J (2012) Forensic botany: a practical guide. Wiley-Blackwell, Chichester

Hardy CR, Martin DS (2012) Plants and wildlife forensics. In: Huffman JE, Wallace JR (eds) Wildlife forensics: methods and applications. Wiley, London, UK

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim K-J, Kress WJ, Schneider H, van AlphenStahl J, Barrett SCH, van den Berg C, Bogarin D, Burgess KS, Cameron KM, Carine M, Chacón J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TAJ, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim Y-D, Lahaye R, Lee H-L, Long DG, Madriñán S, Maurin O, Meusnier I, Newmaster SG, Park C-W, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi D-K, Little DP (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106(31):12794. https://doi.org/10.1073/pnas.0905845106

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS One 6(5):e19254. https://doi.org/10.1371/journal.pone.0019254

Houston R, Mayes C, King JL, Hughes-Stamm S, Gangitano D (2018) Massively parallel sequencing of 12 autosomal STRs in *Cannabis sativa*. Electrophoresis 39(22):2906–2911. https://doi.org/10.1002/elps.201800152

Itamiya H, Kumisaka K, Kikkawa H, Sugita R (2020) Study on recovery of plant fragments from adhesive sheet and tape in forensic examination. Jpn J Forensic Sci Technol. https://doi.org/10.3408/jafst.777

Jagielski T, Gawor J, Bakuła Z, Zuchniewicz K, Żak I, Gromadka R (2017) An optimized method for high quality DNA extraction from microalga *Prototheca wickerhamii* for genome sequencing. Plant Methods 13:77–77. https://doi.org/10.1186/s13007-017-0228-9

Jhang T, Shasany AK (2012) Random amplified marker technique for plants rich in polyphenols. Methods Mol Biol 862:61–74. https://doi.org/10.1007/978-1-61779-609-8_6

Jolivet C, Degen B (2012) Use of DNA fingerprints to control the origin of sapelli timber (*Entandrophragma cylindricum*) at the forest concession level in Cameroon. Forensic Sci Int Genet 6(4):487–493. https://doi.org/10.1016/j.fsigen.2011.11.002

Kelley L, Rose E, McCullough B, Martinez M, Baudelet M (2020) Non-destructive DNA analysis of single pollen grains. Forensic Chem 20:100275. https://doi.org/10.1016/j.forc.2020.100275

Koopman WJM, Kuiper I, Klein-Geltink DJA, Sabatino GJH, Smulders MJM (2012) Botanical DNA evidence in criminal cases: knotgrass (*Polygonum aviculare L.*) as a model species. Forensic Sci Int Genet 6(3):366–374. https://doi.org/10.1016/j.fsigen.2011.07.013

Korpelainen H, Virtanen V (2003) DNA fingerprinting of mosses. J Forensic Sci 48(4):804–807. https://doi.org/10.1520/JFS2002435

Kress WJ (2017) Plant DNA barcodes: applications today and in the future. J Syst Evol 55(4):291–307. https://doi.org/10.1111/jse.12254

Lancaster C, Espinoza E (2012) Analysis of select Dalbergia and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. Rapid Commun Mass Spectrom 26(9):1147–1156. https://doi.org/10.1002/rcm.6215

Lee C-L, Huang Y-H, Hsu IC, Lee HC (2020) Evaluation of plant seed DNA and botanical evidence for potential forensic applications. Forensic Sci Res 5(1):55–63. https://doi.org/10.1080/20961790.2019.1594599

Lesiak AD, Cody RB, Dane AJ, Musah RA (2015) Plant seed species identification from chemical fingerprints: a high-throughput application of direct analysis in real time mass spectrometry. Anal Chem 87(17):8748–8757. https://doi.org/10.1021/acs.analchem.5b01611

Marciano MA, Panicker SX, Liddil GD, Lindgren D, Sweder KS (2018) Development of a method to extract opium poppy (*Papaver somniferum L.*) DNA from heroin. Sci Rep 8(1):2590. https://doi.org/10.1038/s41598-018-20996-9

Meiklejohn KA, Jackson ML, Stern LA, Robertson JM (2018) A protocol for obtaining DNA barcodes from plant and insect fragments isolated from forensic-type soils. Int J Legal Med 132(6):1515–1526. https://doi.org/10.1007/s00414-018-1772-1

Meiklejohn KA, Damaso N, Robertson JM (2019) Assessment of BOLD and GenBank – their accuracy and reliability for the identification of biological materials. PLoS One 14(6):e0217084. https://doi.org/10.1371/journal.pone.0217084

Minich JJ, Zhu Q, Janssen S, Hendrickson R, Amir A, Vetter R, Hyde J, Doty MM, Stillwell K, Benardini J, Kim JH, Allen EE, Venkateswaran K, Knight R (2018) KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. mSystems 3(3):e00218–e00217. https://doi.org/10.1128/mSystems.00218-17

Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, Sundaresan V (2016) DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnol J 14(1):8–21. https://doi.org/10.1111/pbi.12419

Moore MK, Frazier K (2019) Humans are animals, too: critical commonalities and differences between human and wildlife forensic genetics. J Forensic Sci 64(6):1603–1621. https://doi.org/10.1111/1556-4029.14066

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8(19):4321–4325. https://doi.org/10.1093/nar/8.19.4321

Musah RA, Espinoza EO, Cody RB, Lesiak AD, Christensen ED, Moore HE, Maleknia S, Drijfhout FP (2015) A high throughput ambient mass spectrometric approach to species identification and classification from chemical fingerprint signatures. Sci Rep 5(1):11520. https://doi.org/10.1038/srep11520

Ng KKS, Lee SL, Tnah LH, Nurul-Farhanah Z, Ng CH, Lee CT, Tani N, Diway B, Lai PS, Khoo E (2016) Forensic timber identification: a case study of a CITES listed species, *Gonystylus bancanus* (Thymelaeaceae). Forensic Sci Int Genet 23:197–209. https://doi.org/10.1016/j.fsigen.2016.05.002

Ng CH, Ng KKS, Lee SL, Tnah LH, Lee CT, Zakaria N-F (2020) A geographical traceability system for Merbau (*Intsia palembanica Miq.*), an important timber species from peninsular Malaysia. Forensic Sci Int Genet 44:102188. https://doi.org/10.1016/j.fsigen.2019.102188

Nickrent DL (1994) From field to film: rapid sequencing methods for field-collected plant species. BioTechniques 16(3):470–475

Osathanunkul M, Madesis P (2019) Bar-HRM: a reliable and fast method for species identification of ginseng (*Panax ginseng, Panax notoginseng, Talinum paniculatum* and *Phytolacca Americana*). PeerJ 7:e7660. https://doi.org/10.7717/peerj.7660

Paranaiba RTF, Carvalho CBV, Freitas JM, Fassio LH, Botelho ÉD, Neves DBJ, Silva RCJ, Aguiar SM (2019) Forensic botany and forensic chemistry working together: application of plant DNA barcoding as a complement to forensic chemistry-a case study in Brazil. Genome 62(1):11–18. https://doi.org/10.1139/gen-2018-0066

Paranaiba RTF, Carvalho CBV, Paiva RS, Trindade BR, Barros MG, Souza EP, Gontijo AB, Silveira D (2020) DNA from wood – a simple approach facing a challenging matrix – a preliminary study. Forensic Sci Int 314:110371. https://doi.org/10.1016/j.forsciint.2020.110371

Pastore T, Braga J, Coradin V, Magalhães W, Okino E, Camargos J, Muniz G, Bressan O, Davrieux F (2011) Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: discrimination of true mahogany, cedar, andiroba, and curupixá. Holzforschung 65. https://doi.org/10.1515/hf.2011.010

Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R (2009) DNA from processed and unprocessed wood: factors influencing the isolation success. Forensic Sci Int Genet 3(3):185–192. https://doi.org/10.1016/j.fsigen.2009.01.002

Roman MG, Houston R (2020) Investigation of chloroplast regions rps16 and clpP for determination of *Cannabis sativa* crop type and biogeographical origin. Legal Med 47:101759. https://doi.org/10.1016/j.legalmed.2020.101759

Schield C, Campelli C, Sycalik J, Randle C, Hughes-Stamm S, Gangitano D (2016) Identification and persistence of *Pinus* pollen DNA on cotton fabrics: a forensic application. Sci Justice 56(1):29–34. https://doi.org/10.1016/j.scijus.2015.11.005

Schulte F, Lingott J, Panne U, Kneipp J (2008) Chemical characterization and classification of pollen. Anal Chem 80:9551–9556. https://doi.org/10.1021/ac801791a

Shivanna K, Tandon R (2014) Reproductive ecology of flowering plants: a manual. Springer, New Delhi

Solano J, Rodríguez L, Encina F (2016) Identification case of evidence in timber tracing of *Pinus radiate*, using High-resolution melting (HRM) analysis. Forensic Sci Int Genet 21:e6-e9. https://doi.org/10.1016/j.fsigen.2015.09.003

Solano J, Anabalón L, Figueroa A, Gangitano D (2020) ITS barcoding using high resolution melting analysis of *Cannabis sativa* drug seizures in Chile: a forensic application. Forensic Sci Int 316:110550. https://doi.org/10.1016/j.forsciint.2020.110550

St. John L, Ogle DG, Scianna J, Winslow S, Tilley D, Hoag C (2010) Plant materials technical note no. 1. Plant materials collection guide. USDA-NCRS, Boise

Staats M, Arulandhu AJ, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T, Prins TW, Kok E (2016) Advances in DNA metabarcoding for food and wildlife forensic species identification. Anal Bioanal Chem 408(17):4615–4630. https://doi.org/10.1007/s00216-016-9595-8

Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP (2011) Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep 1(1):42. https://doi.org/10.1038/srep00042

Sucher NJ, Hennell JR, Carles MC (eds) (2012) Plant DNA fingerprinting and barcoding: methods and protocols. Meth Mol Biol. 862. Springer, New York. https://doi.org/10.1007/978-1-61779-609-8

Timpano EK, Scheible MKR, Meiklejohn KA (2020) Optimization of the second internal transcribed spacer (ITS2) for characterizing land plants from soil. PLoS One 15(4):e0231436. https://doi.org/10.1371/journal.pone.0231436

Tnah L, Lee S, Tan A, Lee CT, Ng K, Hong N, Ling S, Zakaria N (2015) DNA barcoding of 100 common medicinal plants in Malaysia for species authentication. Genome 58:229–229

Um J-Y, Chung H-S, Kim M-S, Na H-J, Kwon H-J, Kim J-J, Lee K-M, Lee SJ, Lim JP, Do K-R, Hwang W-J, Lyu Y-S, An N-H, Kim H-M (2001) Molecular authentication of *Panax ginseng* species by RAPD analysis and PCR-RFLP. Biol Pharm Bull 24(8):872–875. https://doi.org/10.1248/bpb.24.872

Uncu AO, Uncu AT (2020) A barcode-DNA analysis method for the identification of plant oil adulteration in milk and dairy products. Food Chem 326:126986. https://doi.org/10.1016/j.foodchem.2020.126986

Vaclavik L, Cajka T, Hrbek V, Hajslova J (2009) Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. Anal Chim Acta 645(1):56–63. https://doi.org/10.1016/j.aca.2009.04.043

Vašek J, Čílová D, Melounová M, Svoboda P, Vejl P, Štikarová R, Vostrý L, Kuchtová P, Ovesná J (2020) New EST-SSR markers for individual genotyping of opium poppy cultivars (*Papaver somniferum L.*). Plants 9(1):10. https://doi.org/10.3390/plants9010010

Vlam M, de Groot GA, Boom A, Copini P, Laros I, Veldhuijzen K, Zakamdi D, Zuidema PA (2018) Developing forensic tools for an African timber: regional origin is revealed by genetic characteristics, but not by isotopic signature. Biol Conserv 220:262–271. https://doi.org/10.1016/j.biocon.2018.01.031

Ward J, Gilmore S, Robertson J, Peakall R (2009) A grass molecular identification system for forensic botany: a critical evaluation of the strengths and limitations. J Forensic Sci 54:1254–1260. https://doi.org/10.1111/j.1556-4029.2009.01196.x

Wilks B, Morgan RM, Rose NL (2017) An experimental study addressing the use of geoforensic analysis for the exploitation of improvised explosive devices (IEDs). Forensic Sci Int 278:52–67. https://doi.org/10.1016/j.forsciint.2017.06.028

Wiltshire PE (2016) Protocols for forensic palynology. Palynology 40(1):4–24. https://doi.org/10.1080/01916122.2015.1091138

Wright SI, Kalisz S, Slotte T (2013) Evolutionary consequences of self-fertilization in plants. Proc R Soc B Biol Sci 280(1760):20130133. https://doi.org/10.1098/rspb.2013.0133

Yoccoz N, Bråthen K, Gielly L, Haile J, Edwards M, Goslar T, Stedingk H, Brysting A, Coissac É, Pompanon F, Sønstebø JH, Miquel C, Valentini A, Bello F, Chave J, Thuiller W, Wincker P, Cruaud C, Gavory F, Taberlet P (2012) DNA from soil mirrors plant taxonomic and growth form diversity. Mol Ecol 21:3647–3655. https://doi.org/10.1111/j.1365-294X.2012.05545.x

Young JM, Austin JJ, Weyrich LS (2017) Soil DNA metabarcoding and high-throughput sequencing as a forensic tool: considerations, potential limitations and recommendations. FEMS Microbiol Ecol 93(2). https://doi.org/10.1093/femsec/fiw207

Young B, Roman MG, LaRue B, Gangitano D, Houston R (2020) Evaluation of 19 short tandem repeat markers for individualization of *Papaver somniferum*. Sci Justice 60(3):253–262. https://doi.org/10.1016/j.scijus.2019.12.002

Molecular Markers and Genomics for Food and Beverages Characterization

40

Rita Vignani, Monica Scali, and Pietro Liò

Contents

Abstract

Abstract We summarize and discuss latest trends and works on methods and most significant case studies concerning molecular-based food and beverages quality and authenticity assessment. DNA-based technology may work as a wild card between food authentication, gut microbiome profiling, and the evaluation of food-borne microbiological risk. The central axiom of food security necessarily passes through

R. Vignani (✉)
Department of Life Science, University of Siena, SIENA, Italy

Serge-genomics, SIENA, Italy
e-mail: rita.vignani@unisi.it

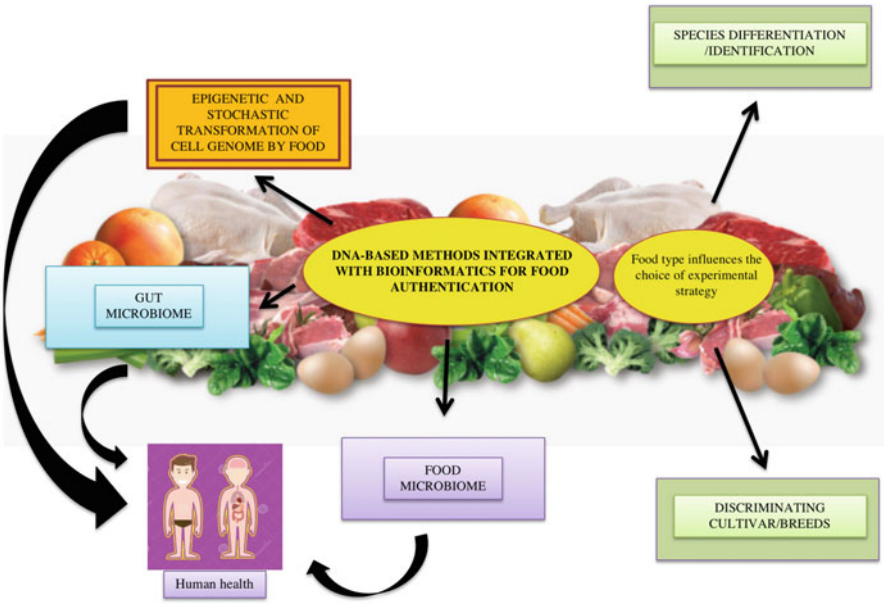
M. Scali
Department of Life Science, University of Siena, SIENA, Italy
e-mail: monica.scali@unisi.it

P. Liò
Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
e-mail: pl219@cam.ac.uk

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_43

the validation of the concept of food authenticity that in turn depends on the analytical assessment of ingredients. With increasing molecular detail, literature demonstrates how the food intake is associated with specific epigenetic effects on genomes leading to gene expression regulation. Similarly, food intake deeply affects human health by influencing secondary metabolite production regulated by the gut microbiome. In the last decade, biology moved rapidly from small numbers to big data, keeping the DNA as a major tagged molecule for food characterization. Simple molecular markers have the advantage to be used by small-medium laboratories and may be effective in addressing simple questions, while the integrated genomics approaches address exhaustive description of food and beverages quality traits. More significant case studies of molecular characterization of plant and animal-derived food and grapevine-derived beverages are also reported. Finally we conclude on how the interdisciplinary and multidisciplinary approaches could improve our relation with food and beverages from the point of view of health and quality of life and against frauds.

Visual Abstract DNA-based methods integrated with bioinformatics for genetic identification may act as a wild card under several molecular approaches that intercept human health issues. Food has a deep impact on human health on a direct basis through its nutritional properties, its sanitary status, affecting the gut microbiome and the cell epigenome, and likely, under certain circumstances, a stochastic mechanism even by transforming the cell genome itself. DNA-based authentication can be effective in profiling food microbiome related to the potential risk of causing food-borne diseases and food intrinsic qualities related to raw ingredients and origin. The DNA test aimed to address food authentication should be planned according to the specific goal and type of food or beverage, where species identification or intraspecific identification are selectively required.



Keywords

Food DNA analysis · Bioinformatics · Molecular authentication · Product traceability

Introduction

Food security is a topic of increasing concern in the food industry and is deeply interlinked to food labeling regulations. Fraudulent activities are characterized by their intentional nature, including the aim to make an economic gain, in violation of legal rules and at the expense of the immediate customer or the final customer, which are potentially harmful to human, animal, or plant health and environment. Since the creation of the Administrative Assistance and Cooperation System for Food Fraud (AAC-FF system) in 2016, European Member States requests for cooperation concerning suspected cases of fraud in the agri-food chain have significantly risen. Between 2016 and 2019, a total of 861 notifications were sent through the system, according to the 2019 report by the AAC-FF organization, showing an 85% increase (The EU Food Fraud Network and the Administrative Assistance and Cooperation System, Annual Report 2019). Recognizing and investigating fraudulent activities in the food sector remains a challenging task, due to the complexity of an internationalized food trade and to the inventiveness of those who proceed in acting frauds with a myriad of methodologies aimed at obtaining an economic gain. Having confidence in the authenticity of food ingredients is essential to protect brands against food fraud. In 2018 according to the annual report by the European Union, fraud costs the food industry around 30 billion euros every year (The EU Food Fraud Network and the System for Administrative Assistance – Food Fraud, Annual Report 2018). Since the complexity of the food supply chain is challenging the abilities of analytical tools used for traceability of ingredients, various emerging molecular techniques are currently used for food and beverages authentication and assessment of compliance of composition to the regulatory framework and to the ingredients declared in labels. In a market characterized by growing competitiveness and ever-expanding borders, it is necessary to inspire a positive interest in healthy food and beverages consumption. Consumers are increasingly attracted and influenced by a number of attributes that affect product value, such as guarantees declared on labels like DOP (Protected denomination of origin), IGP (Protected Geographical Indication), DOC (Denominazione di Origine Controllata), and DOCG (Denominazione di Origine Controllata e Garantita) for Italian wines and AOC (Appelation d'Origine Controlee) for the French wines and organic certifications, product name, genuineness, and transparency. Despite the prevalence of food fraud and its sometimes devastating consequences for consumer well-being and the sectors involved, a systematic economic analysis of food fraud is virtually absent. Previous research on food fraud has focused mainly on product mislabeling and its impact on consumers (Meerza et al. 2019).

The higher the economic value of the product on the market, the greater the risk that criminal organizations may apply fraudulent practices on them to maximize

illicit profits. Wines, olive oil, and a few plant condiments such as saffron and truffle are among the most counterfeited products in the market (White 2017). Regarding the welfare disaggregation effects of food fraud, it has been shown within a theoretical framework that comparing the market effects of mislabeling and food adulteration, there would be a tendency by producers to mislabeling rather than adulterating their products (Danezis et al. 2016). The phenomenon of counterfeiting and adulteration are dangerous to both producer reputation and human health. Technological analysis is a tool that allows producers to protect their products and strengthen their relationship with consumers. For this reason, anti-counterfeiting technologies are increasing as effective and incontrovertible strategic tools for protecting local and niche food lovers. However, in a globalized market, analytical traceability is still only voluntarily adopted. Controls are not regularly undertaken on imported food products or on national products. Therefore, the systemic introduction of analytical methods for assessing food integrity and origin genuineness before market delivery seems essential. It is likely that governmental organizations' deficient control systems will lead to widespread frauds on the worldwide food market. Identifying the ingredients in a food product helps to understand if this was produced according to local and national regulations. In other words, food and beverages authentication is an analytical methodology that validates the specification on labels regarding food ingredients, origin, and production process. Depending on the specific purpose, an authentication plan may require different levels of sensitivity, where different analytical methods are integrated, from visual and morphological inspection to the metabolomic profile, to characterization through elementary isotopic ratios up to profiling based on the study of DNA to reconstruct genetic identity of the main ingredients. Under this point of view, it should be pointed out how the DNA analysis may be suitable to target species, e.g., DNA barcoding (Villa et al. 2016) or NGS (Haiminen et al. 2019) as well as intraspecific variability, cultivar, or varieties identification, e.g., DNA molecular markers, SNPs, SSRs, etc. (Pereira et al. 2018).

The identification of the varieties or animal breeds is of great importance from the point of marketing of the product as the regulations for the special classification of the products specified above (DOP, DOC, etc.) regulate the type of varietal ingredients and sometimes also the relative proportion in the final product. This is the case of many wines of qualitative excellence where the blend is established by the production regulations by allowing specific percentages of each grape variety in the final blend, requiring the application of molecular approaches such as the real-time PCR (RT-PCR). Thus, it seems important to calibrate the type of analysis according to the principle of addressing with the minimal analytical effort the higher probability of getting adequate responses to a major question. In the following sections, several case studies, including wine and Aceto Balsamico di Modena, will be described pointing out the analytical approaches that in recent times were addressed at obtaining the identity profile of each product.

Epigenetic Influence of Food

By late antiquity and medieval times, the dominant approach became Aristotle's teleology, whereby human nature was believed to exist somehow independently of individuals causing humans to simply become what they become. This, in turn, seemed to be the epiphany of a special connection between human nature and divinity. Accordingly, nature itself has intentions and goals, including the goal for humanity to live naturally. However, the existence of this invariable and metaphysical human nature is subject of much historical debate, continuing into modern times and influencing a sort of static nature of human beings bringing in the essential concept of man as a central leader in nature. As a consequence, medicine and environmental science were organized for centuries under this theoretical dogmatic assumption. In the post-genomic era, the philosophical idea of human health changed in that man is no longer the standing-alone protagonist, but rather human health is clearly dependent on complex interactions between his physiology and external factors. Food intake has remarkable effects on our health by regulating the energetic balance of our bodies and due to increasing evidence, because it has a significant impact on our gut microbiota and on our epigenome (Wu et al. 2019). Effects on microbiota were investigated recently by evaluating stool composition in rats fed for 6 months with up to 33% content of genetically modified food (GM-food). The authors demonstrated that bacterial community richness was not altered by the test diets, but the fecal microbes were associated with host systemic metabolic changes found in plasma in a gender-specific manner (Mesnage et al. 2019).

Zhang and Kutateladze (2018) reported finer impacts on molecular equilibrium of living cells induced by food-derived metabolites by which foreign DNA entering our digestive apparatus can stochastically lead to genome-wide alterations of transcriptional and CpG DNA methylation status. Furthermore, a new concept for general pathogenesis has been proposed by Doerfler (2020) by which DNA ingested with food would persist transiently in the gastrointestinal tract in mice and can be detected in somatic cells, namely, white blood and spleen cells. This innovative concept deserves rigorous pursuit especially if referred to the consumption of industrial food that may contain percentages of plant GMO derivatives. Remarkable breakthroughs in food consumption habits established how a balanced diet enhances life expectancy and improves health quality. However, the molecular mechanisms underlying these positive effects are only partially understood, and they seem to be linked to dietary factors such as secondary metabolites that may regulate epigenetic marks in the genome, modulating gene expression. For instance, specific grapevine (*Vitis vinifera* L.) phytochemicals, such as dihydrocaffeic acid (DHCA) and malvidin-3'-O-glucoside (Malgluc), are able in mice to attenuate depression-like behaviors by modulating epigenetic marks in the genome, influencing inflammation and brain synaptic plasticity promoting resilience against stress (Wang et al. 2018). Many other factors driven by food-derived products were demonstrated, including

hijacking the conventional epigenetic marking pathways on histones operated by organic derivatives of sodium benzoate, a preservative approved by FDA that is widely used by food industry that promoting histone lysine benzoylation leads to a modification in gene expression (Huang et al. 2018). Long-term effects deriving from food consumption habits in early life, such as breast-feeding, likely induce an epigenetic memory resisting till adulthood. Namely, milk fats are known to activate the nuclear receptor peroxisome proliferator-activated receptor (PPAR alpha) which is an important transcriptional factor regulating the liver metabolism. In mice the transcriptional activation of PPAR alpha target genes such as the fibroblast growth factor-21 (Fgf21) induced by demethylation would remain as a epigenetic mark later on in life preventing weight gain and obesity (Yuan et al. 2018). Another example of effects of food on epigenetic balance of host cells is represented by the association of consumption of highly processed, carbohydrate, and fat-rich food characterizing the so-called western diet (Pollan 2008). The western diet limits the production of short fatty acids (SFCAs) by the gut microbiome increasing the polyacetylation of histones, thus inducing modifications in the host chromatin state (Krautkramer et al. 2016). Additional studies concerning the effects of diet on epigenome were recently summarized by Miceli et al. (2014) and Zhang and Kutateladze (2018). The topic may be seen as a new stimulating border where the new frontiers of molecular biology build up the scientific basis of a centenarian, traditional knowledge that is culturally inherited from generation to generation, concerning the importance of a balanced and naturally regulated food intake for the maintenance of health.

Microbiome in Food

Preventing microbial food spoilage is a major concern for health authorities, regulators, consumers, and the food industry. The contamination of food is sometimes difficult to prevent, especially taking into account that a large part of food supplies will be likely produced in the next 30 years largely in developing countries to face demographic increment and subsequently marketed worldwide. The contamination of food products along production chains and subsequent phases of post-harvesting of plant and animal species for marketing purposes is difficult to control because there are several potential sources during production, processing, storage, distribution, and consumption, where microorganisms come in contact with the product. Next-generation sequencing (NGS) techniques including GS high-throughput full-length 16S rRNA gene sequencing provide insights into bacterial community structure possibly guiding targeted prophylactic and disinfection policy (Zwirzitz et al. 2020). With the wide spreading of alternative vegetarian and vegan diets, the markets have increased and differentiated offers of several types of food supplements, claimed to contain probiotics, additives of nutraceutical properties highlighting the need to develop methods to assess their safety and authenticity. Since 2015 the Members of the Sequence the Food Supply Chain Consortium (IBM Research and Mars, USA) are working to characterize the microbiome at a baseline and after food processing. To achieve this, the group is creating reference databases through

metagenomics and computational, analytical workbenches capable of verifying the effectiveness and security of food along the entire productive chains by use of RNA and DNA high-throughput sequencing technologies of microbial communities in food (Weimer et al. 2016).

Tailoring DNA-Based Methods for Food Characterization

The adoption of tailored techniques for food or beverages matrices characterization is regulated by several key factors including the complexity of food in terms of type of processing of ingredients in food or beverage aging and number of biological sources contributing to food. Each factor among those mentioned has a great influence on what can be defined as the universal limiting feature of any molecular-based analysis aimed at identifying the plant or animal origin of food: the chance to optimize a suitable DNA purification protocol. Suitability essentially must comply with the chance to use genotyping or sequencing methodologies depending on the specific level of discrimination the analysis should achieve. In particular, several types of food or beverage authentication are fulfilled by DNA analysis that will identify species; in other cases the analysis should respond to an intraspecific degree of sensitivity detecting varieties or breeds.

Overview of Molecular Methods for Food Authentication

Food security cannot be postulated if the identity of ingredients in food cannot be guaranteed. Many questions concerning food and beverages traceability remain an open question, despite the flourishing of multiple analytical techniques. During the early 1970s, attempts to track and give food an identity card came from the use of chemistry and biochemistry; after the molecular biology boosting, many analytical DNA-based techniques were set for the intraspecific analysis of genetic variability of populations. The DNA fingerprinting techniques were the first step to build up reference databanks as open sources for downstream applications and study of the genetic variability in animal and plant populations of agro-economical interest, although the transfer of these techniques to the identification of food or beverages ingredients is limited by the bottleneck of DNA extraction out of complex matrices. Molecular authentication efficiency maximizes probabilities of success when applied to natural food and beverages characterized by short-chain productions, and it diminishes in industrial food due to the progressive degradation of DNA from original ingredients caused by aging, a large-scale distribution scheme in the food supply chain, intensive processing of ingredients, creation of DNA admixtures from a relative high number of biological sources, additives addition, and chemical modifications of DNA due to the presence of preservative, coloring agents, or other additives. For these reasons, the molecular authentication of industrial food or beverages is more difficult to achieve than that of local, natural food distributed along local small-scale productive chains (Fig. 1).

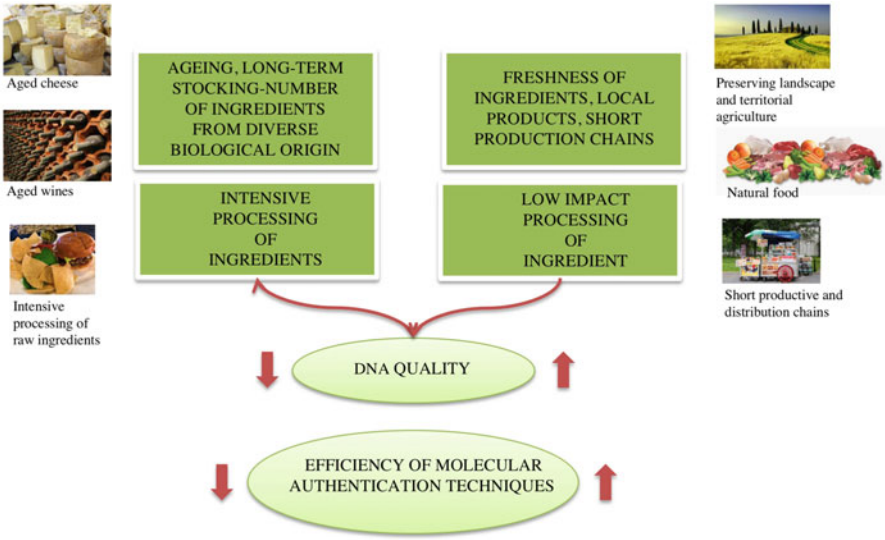


Fig. 1 Molecular authentication efficiency maximizes probabilities of success when applied to natural food and beverages products characterized by short-chain productions. Molecular authentication efficiency diminishes in industrial food due to the progressive degradation of DNA from original ingredients deriving from aging, a large-scale distribution scheme food chain, intensive processing of ingredients, creation of DNA admixtures from a relative high number of biological sources, additives addition, and chemical modifications of DNA due to the presence of preservative, coloring agents, or other additives. For these reasons, the molecular authentication of industrial food or beverages is more difficult to achieve than that of local, natural food distributed along local small-scale productive chains

Consumers and producers need transparency and genuineness in order to be guaranteed the identity of food and beverages in an internationalized market. Traceability is the quality of having an origin or course of development that may be found; referring to food chains, it means a detailed overview of each step of production (origin, logistics, distribution, stocking) giving the power to control the critical steps in order to prevent spoilage, fraud, and risks for human health. It was only after the mid-1980s in concomitance with easiness of many molecular biology tools that the DNA-based analyses were set to characterize food and beverages identity, although the transfer of most of the genotyping methods from plant or animal to food or beverages is depending on DNA quantity and quality that in general in food is a limiting factor.

Due to an increased awareness matured in recent times, consumers are demanding transparency as to where their food comes from and what is in it. In addition, food authenticity represents a prerequisite to address the issue of food integrity. The vast majority of food fraud incidents do not cause a public health threat, but they could. Food fraud is a frequent root cause of food safety incidents and possible threats for human health. Analytical traceability methodologies are necessary especially in processed food where the origin and identity of ingredients were rendered

indistinguishable from original ingredients due to cooking, mixing with multiple ingredients having diverse biological origins. Food and beverages molecular authentication becomes a priority in establishing the food composition and origin (Georgiou and Danezis 2017).

Analytical traceability recent history is additionally used with respect to documental traceability, and it renders controls by organization in charge of actuating control policies feasible. Integration of data deriving from multidisciplinary inputs obtained by molecular, genomics, metabolomics, historical, and cultural fields concerning food and beverages may lead to future portals dedicated to food and beverages ontology (The FoodOn Food Ontology 2020). Bioinformatics is a common, essential tool for the validation of multiple, analytical approaches for authentication of food, wine, and other specialties like the Aceto Balsamico di Modena (Fig. 2). Molecular, chemical, and metabolomic profiling merge into comprehensive food and beverages ontology databases fed by bioinformatics tools, similarly to what already happens in pharmaceuticals (Dimitri and Liò 2017).

Currently there are many different methodologies based on nucleic acids analysis for food and beverages authentication that are set to focus on the biological

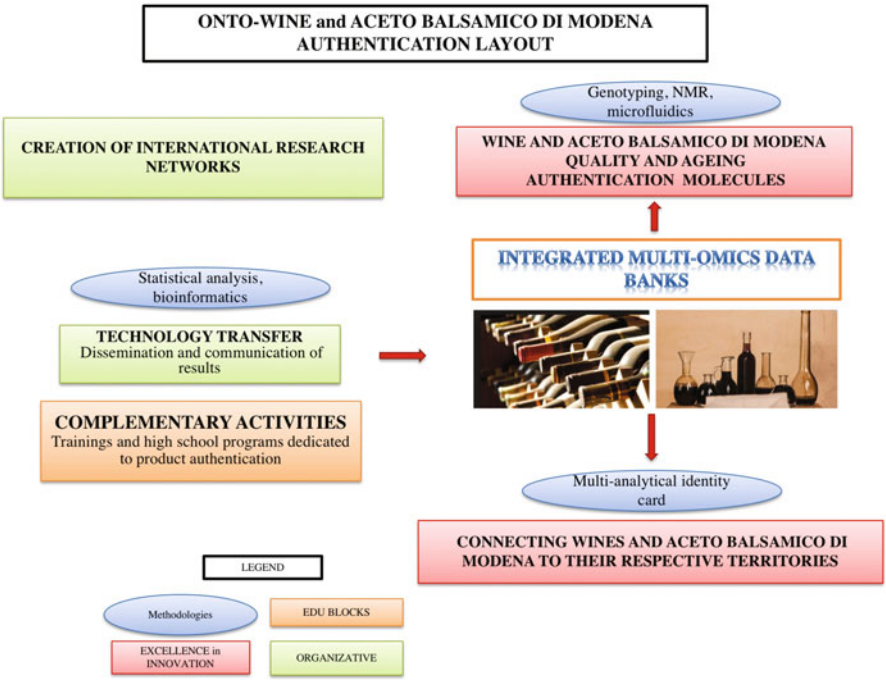


Fig. 2 International research networks for high-quality grapevine – derived products, including DOC and DOCG, AOC wines and Aceto Balsamico di Modena authentication programs integrate multianalytical procedures with the scope of defining products ontology databases. Under such international initiatives, technology transfer actions and educational activities help to connect producers, consumers, possible stakeholders, and research

identification of the main ingredients. DNA analysis is often preferred with respect to RNA analysis (Valdés et al. 2014) thanks to the relative stability of the DNA molecule even after food processing.

Digital polymerase chain reaction (dPCR) (Quan et al. 2018) is a novel method for the absolute quantification of target nucleic acids that took advantage of advances in microfluidics that rendered possible the dispensing of droplets acting singularly as PCR microreactors following a random, Poisson distribution. Differently from real-time PCR, it is an end-point reaction where the absolute quantification relies on the proportion of PCR-positive partitions that suffices to determine the concentration of the target sequence without a need for calibration. dPCR has been successfully used for routine analysis of GMOs on a panel of positive reference material with varying GM content from 0.1% to 10%, although tolerance to inhibitors must be empirically evaluated on a case-by-case basis (Iwobi et al. 2016).

Generally speaking, the molecular techniques lack the ability to address geographical derivation of food and beverages ingredients and their connection with the nutritional properties that in the finished product derive from the secondary metabolic traits of raw ingredients, either from plant or animal origin.

Case Studies

Molecular Authentication Methods for Species Determination in Food

Filled pasta (ravioli, tortellini, lasagna), olive oil, fish and seafood, plant supplements, honey, and wheat are only a few examples of food and beverages where authentication at species level demonstrated to be a useful approach to guarantee the compliance of ingredients to labels. The integration of multiple DNA-based techniques, including the DNA-barcoding, molecular markers, and NGS have been suggested as the most suited for species identification purposes, being already applied for the authentication of several products (Grazina et al. 2020; Galimberti et al. 2013).

The 2013 horse meat scandal was a food industry scandal in parts of Europe in which foods advertised as containing beef were found to contain undeclared horse meat (Castle 2013; O'Mahony 2013). The horse meat scandal in filled pasta has given the regulatory governmental entities the opportunity to apply the molecular authentication methods in extensive monitoring checks. A smaller number of products also contained other undeclared meats, such as pork (Cai et al. 2017). The analysis stated that 23 out of 27 samples of beef burgers contained pig DNA that is a taboo food in the Muslim and Jewish communities (Food Safety Authority of Ireland 2013). While the presence of undeclared meat was not a health issue, the scandal revealed a major breakdown in the traceability of the food supply chain and the risk that harmful ingredients could have been included as well. Namely, the hypothesis that sports horses could have entered the food supply chain represented a potential threat since their meat could contain the drug phenylbutazone which is allowed for

veterinary only, but it is banned in food animals. The scandal later spread to 13 other European countries, and European authorities decided to find an EU-wide solution, and very soon the international network on food authentication promoted and optimized protocols dedicated to the detection of animal species in processed food.

The DNA mini-barcoding is a variation on the theme of the DNA-barcoding that appears most suitable for food authentication where DNA is degraded and present in trace. It mainly consists in a reduction in size of the traditional plastid targets characterizing the DNA barcoding that may easily be detected with high-resolution melting (HRM) analysis as a fast detection protocol (Gao et al. 2019; Villa et al. 2016).

Concerning the quantitative determination of species targets detected by real-time PCR, tests (TaqMan® probes assays) were developed for detection of beef, pork, lamb, chicken, and turkey meat. Assays were developed to detect small amplicons (<150 base pairs) targeting regions in the mitochondrial cytochrome b (cytb) gene (Dooley et al. 2004).

Olive oil, and especially Protected Designation of Origin of extra virgin olive oil (PDO EVO oil) is a high-quality condiment that is thought essential in a balanced Mediterranean style diet, where it brings antioxidants that are important to keep us healthy. The case of olive oil authentication can be faced under two main points of view: the first regards the numerous attempts of frauds that have occurred in recent times; the second concerns the need of picturing the authenticity in terms of botanical origin of *O. europaea* varietal components. The first case is discussed here since it deals with species-specific molecular authentication techniques, from the moment that fraud attempts often consisted in adding low-quality oils deriving from other botanical species (e.g., arachis, sun flower, etc.). The second case will be treated in the following paragraph where intraspecific methods are discussed and regards the availability of methods addressing the genetic profiling of PDO EVO oil to guarantee product traceability and to certify its quality.

In 2017 according to investigators, the illegal activities promoted the import of olive pomace oil that it is an oil extracted from already-pressed fruit pulp using chemical solvents, labeled as EVO oil and exported to the USA. Those products were sold through retail chains in New York, Boston, and Chicago (Smith 2017). In 2019 Europol's intellectual property crime unit seized 150,000 liters of adulterated olive oil heading for German restaurants (Al-Zoubi 2019). A typical case of olive oil fraud is the unlawful addition of oil seeds, and it was faced in the past by chemical profiling and other chemometric tools (Monfreda et al. 2012).

DNA-barcoding techniques applied to olive oil, contaminated with canola and sunflower, analyzed for single-nucleotide polymorphism (SNP) variations in non-coding spacer region between psbAtrnH and partial coding region of matK of plastid genome allowed the detection of oils other than from *O. europea*, with 5% of sensitiveness (Kumar et al. 2011).

The traceability of seafood has largely used the info built up through the consortium for barcoding of life (http://www.boldsystems.org/ or https://ibol.org/) that has a collection of about 8000 different fish species, targeting the mitochondrial gene cytochrome c oxidase I (COI). DNA barcoding represents a widespread tool for fish

and fish products traceability; however major limits in the use of classical barcoding are mainly due to the existence of complex DNA admixtures that unable the recognition of undeclared substitutions especially with local fish breeds and the difficulty in having total DNA of sufficient quantity and quality so that it will produce comprehensive DNA sequences representing even minor breed components. Animal tubulin-based polymorphism (aTBP), an intron length polymorphism method, was recently applied to track identity of *Sparus aurata*, *Dicentrarchus labrax*, *Oncorhynchus mykiss*, *Acipenser naccarii*, *Thunnus thynnus*, *Salmo carpio*, and *Salmo trutta* f. *fario* (Gianì et al. 2020).

Herbal dietary supplements (HDS) are commonly consumed to treat several symptoms including menopausal and are becoming increasingly popular in association with specific dietary habits, such as the vegetarian or vegan style diets. HDS containing black cohosh (*Actaea racemosa*) may accidentally be misidentified or may be deliberately adulterated with low quality-related botanical species. In addition, black cohosh is morphologically related to species that are known to be toxic to humans, and there have been reports of adverse events and toxicities associated with HDS supplements containing black cohosh that deserve deepening of analytical traceability following high standard control policies for guaranteeing safety for human consumption. Among the examples of molecular traceability, a DNA barcoding approach was used (Baker et al. 2012) as well as multiple DNA-based analyses including SCARS, HRM, and NGS (Masada 2016; Grazina et al. 2020).

Honey is an example of food having a mixed biological origin from plants and animal species, the honeybees (*Apis mellifera* L.). Modern techniques enable the results of honey quality testing to be obtained in a short time and move toward the integration of multiple analytical methods from nuclear magnetic resonance (NMR) to molecular methods. The quality of the honey is related with the plant species, and to gain this scope the melissopalynological analyses may be flanked by PCR-based investigation that can address simultaneously the type of honeybee varieties and plant species or variety (Puścion-Jakubik et al. 2020).

Among the most innovative molecular authentication approaches reported in literature, the dPCR is rapidly expanding. Recently, dPCR has been used along the industrial pasta pipeline production chain to detect contaminations of bread wheat (*T. aestivum*) against durum wheat (*T. durum*) at 1,5% sensitiveness. Interestingly the authors address the dPCR to detect a *T. aestivum*-specific locus, the Pinb-D1 (puroindoline b) and the TaHd1 (gene putatively coding for a zinc binding protein), respectively, where the target in the TaHd1 consisted in a conserved within the *Triticum* genus (Morcia et al. 2020).

Molecular Authentication Methods for Variety or Breed Determination in Food, Condiments, and Beverages

Preserving biodiversity of agronomically important species has a fundamental importance for the ability of living organisms to adapt to environmental and climate changes, as well as for the resilience of ecosystems and protection of landscapes.

Biodiversity preservation has reached an increasingly importance in the field of food and beverages authentication. In order to picture the origin of national or regional food, it is necessary to use molecular approaches that are able to predict the varietal or breed intraspecific diversity that rely on robust databanks originating from the detailed prospection of autochthonous genetic resources.

PDO EVO oil requires compliance with precise parameters such as cultivar, geographical origin, agronomic practice, production technology, and organoleptic qualities (Giménez et al. 2010).

Each trait has to be investigated to keep track of PDO EVO oil and to prove its biological origin and intrinsic qualities. Thus, the introduction of legal certifications has required the implementation of molecular traceability procedures. More than a decade ago several studies pointed out the importance of methods based on DNA analysis for olive oil certifications focusing on the combination of multiple techniques such as random amplified polymorphic DNAs (RAPD), interspersed simple sequence repeats (ISSR), and simple sequence repeats (SSRs) (Martins-Lopes et al. 2008).

Among the most widely used markers having a universal field of application, starting from olive oil DNA admixtures, the SSRs analysis allow the detection of multiallelic variants useful to reconstruct the cultivar genotype based on observed allelic sizes and frequencies (Muzzalupo et al. 2015).

More recently, SNPs were also used for PDO EVO oil authentication with respect to the varietal components (Bazakos et al. 2016).

Methods of analysis such as quantitative PCR (qPCR) are also now being exploited with a special emphasis placed on the method of high-resolution melting (HRM), a post-polymerase chain reaction method, which enables rapid, high-performing identification of genetic variants in the DNA regions of interest without sequencing and may distinguish very similar cultivars which differ as little as in only one nucleotide in a specific locus (Batrinou et al. 2020). Together with *O. europea*, and among the woody fruit crop species, the grapevine has certainly a fundamental economic importance for the Mediterranean countries. The wine sector is one of the most strategic agro-food businesses resulting in highly expensive wines that attract fraudulent practices.

Wine is a complex mixture of tens of thousands of chemical compounds. A wine's composition is determined by the compounds present in the fruit, the microbial populations in the juice and wine, the techniques and additions used during winemaking, contributions from the fermentation or aging vessel, and the effects of oxidation and aging.

The existing wine traceability system has some limitations, especially in that the vast majority of controls are documental and the analytical-based certification according to the EU 22005 is foreseen only on a voluntary basis. Grapevine-derived beverages including DOC and DOCG wines and specialty condiments like the unique Aceto Balsamico di Modena are special and important cases not only economically but also from the point of view of frequent attempts of fraud. Recently, the use of a molecular SSR-based fingerprinting of wine in association with bioinformatics, resulting in the term Wine DNA Fingerprinting (WDF), has revealed to be very effective to depict the varietal nature of commercial monovarietal and blended

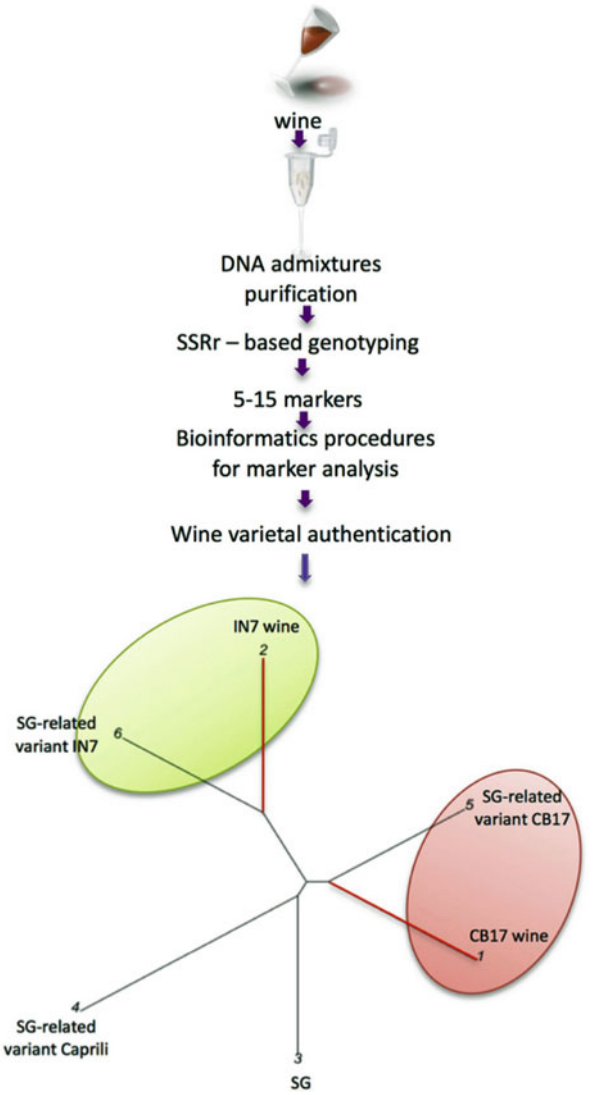


Fig. 3 Wines DNA fingerprints (WDF) technique rely on integration of bioinformatics and molecular data obtained from SSRs that results in a graphical output where the genetic distances among wines and grapevines describe the varietal nature of the wine. In the dendrogram, small-scale fermented experimental Sangiovese wines were analyzed by WDF. The WDF allowed to associate each wine profile with the clonal, local Sangiovese variants CB17, IN7. Wines: 1 = small scale fermented CB17; 2 = small scale fermented IN7. Grapevines: 3 = Standard Sangiovese (SG); 4 = SG-related variant Caprili; 5 = SG-related variant CB17; 6 = SG-related variant IN7. From Vignani et al. (2019).

wines (Fig. 3) (Vignani et al. 2019; Pereira et al. 2018). The WDF arises from the info gathered in grapevine-dedicated databanks, and the method has great similarities with the genotyping protocol operated on plants (Fig. 4). The WDF method was also used to authenticate musts contributing to the Aceto Balsamico di Modena leading to precise quality and compliance assessments based on the probability of presence or absence of specific table grape varieties (Vignani unpublished work). Similarly to the case of PDO EVO oil, the SNP markers are widely used for varietal characterization of wines. After functional characterization of eleven SNPs in two

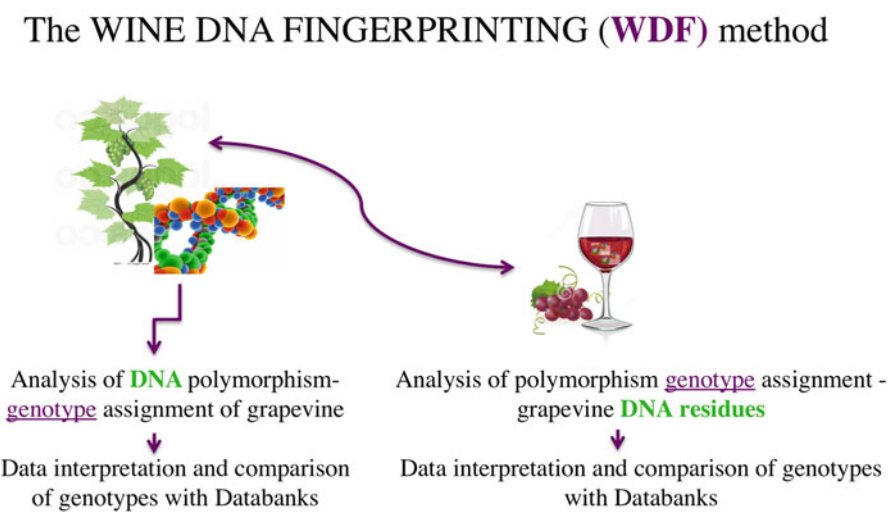


Fig. 4 The Wine DNA fingerprinting (WDF) is based on SSR-based genotyping of wine DNA and plant references. The genotyping principle from the plants to the wine has strong similarities and refers to the alleles or variant forms of a gene that are carried by an organism. Grapevines are diploid organisms, which means that they have two alleles at each genetic position, or locus, with one allele inherited from each parent. Thus, grapevines genotypes generally carry two alleles, excluding cases of localized chimeric tissues. This renders quite immediate the interpretation of the monovarietal nature of a wine through genetic profiling, while the presence of a blend requires more articulated interpretation of wine and plant profiles. The bioinformatics tools developed for the WDF eases the assignment of genetic distances between plant and wine profiles, resulting in a final probability estimate of presence/absence of specific grapevine types in a blend

genes, flavanone 3-hydroxylase (*F3H*, EC: 1.14.11.9) and leucoanthocyanidin dioxygenase (*LDOX*, EC 1.14.11.19; synonym anthocyanidin synthase, *ANS*) active in the anthocyanin pathway, an HRM assay fulfilled the objective of differentiating five genotypes among 20 different cultivars. Three probes, with different lengths and sequences, were used as bio-recognition elements in an optical biosensor platform based on a long period grating (LPG) fiber-optic sensor (Gomes et al. 2018). The use of biosensors that link a biological trait such as a DNA mutation to a physical principle of detection constitutes the basis for a fast, high-throughput traceability system able to avoid the nucleic acid amplification step by PCR. Recent work uses a label-free platform detecting SNPs characterizing a few grapevine varieties and represents a potential powerful innovative authentication method for wines (Barrias et al. 2019). The whole-genome sequencing of three "Nebbiolo" clones (CVT 71, CVT 185, and CVT 423) opened new perspectives for clonal SNP-based genotyping (Gambino et al. 2017). SNP-based grapevine identification methods were recently demonstrated to differentiate three clones of Italian Nebbiolo, the famous grapevine out of which the prestigious wine Barolo is produced. Two SNPs were sufficient to identify "Nebbiolo" from 1157 genotypes. The SNP TaqMan® genotyping assays developed in this work successfully identified "Nebbiolo" in musts and wines collected at different experimental wine-making steps. The high

sensitivity of the assays allowed identification of must mixtures at 1% and wine mixtures at 10–20% with non-"Nebbiolo" genotypes. However, in commercial wines, the amplification efficiency was limited by the low amount of grapevine DNA and the presence of PCR inhibitors (Boccacci et al. 2020). Many SNP-based genotyping techniques, such those on chip allowing thousands of markers to be tested simultaneously as by use of the commercial array, Vitis18kSNP (Illumina Inc., San Diego, CA), are theoretically more prone than SSR-based authentication methods to quantitative varietal determinations of wines, but currently there are many potential issues that need implementation. First of all, assays on SNPs chips require DNA characteristics in terms of quantitative and quality even from single source biological organisms that are not generally extracted out of commercial wines, especially the aged ones. In addition, the assays on chips are done on technological platforms that may be expensive for medium and small laboratories. Furthermore, when a minimal panels of SNP markers are used on a small-scale basis, such as in a TaqMan® PCR assay, this is generally capable of distinguishing among the grapevine varieties that were sequenced in the specific region targeted by the TaqMan® probe, thus rendering the predictive ability of the assay in identifying unknown cultivars potentially ineffective. Least, but not last due to the high sensitivity of quantitative PCR to inhibitors, limitations in wine DNA purity may contribute negatively to the amplification efficiency.

Despite the fact that the SSR-based fingerprinting of grapevine products is not a quantitative technique, with respect to other molecular approaches, it has the undoubted advantage of being effective in varietal identification even using a small number of markers and the universality trait of application in the grapevine world, based on the richness and consistency of international genotype databanks that can merge into a certain food or beverages ontology identity cards. Among the in vitro factors affecting WDF efficiency, we can mention wine DNA quality, reliability of the SSR-marker panel that must retain robustness on degraded DNA admixtures, and number of contributors in the wine varietal blend (Fig. 5). Finally, as

Fig. 5 Among the in vitro factors affecting WDF efficiency, we can mention wine DNA quality, reliability of the SSRs-marker panel that must retain robustness on degraded DNA admixtures, number of contributors in the wine varietal blend

a general consideration, the WDF procedure is suitable for small and medium laboratories, but it is quite laborious especially in the initial phase, the sample preparation for DNA extraction, and the elaboration of wine genetic profiles in comparison to plant references. The universality trait of SSR-based genotyping for food authentication is also proven by the varied scientific production over the last 20 years, immediately after or in concomitance to the organization of international databases for the evaluation of biodiversity of plant and animal populations gathering the basic information for genetic identification of varieties or breeds. As an example, an SSR-based approach was used to identify the *Sus scrofa* Cinta Senese breed, and the method gave origin to a patented authentication protocol (Scali personal communication). Although the Cinta Senese is a local Tuscan breed, it is quite famous since it is identified with the immortal allegorical fresco by Ambrogio Lorenzetti on the effects of good governance, which is located in the Town Hall of Siena (Tuscany, Italy). The study was beneficial for promoting the Cinta Senese cured meat as high-quality standard DOP products containing 100% Cinta Senese meat only. The molecular method differentiated Cinta Senese pigs from white pig breeds Duroc, Hampshire, and Landrace (Scali et al. 2012) and allowed the identification of private Cinta Senese genotypes for tracking the origin of hams and milled meat samples containing a prevalence of Cinta Senese with different percentages of white pig meat up to 5% (Scali et al. 2015).

The Growing Contribution of Bioinformatics in Food and Wine Traceability

Nucleic acids could be used to characterize wines as shown in Vignani et al. (2019) through the implementation of Bioinformatics softwares. Bioinformatics has a leading role in the post-genome era and provides approaches and methods for the extraction of new knowledge from all the available molecular data. Multiple sequence alignment is of central importance for many of them. For instance, it serves as the basis for the detection of detecting conserved regions and as an important prerequisite for the construction of phylogenetic trees. Important softwares for alignment are Clustal Omega, MAFFT, MUSCLE, T-Coffee, and ClustalW. The most used package is probably Clustal Omega; its source code and executables can be downloaded from http://www.clustal.org/omega/. The basic use of Clustal Omega involves aligning a number of unaligned sequences that are all contained in a single file. For example, if the file wine.fa contains two or more unaligned sequences in fasta format, then "clustalo -i wine.fa" will read in the file, align the sequences, and output the alignment to screen (default) in the default (fasta) format.

For phylogeny inference, a great webplace for downloading software is Joe Felsenstein's phylip site http://evolution.genetics.washington.edu/phylip/software.html where one could find Maximum Likelihood packages such as PhyML, RAxML (https://sites.google.com/site/raxmlgui/), Bayesian inference such as MrBayes and BEAST (http://beast.bio.ed.ac.uk/Main_Page) and TNT for parsimony (http://www.zmuc.dk/public/Phylogeny/TNT/) that can be recommended. The portal www.

phylogeny.fr is also a good starting point for beginners as it guides from alignment to curation of the alignment to tree. Try FigTree (tree.bio.ed.ac.uk) for publication quality tree editing.

Conclusion

As global pressures for food availability and sustainable cultivation are growing rapidly, there is a technology convergence on food quality assessment. In particular molecular techniques, leveraging biological, chemical, and physical methodologies are widely used.

DNA-based tracking of food and beverages is a mark of the biological origin of the food and beverages, and in a future perspective, it may help in distinguishing genuine from synthetic "edible like substances" (Pollan 2008) against synthetic or chemical food.

There are many food products that are of superior quality (taste, texture, fragrance, etc.) because of the locale in which they are cultivated. Environmental conditions, such as local climate and soil characteristics, combine to yield crops that exhibit specific traits. Some food and beverages are recognized as being of higher quality because they derive from well-defined geographical areas, command better prices, and generally are legally protected. Indeed, the aim of Protected Designations of Origin (PDO), Protected Geographical Indication (PGI), and Traditional Specialty Guaranteed (TSG) is to add value to certain specific high-quality products from a particular origin. In perspective, molecular testing could be introduced as a voluntarily adopted method for proving intrinsic quality of many regional food products.

Bioinformatics itself is a vibrant discipline at the intersection of biology, computer science (machine learning), chemistry, and statistics. Bioinformatics uses biological and chemistry knowledge and statistical and computational techniques to support and explain cellular and molecular biology experiments. Integration of wet lab methods with bioinformatics is a promising expanding field where intrinsic quality of food and beverages merge into innovative identifying, unique profiles describing the product ontology.

References

Al-Zoubi J (2019) 150,000 liters of fraudulent oil seized by Europol. https://www.oliveoiltimes.com/business/150000-liters-of-fraudulent-oil-seized-by-europol/68065

Baker DA, Stevenson DW, Little DP (2012) DNA barcode identification of black cohosh herbal dietary supplements. J AOAC Int 95(4):1023–1034. https://doi.org/10.5740/jaoacint.11-261. PMID: 22970567

Barrias S, Fernandes JR, Eiras-Dias JD, Brazão J, Martins-Lopes P (2019) Label free DNA-based optical biosensor as a potential system for wine authenticity. Food Chem 270:299–304. https://doi.org/10.1016/j.foodchem.2018.07.058

Batrinou A, Strati IF, Houhoula D, Tsaknis J, Sinanoglou VJ (2020) Authentication of olive oil based on DNA analysis. Grasas Aceites 71(3):e366. http://grasasyaceites.revistas.csic.es/index.php/grasasyaceites/article/view/1833

Bazakos C, Khanfir E, Aoun M, Spano T, El Zein Z, Chalak L et al (2016) The potential of SNP-based PCR-RFLP capillary electrophoresis analysis to authenticate and detect admixtures of Mediterranean olive oils. Electrophoresis 37(13):1881–1890. https://doi.org/10.1002/elps.201500537

Boccacci P, Chitarra W, Schneider A, Rolle L, Gambino G (2020) Single-nucleotide polymorphism (SNP) genotyping assays for the varietal authentication of 'Nebbiolo' musts and wines. Food Chem 312:126100. https://doi.org/10.1016/j.foodchem.2019.126100

Cai Y, He Y, Lv R, Chen H, Wang Q, Pan L (2017) Detection and quantification of beef and pork materials in meat products by duplex droplet digital PCR. PLoS One 12(8):e0181949. https://doi.org/10.1371/journal.pone.0181949

Castle S (2013) Nestlé removes 2 products in horse meat scandal. The New York Times. http://businessethicscases.blogspot.com/2014/02/nestle-removes-2-products-in-horse-meat.html

Danezis GP, Tsagkaris AS, Brusic V, Georgiou CA (2016) Food authentication: state of the art and prospects. Curr Opin Food Sci 10:22–31. https://doi.org/10.1016/j.cofs.2016.07.003

Dimitri GM, Liò P (2017) DrugClust: a machine learning approach for drugs side effects prediction. Comput Biol Chem 68:204–210. https://doi.org/10.1016/j.compbiolchem.2017.03.008

Doerfler W (2020) Clinical epigenetics essential concepts are missing: foreign DNA in food invades the organism's cells and can lead to stochastic epigenetic alterations with a wide range of possible pathogenic consequences. Clin Epigenetics 12:21. https://doi.org/10.1186/s13148-020-0813-z

Dooley J, Paine KE, Garrett SD, Brown HM (2004) Detection of meat species using TaqMan® real-time PCR assays. Meat Sci 68(3):431–438. https://doi.org/10.1016/j.meatsci.2004.04.010

Food Safety Authority of Ireland (2013) Horse and Pork DNA in meat products. https://www.fsai.ie/faq/horse_pork_dna_meat_products.html

Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M et al (2013) DNA barcoding as a new tool for food traceability. Food Res Int 50:55–63. https://doi.org/10.1016/j.foodres.2012.09.036

Gambino G, Dal Molin A, Boccacci P, Minio A, Chitarra W, Avanzato CG et al (2017) Whole-genome sequencing and SNV genotyping of 'Nebbiolo' (*Vitis vinifera* L.) clones. Sci Rep 7(1):17294. https://doi.org/10.1038/s41598-017-17405-y

Gao Z, Liu Y, Wang X, Wei X, Han J (2019) DNA Mini-Barcoding: a derived barcoding method for herbal molecular identification. Frontiers in Plant Science. https://doi.org/10.3389/fpls.2019.00987

Georgiou CA, Danezis GP (2017) Food authentication: management, analysis and regulation. https://doi.org/10.1002/9781118810224

Gianì S, Silletti S, Gavazzi F, Morello L, Spinsanti G, Parati K et al (2020) aTBP: a versatile tool for fish genotyping. PLoS One 15(8):e0237111. https://doi.org/10.1371/journal.pone.0237111

Giménez MJ, Pistón F, Martín A, Atienza GS (2010) Application of real-time PCR on the development of molecular markers and to evaluate critical aspects for olive oil authentication. Food Chem 118(2):482–487. https://doi.org/10.1016/j.foodchem.2009.05.012

Gomes S, Castro C, Barrias S, Pereira L, Jorge P, Fernandes JR et al (2018) Alternative SNP detection platforms, HRM and biosensors, for varietal identification in *Vitis vinifera* L. using *F3H* and *LDOX* genes. Sci Rep 8:5850. https://doi.org/10.1038/s41598-018-24158-9

Grazina L, Amaral JS, Mafra I (2020) Botanical origin authentication of dietary supplements by DNA-based approaches. Compr Rev Food Sci Food Saf 19:1080–1109. https://doi.org/10.1111/1541-4337.12551

Haiminen N, Edlund S, Chambliss D, Kunitomi M, Weimer BC, Ganesanet B et al (2019) Food authentication from shotgun sequencing reads with an application on high protein powders. NPJ Sci Food 3:24. https://doi.org/10.1038/s41538-019-0056-6

Huang H, Zhang D, Wang Y, Perez-Neut M, Han Z, Zheng YG et al (2018) Lysine benzoylation is a histone mark regulated by SIRT2. Nat Commun 9:3374. https://doi.org/10.1038/s41467-018-05567-w

Iwobi A, Gerdes L, Busch U, Pecoraro S (2016) Droplet digital PCR for routine analysis of genetically modified foods(GMO)eA comparison with real-time quantitative PCR. Food Control 69:205–213. https://doi.org/10.1016/j.foodcont.2016.04.048

Krautkramer KA, Kreznar JH, Romano KA, Vivas EI, Barrett-Wilt GA, Rabaglia ME et al (2016) Diet-microbiota interactions mediate global epigenetic programming in multiple host tissues. Mol Cell 64:982–992. https://doi.org/10.1016/j.molcel.2016.10.025

Kumar S, Kahlon T, Chaudhary S, Chaudhary S (2011) DNA barcodes for unmasking the adulteration in olive oil. Food Chem 127(3):1335–1341. https://doi.org/10.1016/j.foodchem.2011.01.094

Martins-Lopes P, Gomes S, Santos E, Guedes-Pinto H (2008) DNA markers for Portuguese olive oil fingerprinting. J Agric Food Chem 56(24):11786–11791. https://doi.org/10.1021/jf801146z

Masada S (2016) Authentication of the botanical origin of Western herbal products using Cimicifuga and Vitex products as examples. J Nat Med 70:361–375. https://doi.org/10.1007/s11418-016-1006-0

Meerza SIA, Giannakas K, Yiannaka A (2019) Market and welfare effects of food fraud. Aust J Agric Resour Econ 63(4):759–789. https://doi.org/10.1111/1467-8489.12348

Mesnage R, Le Roy CI, Biserni M, Salles B, Antoniou MN (2019) Relationship between faecal microbiota and plasma metabolome in rats fed NK603 and MON810 GM maize from the GMO90+ study. Food Chem Toxicol 131:110547. https://doi.org/10.1016/j.fct.2019.05.055

Miceli M, Bontempo P, Nebbioso A, Altucci L (2014) Natural compounds in epigenetics: a current view. Food Chem Toxicol 73:71–83. https://doi.org/10.1016/j.fct.2014.08.005

Monfreda M, Gobbi L, Grippa A (2012) Blends of olive oil and sunflower oil: characterisation and olive oil quantification using fatty acid composition and chemometric tools. Food Chem 134:2238–2290. https://doi.org/10.1016/j.foodchem.2012.03.122

Morcia C, Bergami R, Scaramagli S, Ghizzoni R, Carnevali P, Terzi V (2020) A Chip digital PCR assay for quantification of common wheat contamination in pasta production chain. Foods 9(7):911. https://doi.org/10.3390/foods9070911

Muzzalupo I, Pisani F, Greco FC, A. (2015) Direct DNA amplification from virgin olive oil for traceability and authenticity. Eur Food Res Technol 241:151–155. https://doi.org/10.1007/s00217-015-2455-5

O'Mahony PJ (2013) Finding horse meat in beef products – a global problem. QJM 106(6):595–597. https://doi.org/10.1093/qjmed/hct087

Pereira L, Gomes S, Barrias S, Gomes EP, Baleiras-Couto M, Fernandes JR, Martins-Lopes P (2018) From field to the bottle: an integrated strategy for wine authenticity. Beverages 4:71. https://doi.org/10.3390/beverages4040071

Pollan M (2008) In defence of food. New York Times Non-Fiction Best Seller List. ISBN 0-14-314274-7

Puścion-Jakubik A, Borawska MH, Socha K (2020) Modern methods for assessing the quality of bee honey and botanical origin identification. Foods 9:1028. https://doi.org/10.3390/foods9081028

Quan PL, Sauzade M, Brouzes E (2018) dPCR: a technology review. Sensors (Basel) 18(4):1271. https://doi.org/10.3390/s18041271

Scali M, Vignani R, Bigliazzi J, Paolucci E, Bernini A, Spiga O et al (2012) Genetic differentiation between Cinta Senese and commercial pig breeds using microsatellite. Electron J Biotechnol 15(2):1–11. https://doi.org/10.2225/vol15-issue2-fulltext

Scali M, Vignani R, Quintetti M, Bigliazzi J, Paolucci E, Cresti M (2015) Genetic traceability of Cinta Senese pig (*Sus scrofa domesticus* L.): a study of the meat and processed products by microsatellite markers. Food Nutr Sci 6:712–726. https://doi.org/10.4236/fns.2015.68074

Smith M (2017) Italy Arrests 33 Accused of Olive Oil Fraud. https://www.oliveoiltimes.com/business/italy-arrests-33-accused-olive-oil-fraud/55364

The EU Food fraud network and the system for administrative assistance – food fraud, annual report 2018. https://ec.europa.eu/food/sites/food/files/safety/docs/ff_ffn_annual-report_2018.pdf

The EU Food fraud network and the administrative assistance and cooperation system, annual report 2019. https://ec.europa.eu/food/sites/food/files/safety/docs/ff_ffn_annual-report_2019.pdf

The FoodOn Food Ontology, BioPortal (2020). https://bioportal.bioontology.org/ontologies/FOODON

Valdés A, Simó C, Ibáñez C, García-Cañas V (2014) Emerging RNA-Seq applications in food science. In: Virginia García-Cañas V, Cifuentes A, Simó C (eds) Comprehensive analytical chemistry, 64. Elsevier, pp 107–128. https://doi.org/10.1016/B978-0-444-62650-9.00005-1

Vignani R, Liò P, Scali M (2019) How to integrate wet lab and bioinformatics procedures for wine DNA admixture analysis and compositional profiling: case studies and perspectives. PLoS One 14(2):e0211962. https://doi.org/10.1371/journal.pone.0211962

Villa C, Costa J, Meira L, Oliveira MBPP, Mafra I (2016) Exploiting DNA mini-barcodes as molecular markers to authenticate saffron (*Crocus sativus* L.). Food Control 65:21–31. https://doi.org/10.1016/j.foodcont.2016.01.008

Wang J, Hodes GE, Zhang H, Zhang S, Zhao W, Golden SA et al (2018) Epigenetic modulation of inflammation and synaptic plasticity promotes resilience against stress in mice. Nat Commun 9:477. https://doi.org/10.1038/s41467-017-02794-5

Weimer BC, Storey DB, Elkins CA, Baker RC, Markwell P, Chambliss DD et al (2016) Defining the food microbiome for authentication, safety, and process management. IBM J Res Dev 60, (5/6), 1:1–1:13. https://doi.org/10.1147/JRD.2016.2582598

White V (2017) Food Fraud: a challenge for the food and drink industry. https://www.newfoodmagazine.com/article/22854/food-fraud-an-emerging-risk-for-the-food-and-drink-industry/

Wu Y, Wan J, Choe U, Pham Q, Schoene NW, He Q et al (2019) Interactions between food and gut microbiota: impact on human health. Annu Rev Food Sci Technol 10(1):389–408. https://doi.org/10.1146/annurev-food-032818-121303

Yuan X, Tsujimoto K, Hashimoto K, Kawahori K, Hanzawa N, Hamaguchi M et al (2018) Epigenetic modulation of Fgf21 in the perinatal mouse liver ameliorates diet-induced obesity in adulthood. Nat Commun 9:636. https://doi.org/10.1038/s41467-018-03038-w

Zhang Y, Kutateladze TG (2018) Diet and the epigenome. Nat Commun 9:3375. https://doi.org/10.1038/s41467-018-05778-1

Zwirzitz B, Wetzels SU, Dixon ED, Stessl B, Zaiser A, Rabanseet I et al (2020) The sources and transmission routes of microbial populations throughout a meat processing facility. NPJ Biofilms Microbiomes 6:26. https://doi.org/10.1038/s41522-020-0136-z

Use of DNA Barcoding for Plant Species Identification

41

Jaskirandeep Kaur Jossan and Rajinder Singh

Contents

Abstract

In criminal investigations, many a times the investigator can came across evidentiary materials of plant origin including poisonous and hallucinogenic fungi. Its identification can help to determine a sample's geographic origin, ascertain the possession or trade in forbidden or endangered species, and provide links between crime scene and individuals. DNA barcoding has emerged as an important technique for the identification of species over the last two decades. It is the most-preferred technique by forensic scientists over conventional methods across the globe for identifying the species of origin for seized materials owing to its

J. K. Jossan · R. Singh (✉)
Department of Forensic Science, Punjabi University, Patiala, Punjab, India
e-mail: rajinder_forensic@pbi.ac.in

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_60

ability to analyze the forensic samples easily and accurately even in powdered, degraded, and stained forms. In this chapter, various aspects of plant DNA barcoding and usage of specific barcodes for the species identification of diverse evidentiary materials of botanical origin are highlighted and discussed.

Keywords

Forensic botany · DNA barcodes · Botanical evidence · *mat*K · *rbc*L · ITS

Introduction

Linking a suspect to a crime scene is a common forensic goal. Criminal investigations involve the recovery, processing, and analysis of a wide range of evidentiary materials of different origins, natures, and in different forms (Frances et al. 2010). Plant materials such as grasses, shrubs, weeds, seed pods, pollen grains, leaves, etc., are frequently encountered in criminal investigations and can often be found on other evidence types, as bodies, clothes, shoes, vehicles, tools, weapons, etc. Plant identification can help to determine a sample's geographic origin, provide links between crime scene and individuals, and ascertain the possession or trade in forbidden or endangered species (Ferri et al. 2009).

In practical terms, forensic botany can present additional information in many forensic cases involving plant evidence that may be useful to link a suspect, a victim, or a vehicle to the crime scene. During the commission of outdoor crimes, plant material may be transferred from the crime scene to the victim or perpetrator, and this kind of trace may be probative due to restricted and specific geographical distribution. Plant re-colonization (succession) of land that has been disturbed will follow a specific set of patterns that can be useful to estimate the time of death. Identification of vegetable contents in stomach or partially plant digested in feces can support or disprove an alibi. Again, forensic botany can be used to identify primary or secondary crime scene and track drug distribution network.

Types of Cases

There are a number of cases in which plant materials are recovered and play a vital role in solving the crime:

Transferred Trace Evidence in Physical and Sexual Assault Cases

Plant evidence plays an important role in criminal investigations which further helps in solving critical cases and works as a vital evidence in the court. During the investigation, the forensic scientists come across different types of plant materials, so in order to get full information, the specialists need to know the cellular features, plant systematic taxonomy, pollens identification points, plant ecology, plant succession patterns, and limnology. If analyzed properly, these evidences will help to

reconstruct the crime scene (Coyle et al. 2003). The use of plant material as trace evidence is still novel in criminal case study because of lack of knowledge or the limitations of the techniques to identify the samples accurately. Forensic botanists are still working with their traditional identification methods, but as the era is advancing some of the labs are working toward the progressive approach that is DNA barcoding.

Most of the times, in forensic palynology, where pollens are the main source of evidence, including angiosperms and gymnosperms, help in solving criminal cases. Pollens are microscopic and not visible to the naked eyes; they are liable to be missed by the criminal and left behind during the clearance of the crime or in the process of manipulation. Pollens are mostly found on clothing, shoes, and carpets. These types of trace evidences are helpful in locating the geographical origin of the samples. As plant ecology is geographical specific, this makes a specific pollens pattern distribution. Brown et al. (2002) reported a case where the pollens and soil analysis helped in solving the crime. The pollens were found at shoeprints which matched with the pollens on shoes of the culprit. These pollens from the scene of crime and the culprit were similar, and it helped to link the culprit with the crime scene. In the recent past years, there are a number of cases reported which have been solved with the help of palynology.

In 1995, a murder case in Texas which was remained unsolved for 1 year because the victim was gruesomely murdered and identification was not possible. The victim was stabbed 21 times; his hands were chopped so there was no fingerprint for the matching purpose. So, the case was declared as a cold case. Once in 1996, the investigator attended a forensic lecture where he learned about forensic palynology and thought to give it a shot. After that, they collected the pollens from the corpse clothing, socks, and shoes, and control samples were collected from the area of crime. After the identification, the pollens were found to be of Marijuana and were mostly located on the shirt and socks of the victim. It was noted that the quantity of the samples from the victim was comparatively much higher than that of the present at the area of the crime. Along with the Marijuana, pollens of *Alnus glutinosa* and *Shepherdia argentea* were also found which raised the suspicion that the victim had been relocated from the original crime scene as these pollens were not present at the crime scene. Further investigation was preceded and concluded that the primary crime scene was in Kansas, later the victim was dumped in Texas (Bryant 2001).

Cases Related to Forbidden/Controlled Materials (Drugs and Psychoactive Substances)

Any substance that is taken for non-medical reasons, mainly for the mind-altering effect, is known as drugs of abuse. It is a very long tradition to use herbal products as drugs and the most common are cannabis and its products, opium and its derivatives, and cocaine. According to a survey in 2017, conducted by National Survey of Drug Use and Health (NSDUH) Releases (2017) in the USA, it was reported that 24 million Americans were abusers in 2016. Among them, 14.7 million were men and approx. 9.3 million were women. World Health Organization (WHO) also reported in 2018 that cannabis is the main drug that is still in high demand worldwide with

over 192.2 million users. The main consumers are from West and Central Africa (13.2%), North America (12.9%), and Oceania (11.0%). 53% of the total global drug seizure cases are related to cannabis followed by opium (15%) and then cocaine (10%).

In 2015 and 2016, 39 and 98 death cases, respectively, were reported in Germany due to a combination of new psychoactive substances in conventional drugs. New psychoactive drugs named synthetic cannabinoids (Spice), synthetic cathinones (bath salts), and novel opioids were responsible for most of the accidental poisoning cases (Kraemer et al. 2019). Gravensteen et al. (2019) analyzed 1200 individual's data from Norway and Sweden from the year 2008 in order to analyze the cause of death. In the aforementioned cases, 66% and 74% deaths due to suicides were reported in Norway and Sweden, respectively, and in addition 85% and 66% deaths were reported as accidental due to psychoactive drugs.

Cases Related to Protected Plants Under Environment Laws

In the USA alone, 5,600 plant species are trafficked in a year. The plants are Orchids, Cycads, and Agar-wood. In the time period from 2010 to 2014, 20% trade was related to Orchids, 33% related to Cycas, while there was no reported export of agar wood (used for costly perfumes) between 2005 and 2014. The Convention on International Trade in Endangered Species of wild fauna and flora (CITES) consists of three appendices where different species have been categorized on the basis of their rate of extinction. Appendix I includes 300 plant species which are on the verge of extinction, and in Appendix II, 29,000 plant species are included that are not threatened currently, in addition, 18 plant species which are in danger due to commercial exploitation in the local areas. It has been estimated that 70–80% of people chiefly believe in the traditional herbal medicinal system. It has been practiced in China and India from ancient times as Ayurveda, Unani, Siddha, and Tibetan medicinal systems. Many medicinal plant species have reached the point of extinction due to their over-exploitation from the natural habitat. The illegal trading of medicinal plants has a fair share in global industry that is approx. 62 billion USD.

Cases Related to Poisoning

A number of accidental, homicidal, or suicidal cases have been reported due to poisoning by plants. Most of the times people or children are unaware of the nature of the seeds and the fruits they are consuming (Adhikari 2012). The American Association of Poison Control Centers' National Poison Data System (NPDS) in 2013 reported 46,376 cases of plant poisoning and most of them were related to children (29,346). According to World Drug Report (2018), the rate of fatal poisoning is higher in children under 1 year. Dayasiri et al. (2017) published a data regarding the plant poisoning in rural areas of Sri Lanka where among 325 children, 57% were male with 64% being below the age of 5 years. 99.4% among them had ingested the poison accidently. Most of the poisoning cases were due to *Jatropha circus*, *Ricinus communis*, and *Jatropha multifida.*

The most common types of poisonous plants found in India are (1) contact irritant poisons – *Ricinus communis, Calotropis procera, Codiaeum variegatum, Gloriosa*

superba, Semecarpus anacardium, Capsicum annuum; (2) cardiotoxic glycosides – *Aconitum*, and *Colchicum autumnale*; (3) neurotoxic plants – *Strychnosnux-vomica, Gloriosa superba, Calotropis gigantea, Datura stramonium*; (4) hepatotoxic plants – *Azadirachta indica*, and (5) miscellaneous toxic plants and plant products – *Areca catechu* and *Cleistanthus collinus* (Babu et al. 2016).

In all these types of cases, species identification of such evidence is of utmost importance for the successful implementation of the related laws. Conventionally, morphological characteristics based on taxonomic keys are used to identify species. This approach is feasible only when the evidence is recovered in intact or sufficiently intact form. However, it becomes difficult to employ this approach, if the evidentiary material is in powdered, degraded, and stained form. In such situations, DNA barcoding technique is the method of choice for this purpose as it does not rely on morphological characteristics of the sample.

DNA Barcoding for Species Identification in Forensic Botany

In the history of forensic science, DNA genotyping proved to be the milestone for the purpose of individual identification, which progressed rapidly after the invention of DNA fingerprinting (Jaffreys et al. 1985; Jobling et al. 2004). Similarly, many DNA-based techniques such as Restriction fragment length polymorphism (RFLP), Random amplified polymorphic DNA (RAPD), Amplified fragment length polymorphism (AFLP), Microsatellites and single nucleotide polymorphisms (SNP), Cleaved amplified polymorphic sequences (CAPS), and DNA barcoding have evolved over a period of time for the species identification and other scientific purposes. DNA barcoding is a technique in which one or more short gene sequences are taken from a standardized portion of the genome and used to identify the species (Kress and Erickson 2007). In the last two decades, DNA barcoding has proven its importance in forensic science and has been utilized in the identification of many non-human evidence such as animals and their body parts, plants and fungi, etc., (Cassidy et al. 2005).

The pioneer work on DNA barcoding was reported in 2003 when Hebert et al. published species identification using a part of mitochondrial DNA, i.e., Cytochrome c Oxidase subunit 1 (*CO*1 or COX1). The term DNA barcode represents a small gene sequence from a standardized region of DNA which possesses enough variation among different species that can be used for species identification. DNA barcoding relies on a specific segment of DNA which shows variation among different species. After this ground-breaking work, many researchers started working on different forms of life, e.g., plants and fungi. *CO*1 is considered to be a universal barcode for the identification of animal species (Hollingsworth 2008), whereas, in case of plants; such a barcode is yet to be identified. There are certain conditions which must be met for a barcode to be a universal barcode such as universality, sequence quality and coverage, and level of species discrimination.

Universality deals with the application of DNA barcodes to a large number of species and a universal barcode must be able to identify and differentiate them

(Frezal and Leblois 2008). The sequence quality and coverage mean that a universal DNA barcode must be sequenced easily without generating a lot of noise. This will result in efficient interpretation of the result and accurate species identification (Frezal and Leblois 2008). Level of species discrimination means that more the species are differentiated by a barcode, the better is its efficiency. An ideal barcode must differentiate all the species of a particular group such as bacteria, fungi, animals, or plants.

In plants, many DNA markers/loci such as *rbc*La + *mat*K, *psb*A-*trn*H, *trn*L-F, nrITS, *rpl*32-*trn*, and *psa*A-*ycf*3 have been reported for species identification (Boatwright et al. 2015; Daru et al. 2013; Kyalangalilwa et al. 2013; Manning et al. 2014; Maurin et al. 2016). The Consortium for Barcoding of Life (CBOL), plant working group recommended a core barcode of two plastid-coding regions, *rbc*L + *mat*K augmented with additional markers for this purpose. Plastid DNA barcode *mat*K is a rapidly evolving DNA barcode and has become the closest plant analogue of CO1. Many studies have reported very high species identification rates for *mat*K (Cuénoud et al. 2002; Chase et al. 2007; Hollingsworth 2008; Lahaye et al. 2008; Ford et al. 2009). However, it is evident that more work needs be done before accepting *mat*K as a universal DNA barcode for plants. In case of fungi, the second largest group of eukaryotic living organisms, nuclear DNA ITS region is regarded as the universal DNA barcode. It has the highest probability of successful identification for the broadest range of fungi, with the most clearly defined barcode gap between inter- and intraspecific variation (Schoch et al. 2012).

*mat*K (Maturase K)

*mat*K gene is present in the chloroplast DNA and is highly conserved. Nearly, in all the terrestrial plants, *mat*K is encapsulated between *trn*K and tRNA-lysine (exons). *mat*K, earlier known as Orf, is about 1500 bp in most of the angiosperms (Hilu et al. 1999). It has a high evolutionary rate, appropriate length, and inter-specific divergence, as well as low transmission/transversion rate (Min and Hickey 2007; Selvaraj et al. 2008) (Fig. 1.)

However, with the present availability of primer sets of *matK*, it is nearly impossible to amplify it in all the plant species. According to CBOL (2009), *mat*K marker is able to amplify 90% of angiosperms, whereas in gymnosperms it is

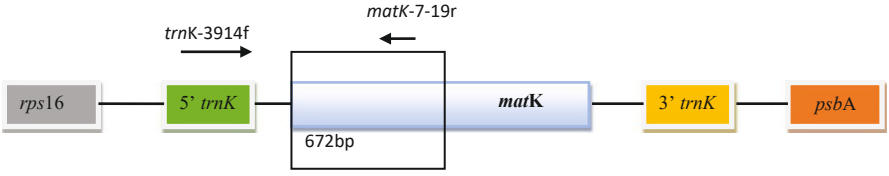


Fig. 1 Diagram showing *mat*K gene nested between introns. The arrows indicate the positions of PCR and sequencing primers

possible to amplify in 83% and in cryptogams it is only 10%. Many authors have tried different primer sets for the amplification of *mat*K, giving the success rate of 100% for angiosperms (Chase et al. 2007; Hollingsworth 2008; Lahaye et al. 2008). For being a universal barcode, the marker should have a high discrimination rate; therefore, in the case of *mat*K, the discrimination rate *Orchidaceae* is more than 90% (Kress and Erickson 2007), while some authors have contradicting results where *mat*K showed 45% of discrimination rate in Nutmeg (Newmaster et al. 2008).

Another plastid DNA barcode, *rbc*L (1400 bp) encodes a large subunit RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) (Fig. 2), which is widely used in the sequencing of plant taxa. It is a bifunctional enzyme which helps in catalyzing both the carboxylation and oxygenation of ribulose-1,5-bisphosphate (RuBP) that leads to fixing of CO_2 as the first step of the Calvin cycle. *rbc*L was the first gene which was used in plants for sequencing purposes (Zurawski et al. 1981). It can be amplified, sequenced, and aligned in most of the land plants with ease, in spite of having low discrimination power. Due to its intensive use in plant studies, more than 10,000 *rbc*L sequences are accessible in GenBank.

Inspite of all these achievements, *rbc*L is only best for identifying the relationship between the species at genus level, but not up to species level (Soltis et al. 1998). The major limitation of the *rbc*L marker is having a large sequence (1409 bp) which does not fulfill the criteria of ideal DNA barcode, especially for forensic purposes. Ideally, DNA barcode should have a short sequence length that can be useful in amplifying the degraded DNA and can be easily sequenced in one pass (Chase et al. 2007). The limitation was overcome by developing the short sequence for plant species identification (Kress and Erickson 2007).

Burgess et al. 2011, identified local floras using the combined DNA barcode *rbc*L + *mat*K recommended by CBOL, where *rbc*L gave the best sequencing rate of 91.4% as compared to other markers. *rbc*L + *mat*K gave high-quality sequences that could be used for identification of temperate flora, giving the success rate of 91.3%. The combination of these two markers gave 100% results in gymnosperms and 92.7% in angiosperms. For *Quercus, rbc*L shows 40% of species resolution (Piredda et al. 2011) while for *Lemnaceae* it gives 100% success rate (Wang et al. 2010). The universality rate of *rbc*L marker was also studied by numerous scientists and reported a sequencing rate of 95% (Kress et al. 2005; Kress and Erickson 2007), 90% (Newmaster et al. 2008), 100% (Fazekas et al. 2008), and 93% (Hollingsworth

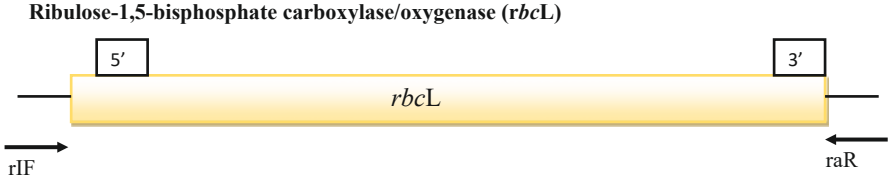


Fig. 2 The orange bar represents the open region frame, *rbc*L and primers are represented by arrows: r1F (*rbcL*1F); raR (*rbcL*aR)

et al. 2009). CBOL has also recommended *rbc*L as a promising gene for plant identification with the combination of *mat*K markers.

While, some authors have turned down *rbc*L as a selected barcode region for plants (Gielly and Taberlet 1994; Renner 1999; Salazar et al. 2003; Mattia et al. 2011). The *rbc*L *marker* has also been employed in the identification of grass species. It gives 100% PCR amplification and sequence recovery rate for 54 species of the grass family (Saadullah et al. 2016). On the contrary, Tahir et al. in 2018 analyzed that *rbc*L marker, which can only identify 90% of grasses; in which, 90% genus level and 30% for species level. Bafeel et al. 2011 matched the BLAST output of all the sequences and concluded that *rbc*L marker can only be assigned to 92% of species up to genus level, whereas only 17.5% of species were identified up to species level. It can be concluded from the previous research that *rbc*L marker is best for identification up to genus level, but for species-level identification it has less discriminatory power than *mat*K.

Internal Transcribed Spacer (ITS)

Along with plastid DNA markers, a nuclear DNA barcode known as ITS is considered as the powerful phylogenetic marker that can resolve a plant species up to species level. Whereas CBOL has recognized the ITS marker as a supplementary marker because of its incomplete concerted evolution, fungal contamination, and not easily amplified and sequenced. Kress (2017) proposed ITS region and plastid region, *trn*H-*psb*A as a potential barcode region for the identification of flowering plants. According to this study, ITS marker has proven to be successful in the case of photosynthetic eukaryotes (except ferns) and fungi (Stoeckle 2003).

Furthermore, ITS can easily be amplified in two smaller fragments, i.e., ITS1 and ITS2 (Fig. 3). China plant barcode of life groups has consented to ITS2 as an alternative marker of ITS because of easy amplification and sequencing for plant species identification. Hence, in the case of plants, ITS2 marker has been used for the identification of *Fabaecae* (Gao et al. 2010), *Poaceae* (Saadullah et al. 2016), medicinal plants (Gao et al. 2010), *Lamiaceae*, Dicotyledons, Monocotyledons Ferns, Mosses (Yao et al. 2010), and many more. ITS2 has a short gene sequence

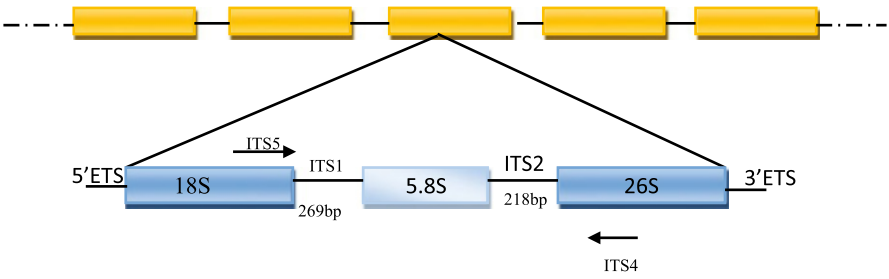


Fig. 3 Structure of ITS region of nuclear ribosomal RNA gene

length, easily amplified with a single pair of primers, and high sequencing efficiency. ITS2 can be used as a complementary locus to *CO*1 for the identification of plants (Chen et al. 2010; Han et al. 2013). ITS2 marker has conquered the limitation associated with ITS marker, where PCR amplification makes multiple copies. Due to its secondary structure, it becomes a systematic tool for the identification of botanical species (Han et al. 2013). The ITS2 marker is considered as the best marker for amplification from degraded samples.

According to Han et al. 2013, ITS2 is small in size and has the ability to be a universal primer even for degraded DNA samples with great success. In the case of *Zingiberaceae*, ITS2 works better than *rbc*L and *mat*K markers for the discrimination upto species level (Dhivya 2012). ITS2 marker is also considered as the universal marker for medicinal plants and their materials (Ward et al. 2009; Grebenstein et al. 1998).

Protocol for Plant Species Identification Using DNA Barcodes

In order to achieve accurate and reliable results, each and every step must have a standardized protocol. The forensic evidences are often recovered in contaminated, degraded, and in very less amount. Therefore, an accurate and standardized procedure should be followed for reliable results. Various steps involved in the plant species identification using DNA barcoding technique are given in Fig. 4.

Collection of Plant Samples for DNA Isolation

There are a number of ways to collect the botanical evidence from the outdoor crime scene. The plant tissues can be stored in liquid nitrogen at −80 °C. But in most of the developing countries, this procedure comes with a heavy cost, so lyophilization or freeze-drying is a good alternative for storage. Plant tissues can be collected and stored in silica gel for a cheaper option (Chase and Hills 1991; Liston et al. 1990). The silica gel helps to remove moisture from fresh plant tissues. There are many dried tissues in plants (wood) that don't need preservation and can be stored for weeks and months even at room temperature prior to DNA isolation. The proper collection procedure consists of various tools that might be required during the process are listed in Table 1.

Precaution must be taken that the samples should never be collected in plastic bags or non-porous surfaces, for the reason that plastic material enhances the growth of fungi and bacteria, which will lead to the ruining of botanical evidence. Sampling depends upon the size and the type of plant material found, if the plant is small in size it can be collected as a whole, but if the plant is large, it can be folded, zigzag, or accordion style.

In the conditions, where the plant cannot be folded, the investigator can cut the sample into half or into sections. Each part of the same plant material should be collected separately in paper bags but with the same evidence number. The evidence

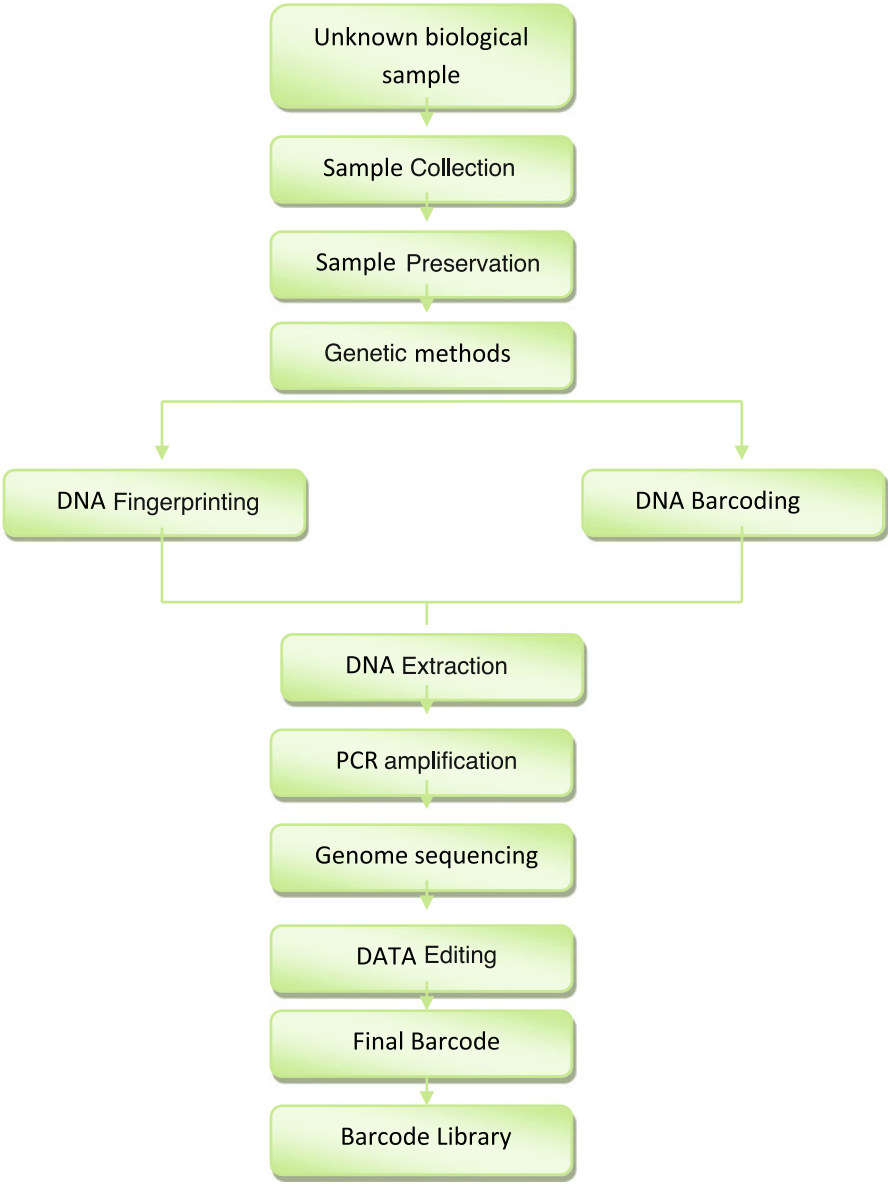


Fig. 4 Flow chart of the procedure for plant DNA barcoding

should be cut with a sharp knife because the blunt tools can damage the taxonomical characteristics of the plant. Minimum five samples of each species should be collected for removing any chance of error and to account for intra-species variations. Sufficient numbers of fruits and flowers should be collected so that they can be analyzed without any error. Seeds contain less quantity of polyphenols that act as a

Table 1 Tools required for storage of plant samples and their functions

Requirements	Function
Silica gel	Drying of samples
Paper envelopes	Collection of plants and plant materials
Permanent markers	Labeling
Scissors or sharp knife	Cutting of large samples
70% ethanol	Sterilization and preservation of samples
Camera	Photographing the crime scene

Table 2 Different preservatives used for preservation of plant samples

Preservatives	Samples	References
Modified Carnoy's solution	Leaf samples	Doyle and Dickson 1987
Chloroform: ethanol		
Brine solution		
Dry ice or liquid nitrogen	Plant tissue	Kilpatrick 2002; Wehausen et al. 2004; Reusch 2006
Buffer solutions containing silica gel		
Ethanol	Plant leaves	Bressan et al. 2014

PCR inhibitor and also have endosperm which possess a sufficient amount of DNA. It is non-destructible evidence that does not need much of preserving chemicals, making it the best plant material for DNA analysis. It can be collected in a packet made from a folded piece of paper, called a druggist fold. While the samples like branches, roots, bark, and stems can be collected in cardboard boxes. It is very important to look out for the samples in the form of stains on the victim's cloth, weapon, and even on the footwear. Mostly they are overlooked because of their size and amount.

Another evidence that is also important but is often neglected are pollens. Pollens can very easily stick on the suspect-related items and can transfer from one place to another without being noticed. Pollens can also be located in the soil, dust, and mud. So, it is very important to locate and then collect them carefully. The samples can be collected with the help of cotton swab, brushing, vacuum cleaning, and electrostatic techniques (Coyle et al. 2005).

Preservations of Samples

In order to achieve successful results, it is paramount to preserve the plant sample accurately to prevent sample degradation and contamination due to environmental factors and microbial activity including fungi. Various preservatives used for preserving botanical samples for DNA analysis are given in Table 2. Storage in molecular-grade ethanol, silica gel, and cold conditions such as in dry ice or liquid nitrogen are most widely agreed upon methods for this purpose.

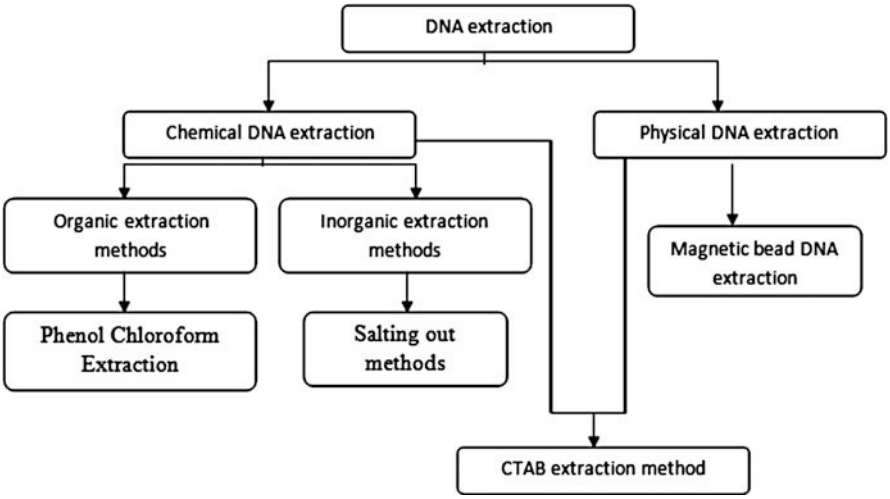


Fig. 5 Flow chart of the different types of DNA extraction methods

Plant DNA Isolation

The availability of high levels of endogenous phenolics, polysaccharides, or other secondary metabolites leads to difficulty in the isolation of DNA from plants. The extraction of good quality DNA depends upon the collection of the sample, the preparation of plant materials, type of protocol, and the most important, the usage of chemicals in the procedure. Good quality DNA is the prerequisite for its downstream analysis. Various protocols have been reported in the literature for this purpose depending upon the type of plant sample used like dry plant tissue from herbarium specimens and have been described (Wittzell 1999; Ristaino et al. 2001; Rogers and Bendich 1994). Various commonly used methods for this purpose are summarized in Fig. 5.

Organic Extraction Method

Organic extraction method or phenol-chloroform extraction method is the most common procedure for DNA extraction that utilizes chemicals and buffers for high-quality DNA from the biological samples. The procedure consists of PCI solution (phenol, chloroform, and isoamyl alcohol in the ratio 25:24:1, respectively) and lysis buffer (Tris, EDTA, $MgCl_2$, NaCl, SDS, and other salts). Various buffers and chemicals are used for this purpose, and their respective functions are summarized in Table 3.

Cetryl Trimethylammonium Bromide (CTAB) Method

Plants consist of high amounts of secondary metabolites, proteins, polysaccharides, and polyphenolic compounds, which make DNA extraction from plant materials

Table 3 Buffers and chemicals used in DNA extraction, and their respective functions

Buffer/chemical	Function
Tris buffer	Helps in maintaining the pH of solution and combines with lipopolysaccharides of the plant cell membrane and to cut it by making it permeable
EDTA	It's a chelating agent and inhibits the DNase activity by blocking its binding site.
SDS (Sodium dodecyl sulfate)	An anionic detergent, used for the lysis cell and nuclear membrane
NaCl	Na^+ forms an ionic bond with negatively charged DNA molecule for neutralization which leads to the prevention of denaturation of combined DNA
$MgCl_2$	Protects DNA from combining with other organelles by blocking the negative charge of lipoproteins.
Chilled alcohol	Precipitation of DNA
TE buffer	Storage and preservation of DNA

difficult. The CTAB extraction method is based on chemical and physical extraction procedures for the extraction of DNA in difficult plant tissues. CTAB is a cationic detergent that helps in the separation of polysaccharides in the process of purification along with polyvinylpyrrolidone (PVP), which aids in the removal of polyphenols.

Murray and Thompson in 1980 developed an inexpensive and easy DNA extraction protocol from plants by using CTAB. In order to achieve good quality and high yield of DNA using the CTAB method, Aboul-Maaty and Oraby in 2019 made some modifications in the conventional CTAB method. The reagents used in the protocol are CTAB, β-mercaptoethanol, chloroform: isoamyl alcohol, NaCl, potassium acetate, ice-cold 100% isopropyl alcohol, 70% ethanol, 1× TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM, EDTA, pH 8.0, autoclaved), and agarose (molecular grade). The sample is ground in liquid nitrogen and added into preheated CTAB extraction buffer containing β-mercaptoethanol. Then, the sample is subjected to water bath which is followed by addition of Chloroform: isoamyl alcohol. The centrifugation step is a must for the separation of aqueous layer. After the separation of aqueous layer from rest of the solution, it is mixed with NaCl. Again, the centrifugation is carried out for precipitation of DNA pellet. For washing, 70% ethanol is required to remove any unwanted chemical left behind. At last, the DNA is preserved in TE buffer.

CTAB has proved to be a reliable method for the isolation of plant DNA, contributing to 37% of research work reported so far (Fig. 6).

Quality and Quantity Assessment

Whether the extracted genomic DNA will give good-quality sequencing results depends upon its purity, quality, and quantity. The optimal size of DNA required with the latest technology is over 10 μg with 30–40 kbp template size for long-read sequencing. NanoDrop and UV spectrophotometers (Thermo Fisher Scientific, Waltham, Massachusetts, USA) are used for the quality and quantity estimation of

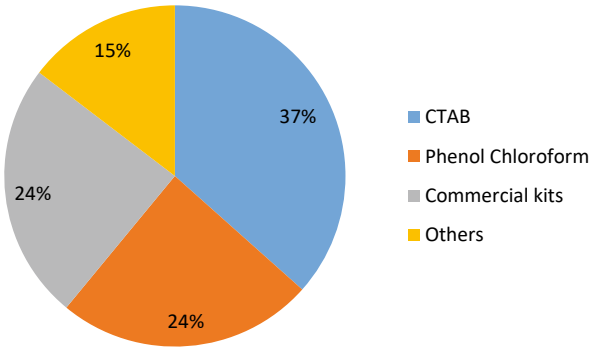


Fig. 6 Percentage distribution of DNA extraction protocols

the DNA. The absorbance reading should be 260/280 1.8–2.0. The value of 260/280 should be 1.8–2.0. DNA, proteins, and salts absorb UV rays at 260 nm, 280 nm, and 230 nm, respectively. The value ranges should fall between 1.8–2.0 at 260/280 nm for purified DNA.

Currently, real-time PCR (RT-PCR), being simple and reliable, is the most widely utilized technique for the quantification of genomic DNA. As compared to, conventional PCR techniques, RT-PCR technique is more suitable because of its quantitative performance, use of closed-tube assays and greater sensitivity. In RT-PCR two quantification strategies are performed which are absolute and relative quantification. In absolute quantification, a calibration curve is used to input PCR signal, whereas, relative changes in mRNA expression levels are measured by relative quantification. The efficiency of absolute quantification is dependable on a condition "identical" amplification efficiency of a native target and calibration curve considered. The calibration curve and data are always reproducible as well as sensitive. The relative quantification is easier to process as compared to absolute quantification as no calibration curve is needed. Moreover, relative quantification is dependent on expression levels or target DNA against the reference gene. (Livak 1997).

In order to perform Rt-PCR, a specific control is used which is complementary to target DNA. These are known as pre-defined assay reagents which contain primers and probes specific to target DNA. The cloned products are then diluted in 1:1000. The standard for cDNA quantification, the exact copy number, and concentration should be known of DNA so that, precise number of molecules could be calculated in RT-PCR. The calibration curve of natural log of the threshold cycle (CT) vs the natural log of the number of molecules will ultimately reveal the quantity of DNA.

PCR Amplification of the DNA Barcodes

Polymerase chain reaction (PCR) is the amplification of target DNA template to produce specific DNA fragment in vitro. It was developed by Kary Mullis and his coworkers in 1985, who works as a milestone in the forensic DNA profiling.

PCR amplification is totally dependent upon the concentration of the components used and the temperature provided. There are three temperature-controlled phases: Denaturation, Annealing, and Extension. Denaturation is carried out at 94–95 °C to open the DNA strands. Standardizing and maintaining appropriate annealing temperature (Tm) is the key to success in PCR as it is helpful in binding the primer with the DNA template. If the annealing temperature is high, the primers will bind poorly to the template and if it is low then undesired segments of DNA will be annealed. The extension of primers is the last and final step in this process and is done at 72–78 °C. In the literature, numerous primers with different annealing temperatures have been reported to amplify the plant DNA barcodes, which are summarized in Table 4.

Plant Genome Sequencing

DNA sequencing is the process to determine the sequence of nucleotides in the target DNA. Maxam-Gilbert chemical cleavage and the Sanger chain termination method were the two methods developed in the mid-1970s for directly sequencing DNA. Sanger and Coulson introduced the plus and minus method for sequencing in 1975, which was soon replaced by the chain termination method. Chain termination method was the basic technique that led to modern DNA sequencing techniques. Sanger, Nicklen, and Coulson in 1977 developed an efficient technique which requires fewer amounts of toxic and radioactive chemicals. Dideoxynucleotidetriphospahtes (ddNTPs) are used as the DNA chain terminator in the Sanger method. *The chain termination method requires a DNA template, primer, radioactive nucleotides, and ddNTPs for the termination of DNA elongation. Dideoxyribonucleotides lack in 3' OH group which is needed for the addition of another molecule. Thus, the termination of the process at the DNA strand is of different length. After the synthesis of new DNA strands of varied lengths, they are denatured and separated on the basis of their size on polyacrylamide-urea gel. The gel is visualized under UV light and read off under X-ray film. After the exposure of X-ray to the gel, the dark bands are formed depicting the various DNA of different lengths. The dark bands are representative of the DNA fragments after the chain termination* (Sanger et al. 1977).

The first whole-plant genome sequence project was announced in the year 2000 on *Arabidopsis thaliana species targeted for* de novo *genome sequencing. It took 10 years and 100 million USD to complete the project (Goff* et al. 2002*).* After the successful project of *Arabidopsis thaliana, genome of Oryza sativa was sequenced* validating that it was followed by *Arabidopsis thaliana* yielding the "finished quality" genome. *However, with the passage of time the cost reduces, but the read length is still needed to be taken care of.* In the case of plants, it is not necessary that all are equally sequenceable because there are three main factors which will determine the feasibility, such as genome size, repeat structure and age, and heterozygosity. The plant genome is of diverse nature and size ranges from 0.063 to 148.8 Gbp (Greilhuber et al. 2006; Hidalgo et al. 2017). The cost of sequencing depends upon the size of genome. Therefore, a few plants genome larger than 10 Gbp can be

Table 4 Various primers sets that are used for the amplification of plant DNA barcodes

Region	Primer Name	Sequence 5′–3’	Primer length (bp)	Tm	References
ITS	ITS4_KYO1	TCCTCCGCTTWTTGWTWTGC	20	47 °C	(Toju et al. 2012)
	ITS4_KYO2	RBTTTCTTTTCCTCCGCT	18	50 °C	
	ITS4_KYO3	CTBTTVCCKCTTCACTCG	18	53 °C	
	ITS-u1	GGAAGKARAAGTCGTAACAAGG	22	55–58 °C	(Cheng et al. 2015)
	ITS-p5	CCTTATCAYTTAGAGGAAGGAG	22		
*rbc*L	*rbc*La_f	ATGTCACCACAAACAGAGACTAAAGC	26	55 °C	(Bafeel et al. 2012)
	*rbc*Laj634R	GAAACGGTCTCTCCAACGCAT	21	53 °C	(Fazekas et al. 2009)
	*rbc*Lr590 R	AGTCCACCGCGTAGACATTCAT	22	50 °C	(Vere et al. 2015)
	*rbc*La-rev	GTAAAATCAAGTCCACCRCG	20	55 °C	(Kress et al. 2009)
	*rbc*L724R	TCGCATGTACCTGCAGTAGC	20	50 °C	(Fay et al. 1997)
*mat*K	*mat*K2 1a	ATCCATCTGGAAATCTTAGTTC	22	49 °C	(Ford et al. 2009)
	*mat*K_390f	CGATCTATTCATTCAATATTTC	22	50 °C	(Cuénoud et al. 2002)
	*mat*K_Xf	TAATTTACGATCAATTCATTC	21	50 °C	(Ford et al. 2009)
	*mat*K_1326r	TCTAGCACACGAAAGTCGAAGT	22	53 °C	(Cuénoud et al. 2002)
	*mat*K_5r	GTTCTAGCACAAGAAAGTCG	20	45 °C	(Ford et al. 2009)
	*mat*K 3F_KIM	CTTCCTCTGTAAAGAATTC	19	55 °C	(Ford et al. 2009)
*trn*H-*psb*A	*trn*Hf_05	CGCGCATGGTGGATTCACAATCC	23	46–52 °C	(Tate and Simpson 2003)
	*psb*A3 f	GTTATGCATGAACGTAATGCTC	22	53 °C	(Sang et al. 1997)

sequenced and the sequencing of allopolypoid or autopolyploid genome is difficult because of additional haplotypes. Furthermore, heterozygosity is also considered in addition to genome size. While assembling short gun sequences, the graph structure is complicated because heterozygous sequences and phasing haplotype become difficult (Li and Harkess 2018).

The first-ever non-Sanger platform on the genome scale was hybridization-based re-sequencing *next-generation sequencing (NGS) in 2007 was introduced that was capable of sequencing single molecules without previous amplification (Türktaş et al. 2015). The advantage of NGS is that (1) it utilizes different chemistries for sequences and detection; (2) Its throughput and cost are less than the Sanger technology. Lots have been changed since the first genome project as it all started from a single wild mustard plant. With the passage of time the researchers have adopted new algorithms and technologies such as* Roche/454 Life Sciences sequencing, ABI/SOLiD sequencing, Solexa/Illumina sequencing, and Long-Read Sequencing. The NGS method uses sequencing by synthesis (SBS) method, which means massively parallel sequencing (MPS) technique where fluorescent dye and terminator caps are used together. Using MPS, a large number of DNA strands can be sequenced simultaneously. Hence, proving it to be less time-consuming and more sensitive than previous sequencing techniques.

Database

Currently, the Consortium for Barcoding of Life (CBOL) is an international initiative that is devoted to the development of DNA barcoding as a global standard for species identification. CBOL works toward bridging our knowledge by generating an online database for different species. CBOL, along with major genomics repositories, e.g., National Center for Biotechnology Information (NCBI), biodiversity organizations, e.g., Global Biodiversity Information Facility (GBIF), major barcoding centers, and multiple taxonomic communities started a project to generate barcodes for all species.

This project is known as The Barcode of Life Data System (BOLD). BOLD currently includes databases for 11,926,218 animals, 480,917 plants, and 173,011 fungi species. The current format for data submission to BOLD is composed of five levels of voucher specimen characterization: (1) the specimen identifier, i.e., catalog, collection codes, and the institution responsible for providing the specimen samples; (2) the taxonomic status; (3) the specimen characteristics, i.e., sex, life stage, and reproduction; (4) the collection data, i.e., collector, collection date, and location with GPS coordinators; (5) the DNA barcode sequence detail, i.e., gene name and location, trace file, alignment details, and primers used to generate the amplicon (http://www.boldsystems.org 2018). Many barcodes such as *rbc*L, *mat*K, and ITS are often used for genomorphic identification of plant species.

Before subjecting to match in the BOLD, the DNA sequences of the unknown samples are edited with software like BioEdit and then, the sequence alignment is carried out using MEGA software. The aligned sequence is stored in *FASTA format, which can be uploaded in the database and is utilized for identification/

comparison purpose. The sequence is written in an output file, after that the word size/length is matched with the initial data available. At last, reward/penalty for a nucleotide match or mismatch is assigned. The final output comprises the species which is identified accurately present at the top match static along with few more information like, the lowest e-value, number of hits one can expect to see, the highest bit score, sequence similarity measurement, query sequence length, the highest percentage identity, and finally percentage of similarity between the known and unknown sequence. The accurate identification is only considered; if it qualifies the reproducibility of the results (Meiklejohn et al. 2019). In BOLD, *rbc*L, and *mat*K have >95,000 and >70,000 sequences, respectively, consisting of >500 bp length, while ITS has >15,000 sequences consisting of >100 bp size.

A database is a structured collection of records or data that is stored on a computer so that it can be consulted by a program to answer queries, i.e., information generated through known analysis are used to predict the unknown (Berrington 2017). Every data recorded in databases has a label and a value. In case of DNA barcoding databases, sequences of different plant species at different barcode regions such as *rbc*L, *mat*K, etc., are stored, which can be used to identify the plant species in question. Therefore, name of the plant species is a label and the sequence of specific barcode is the value (Tnah et al. 2019). A database is similar to an excel sheet as it contains rows (records) and columns (fields). Each row will represent sequences for different DNA barcodes for a single species, whereas, each column will represent sequences for a single barcode for different plant species. However, there are significant differences too. Firstly, DNA databases have a powerful data-manipulating feature, i.e., they can retrieve, search, match, and identify the added data from the previous recorded data. Secondly, it has the feature of cross-reference from each column or different tables as well. Thirdly, it is important for a database to have unique entries. It means that each entry in the database must be unique in itself for it to be considered as a new entry or else, the database will register it as an old entry (Berrington 2017; Tnah et al. 2019).

However, there are certain problems which are faced when using DNA barcoding in the case of plants. To perform species identification using DNA barcodes, it is imperative to have reference sequences from authentic, reliable, and unadulterated reference materials. In the case of protected and medicinal plants, it is often very difficult to procure such reference samples. In other words, there is limited reference data to compare results (Hebert and Gregory 2005; Hollingsworth et al. 2011). Even in cases where such reference material is available, it might be difficult to include them in databases because of lack of cooperation between agencies. Although, BOLD has solved this to an extent, but work is still to be done so that more and more plant species are included in BOLD. Another issue arises due to the lack of standardized protocols/procedures to generate DNA sequences of uniform quality. It has been observed that more than 22% studies available for species identification have some kind of problem with the methodology used in the studies (Raja et al. 2017). Therefore, the aforementioned aspects must be accounted for to ensure fool-proof usage of DNA barcodes for the plant species identification in forensic casework.

Conclusion

Plant DNA barcoding is still in its development stage. From the first plant genome project till now, a lot has changed, new markers have been developed and new techniques have come into play to proceed with the work more precisely and that too in a less amount of time. In near future, a more exciting advancement is coming that is "metabarcoding" which can be helpful in identifying the species even if the plant materials are found in soil, water, or reef. More or less there has been an agreement on some of the markers to be utilized as standardized DNA barcodes in plant species identification (*rbc*L, *mat*K and ITS2).

It has been in high demand to update the database of plant DNA barcodes and the studies should have the correct and reliable data so that no further complication is faced. Number of organizations are working on finding a single universal plant DNA barcode, and the coming future might hold the answer to all the queries of the present. With the great combination of molecular genetics, current sequencing techniques, and bioinformatics; DNA barcoding will be helpful in solving criminal cases which involve evidentiary material of plant origin.

References

Aboul-Maaty NAF, Oraby HAS (2019) Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. Bull Natl Res Centre 43(1):1–10

Adhikari KM (2012) Poisoning due to accidental ingestion of *Dieffenbachia* plant (Dumb Cane). Indian Pediatr 49:247–248

Babu SP, Chandrika DG, Kulkarni MR (2016) Plant poisoning – an observational study in a tristate region emergency department. Int J Med Sci Public Health 5(10):2158–2161

Bafeel SO, Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Thomas J (2011) Comparative evaluation of PCR success with universal primers of maturase K (matK) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (rbcL) for barcoding of some arid plants. Plant Omics 4(4):195

Bafeel SO, Arif IA, Bakir MA, Al Homaidan AA, Al Farhan AH, Khan, HA (2012) DNA barcoding of arid wild plants using rbcL gene sequences Genetics and Molecular Research 11(3):1934–1941. https://doi.org/10.4238/2012.july.19.12

Berrington J (2017) Databases. Anaesth Intensive Care Med 18(3):155–157. https://doi.org/10.1016/j.mpaic.2016.11.016

Boatwright JS, Maurin O, van der Bank M (2015) Phylogenetic position of Madagascan species of Acacia s. l. and new combinations in Senegalia and Vachellia (Fabaceae, Mimosoideae, Acacieae). Bot J Linn Soc 179:288–294. https://doi.org/10.1111/boj.12320

Bressan EA, Rossi ML, Gerald LT, Figueira A (2014) Extraction of high-quality DNA from ethanol-preserved tropical plant tissues. BMC Res Notes 7(1):1–6

Brown T, Smith A, Elmhurst O (2002) The combined use of pollen and soil analyses in a search and subsequent murder investigation. J Forensic Sci 47(3):614–618. https://doi.org/10.1520/JFS15302J

Bryant MV (2001) Forensic palynology: why it works. TAMU 4352 Palynology laboratory. Texas

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Barrett SC (2011) Discriminating plant species in a local temperate flora using the *rbc*L+ *mat*K DNA barcode. Methods Ecol Evol 2(4):333–340. https://doi.org/10.1111/j.2041-210X.2011.00092.x

Cassidy BG, Gonzales RA, Roberts S (2005) Special section DNA testing in animal forensics. J Wildl Manag 69(4): 1454–1462. http://www.ncbi.nih.gov/Genomes/

CBOL (2009) A DNA barcode for land plants. PNAS 106(31):1–4

Chase MW, Hills HH (1991) Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon 40(2):215–220

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madriñán S, Petersen G, Pedersen N (2007) A proposal for a standardised protocol to barcode all land plants. Taxon 56(2):295–299. https://doi.org/10.1002/tax.562004

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Luo K (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5(1):e8613. https://doi.org/10.1371/journal.pone.0008613

Cheng T, Xu C, Lei L, Li C (2015) Barcoding the kingdom plantae: new PCR primers for ITS regions of plants with improved universality and specificity barcoding the kingdom plantae: new PCR primers for ITS regions of plants with improved universality and specificity (July). https://doi.org/10.1111/1755-0998.12438

Coyle HM, Palmbach T, Juliano N, Ladd C, Lee HC (2003) An overview of DNA methods for the identification and individualization of marijuana. Croat Med J 44(3):315–321

Coyle HM, Lee C, Lin W, Lee HC, Palmbach TM (2005) Forensic botany: using plant evidence to aid in forensic death investigation. Croat Med J 46(4):606

Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences American Journal of Botany 89(1):132–44

Daru BH, Manning JC, Boatwright JS, Maurin O, Maclean N, Schaefer H et al (2013) Molecular and morphological analysis of subfamily Alooideae (Asphodelaceae) and the inclusion of Chortolirion in Aloe molecular and morphological analysis of subfamily Alooideae (*Asphodelaceae*) and the inclusion of Chortolirion in Aloe. Taxon 62(1):62–76. https://doi.org/10.1002/tax.621006

Dayasiri MBKC, Jayamanne SF, Jayasinghe CY (2017) Plant poisoning among children in rural Sri Lanka. Int J Pediatr 1–6. https://doi.org/10.1155/2017/6187487

De Vere N, Rich TCG, Trinder SA, Long C (2015) Dna barcoding for plants. Methods Mol Biol 1245(January):101–118. https://doi.org/10.1007/978-1-4939-1966-6_8

Dhivya S (2012) Studies on DNA barcodes to identify the Indian medicinal plant species of Apocynaceae and Zingiberaceae. (Thesis)

Doyle JJ, Dickson EE (1987) Preservation of plant samples for DNA restriction endonuclease analysis. Taxon 36(4):715–722

Fay MF, Swensen SM, Chase MW (1997) Taxonomic affinities of *Medusagyneoppositifolia* (*Medusagynaceae*). Kew Bull 52:111–120

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC et al (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3(7). https://doi.org/10.1371/journal.pone.0002802

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, Barrett SCH et al (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol Ecol Resour 9(SUPPL. 1):130–139. https://doi.org/10.1111/j.1755-0998.2009.02652.x

Ferri G, Alù M, Corradini B, Licata M, Beduschi G (2009) Species identification through DNA "barcodes". Genet Test Mol Biomarkers 13(3):421–426. https://doi.org/10.1089/gtmb.2008.0144

Ford CS, Ayres KL, Toomey N, Haider N, Stahl JVANA, Kelly LJ et al (2009) Selection of candidate coding DNA barcoding regions for use on land plants. Bot J Linn Soc 159:1–11. https://doi.org/10.1111/j.1095-8339.2008.00938.x

Frances F, Castello A, Verdu F (2010) Testing the recoverability of grass DNA transferred textiles for forensic purpose. Maejo Int J Sci Technol 4(3):462–467

Frézal L, Leblois R (2008) Four years of DNA barcoding: current advances and prospects. Infect Genet Evolution 8:727–736. https://doi.org/10.1016/j.meegid.2008.05.005

Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Chen S (2010) Identification of medicinal plants in the family *Fabaceae* using a potential DNA barcode ITS2. J Ethnopharmacol 130(1):116–121. https://doi.org/10.1016/j.jep.2010.04.026

Gielly L, Taberlet P (1994) The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. Mol Biol Evol 11(5):769–777. https://doi.org/10.1093/oxfordjournals.molbev.a040157

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D (2002) A draft sequence of the rice genome. Science 296:92–100. https://doi.org/10.1126/science.1068275

Gravensteen IK, Ekeberg Ø, Thiblin I, Helweg-Larsen K, Hem E, Rogde S, Tøllefsen IM (2019) Psychoactive substances in natural and unnatural deaths in Norway and Sweden–a study on victims of suicide and accidents compared with natural deaths in psychiatric patients. BMC psychiatry 19(1):1–9. https://doi.org/10.1186/s12888-019-2015-9

Grebenstein B, Röser M, Sauer W, Hemleben V (1998) Molecular phylogenetic relationships inAveneae (Poaceae) species and other grasses as inferred from ITS1 and ITS2 rDNA sequences. Plant Syst Evol 213(3–4):233–250

Greilhuber J, Borsch T, Müller K, Worberg A, Porembski S, Barthlott W (2006) Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. Plant Biol (Stuttgart, Germany) 8:770–777

Han J, Zhu Y, Chen X, Liao B, Yao H, Song J, Meng F (2013) The short ITS2 sequence serves as an efficient taxonomic sequence tag in comparison with the full-length ITS. Biomed Res Int. https://doi.org/10.1155/2013/741476

Hebert PDN, Gregory TR (2005) The promise of DNA barcoding for taxonomy. Syst Biol 54(5):852–859. https://doi.org/10.1080/10635150500354886

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc B Biol Sci 270(1512):313–321. https://doi.org/10.1098/rspb.2002.2218

Hidalgo O, Pellicer J, Christenhusz M, Schneider H, Leitch AR, Leitch IJ (2017) Is there an upper limit to genome size? Trends Plant Sci 22:567–573

Hilu KW, Alice LA, Liang H (1999) Phylogeny of Poaceae inferred from matK sequences. Ann MO Bot Gard 86:835–851

Hollingsworth P (2008) DNA barcoding plants in biodiversity hot spots: progress and outstanding questions. Heredity 101:1–2. https://doi.org/10.1038/hdy.2008.16

Hollingsworth ML, Andra Clark A, Forrest LL, Richardson J, Pennington RT, Long DG et al (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour 9(2):439–457. https://doi.org/10.1111/j.1755-0998.2008.02439.x

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode PloS one 26;6(5):e19254. https://doi.org/10.1371/journal.pone.0019254

Jaffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatllite" regions in human DNA. Nature 314:67–73. https://doi.org/10.1038/314067a0

Jobling MA, Gill P, Le L, Court T, Park BB (2004) Encoded evidence: DNA in forensic analysis. Nat Rev 5(October). https://doi.org/10.1038/nrg1455

Kilpatrick CW (2002) Noncryogenic preservation of mammalian tissues for DNA extraction: an assessment of storage methods. Biochem Genet 40:53–62

Kraemer M, Boehmer A, Madea B, Maas A (2019) Death cases involving certain new psychoactive substances: a review of the literature. Forensic science international 298(1):186–267. https://doi.org/10.1016/j.forsciint.2019.02.021

Kress WJ (2017) Plant DNA barcodes: applications today and in the future. J Syst Evol 55:291–307. https://doi.org/10.1111/jse.12254

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One 2(6). https://doi.org/10.1371/journal.pone.0000508

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A 102(23):8369–8374. https://doi.org/10.1073/pnas.0503123102

Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc Natl Acad Sci U S A 106(44):18621–18626. https://doi.org/10.1073/pnas.0909820106

Kyalangalilwa B, Boatwright JS, Daru BH, Maurin O, van der Bank M (2013) Phylogenetic position and revised classification of Acacia s. l. *(Fabaceae: Mimosoideae*) in Africa, including new combinations in Vachellia and Senegalia. Bot J Linn Soc 172:500–523. https://doi.org/10.1111/boj.12047

Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G et al (2008) DNA barcoding the floras of biodiversity hotspots (February). https://doi.org/10.1073/pnas.0709936105

Li FW, Harkess A (2018) A guide to sequence your favorite plant genomes. Applications in Plant Sciences. 6(3):e1030. https://doi.org/10.1002/aps3.1030

Liston A, Rieseberg LD, Adams RP, Do M, Zhu G (1990) A method for collecting dried plant specimens for DNA and isozyme analyses and the results of a field test in Xinjiang. China Ann MO Bot Gard 77:859–863. https://doi.org/10.2307/2399681

Livak KJ(1997). ABI Prism 7700 Sequence detection System User Bulletin #2 Relative quantification of gene expression. http://docs.appliedbiosystems.com/pebiodocs/04303859.pdf

Manning J, Boatwright JS, Daru BH, Maurin O, Bank, M. Van Der (2014) A molecular phylogeny and generic classification of Asphodelaceae subfamily Alooideae: a final resolution of the prickly issue of Polyphyly in the Alooids? Syst Bot 39(1):55–74. https://doi.org/10.1600/036364414X678044

Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M (2011) A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. Food Res Int 44(3):693–702. https://doi.org/10.1016/j.foodres.2010.12.032

Maurin O, Muasya AM, Catalan P, Shongwe EZ, Viruel J, Wilkin P, van der Bank M (2016) Diversification into novel habitats in the Africa clade of Dioscorea (*Dioscoreaceae*): erect habit and elephant' s foot tubers. BMC Evol Biol. https://doi.org/10.1186/s12862-016-0812-z

Meiklejohn KA, Damaso N, Robertson JM (2019) Assessment of BOLD and GenBank–their accuracy and reliability for the identification of biological materials. PLoS One 14(6):e0217084

Min XJ, Hickey DA (2007) Assessing the effect of varying sequence length on DNA barcoding of fungi. Mol Ecol Notes 7(3):365–373. https://doi.org/10.1111/j.1471-8286.2007.01698.x

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

National Survey of Drug Use and Health (NSDUH) Releases (2017) National Survey on Drug Use and Health. https://www.samhsa.gov/data/release/2017-national-survey-drug-use-and-health-nsduh-releases

Newmaster S, Fazekas AJ, Steeves AD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour 8:480–490. https://doi.org/10.1111/j.1471-8286.2007.02002.x

Piredda R, Simeone MC, Attimonelli M, Bellarosa R, Schirone B (2011) Prospects of barcoding the Italian wild dendroflora: oaks reveal severe limitations to tracking species identity. Mol Ecol Resour 11(1):72–83. https://doi.org/10.1111/j.1755-0998.2010.02900.x

Raja HA, Miller AN, Pearce CJ, Oberlies NH (2017) Fungal identification using molecular tools: a primer for the natural products research community. J Nat Prod 80:756–770. https://doi.org/10.1021/acs.jnatprod.6b01085

Renner SS (1999) Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. Am J Bot 86(9):1301–1315. https://doi.org/10.2307/2656778

Reusch TB (2006) Does disturbance enhance genotypic diversity in clonal organisms? A field test in the marine angiosperm Zostera marina. Mol Ecol 15:277–286

Ristaino JB, Groves CT, Parra GR (2001) PCR amplification of the Irish potato famine pathogen from historic specimens. Nature 411(6838):695–697. https://doi.org/10.1038/35079606

Rogers SO, Bendich AJ (1994) Extraction of total cellular DNA from plants, algae and fungi. In: Plant molecular biology manual. pp 183–190. https://doi.org/10.1007/978-94-011-0511-8_12

Saadullah KZ, Ashfaq M, Zaib u Nisa (2016) Identification of the grass family (Poaceae) by using the plant DNA barcodes *rbc*L and *mat*K. J Biodivers Environ Sci 8:175–186

Salazar GA, Chase MW, Soto Arenas MA, Ingrouille M (2003) Phylogenetics of *Cranichideae* with emphasis on *Spiranthinae (Orchidaceae, Orchidoideae)*: evidence from plastid and nuclear DNA sequences. Am J Bot 90(5):777–795. https://doi.org/10.3732/ajb.90.5.777

Sang T, Crawford D, Stuessy T (1997) Chloroplast DNA phylogeny, reticulate evolution and biogeography of Paeonia (*Paeoniaceae)*. Am J Bot 84:1120–1136. https://doi.org/10.2307/2446155

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the national academy of sciences 1;74(12):5463–7

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. PNAS 109(16):6241–6246. https://doi.org/10.1073/pnas.1117018109

Selvaraj D, Sarma RK, Sathishkumar R (2008) Phylogenetic analysis of chloroplast *mat*K gene from *Zingiberaceae* for plant DNA barcoding. Bioinformation 3(1):24. https://doi.org/10.6026/97320630003024

Soltis DE, Soltis PS, Mort ME, Chase MW, Savolainen V, Hoot SB, Morton CM (1998) Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. Syst Biol 47(1):32–42. https://doi.org/10.1080/106351598261012

Stoeckle M (2003) Taxonomy, DNA and the barcode of life. Bioscience 53(9):2–3. https://doi.org/10.1641/00063568(2003)053[0796:TDATBC]2.0.CO;2

Tahir A, Hussain F, Ahmed N, Ghorbani A, Jamil A (2018) Assessing universality of DNA barcoding in geographically isolated selected desert medicinal species of Fabaceae and Poaceae. PeerJ 6:e4499. https://doi.org/10.7717/peerj.4499

Tate JA, Simpson BB (2003) Paraphyly of Tarasa (Malvaceae) and diverse origins of thepolyploid species. Syst Bot 28:723–737. https://doi.org/10.1043/02-64.1

Tnah LH, Lee SL, Tan AL, Lee CT, Ng KKS, Ng CH, Farhanah ZN (2019) DNA barcode database of common herbal plants in the tropics: a resource for herbal product authentication. Food Control 95:318–326. https://doi.org/10.1016/j.foodcont.2018.08.022

Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-Coverage ITS Primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PLoS One 7(7). https://doi.org/10.1371/journal.pone.0040863

Türktaş M, Kurtoğlu KY, Dorado G (2015) Sequencing of plant genomes – a review. Turk J Agric For 39:361–376. https://doi.org/10.3906/tar-1409-93

Wang W, Wu Y, Yan Y, Ermakova M, Kerstetter R, Messing J (2010) DNA barcoding of the Lemnaceae, a family of aquatic monocots. BMC Plant Biol 10:205. https://doi.org/10.1186/1471-2229-10-205

Ward J, Gilmore SR, Robertson J, Peakall R (2009) A grass molecular identification system for forensic botany: a critical evaluation of the strengths and limitations. J Forensic Sci 54(6):1254–1260. https://doi.org/10.1111/j.1556-4029.2009.01196.x

Wehausen JD, Ramey RR, Iie CW (2004) Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. J Hered 95:503–509

Wittzell H (1999) Chloroplast DNA variation and reticulate evolution in sexual and apomictic sections of dandelions. Mol Ecol 8(12):2023–2035. https://doi.org/10.1046/j.1365-294x.1999.00807.x

World Drug Report (2018) (United Nations publication, Sales No. E.18.XI.9)

Yao H, Song J, Liu C, Luo K, Han J, Li Y, Chen S (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS One 5(10):e13102. https://doi.org/10.1371/journal.pone.0013102

Zurawski G, Perrot B, Bottomley W, Whitfeld PR (1981) The structure of the gene for the large subunit of ribulose 1, 5-bisphosphate carboxylase from spinach chloroplast DNA. Nucleic Acids Res 9(14):3251–3270. https://doi.org/10.1093/nar/9.14.3251

Molecular Techniques in Microbial Forensics

42

Neeti Kapoor, Pradnya Sulke, and Ashish Badiye

Contents

Abstract

Microbial forensics is a field that has attracted tremendous interest off late. Bioterrorism, biowarfare, and weaponization of the microbiome have terrified the world. While microbial forensics deals with the study of microbes for legal purposes, its applicability in the other subfields of health and science is irrefutable. It has its application not only to bio-crime investigation but also for clinical and toxicology purposes. There are several risks associated with the handling and analysis of microbes. Hence, it is essential to follow the established guidelines to prevent contamination, cross-contamination, and accidental infections with limited or widespread drastic effects. These guidelines mainly ensure the safety,

N. Kapoor · A. Badiye (✉)
Department of Forensic Science, Government Institute of Forensic Science, Nagpur, Maharashtra, India
e-mail: neetikapoor@ifscnagpur.in; ashishbadiye@ifscnagpur.in

P. Sulke
Government Institute of Forensic Science, Nagpur, Maharashtra, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_44

collection, preservation of microbial forensic samples. This chapter covers a broad overview of epidemiology and microbial forensics, the critical elements of microbial forensics, the sample collection methods and guidelines, the various detection methods (molecular), and the result interpretation.

Keywords

Microbial forensics · Microbial evidences · Processing and detection methods · DNA-based methods · Massive parallel sequencing · STR · VNTR · PCR

Introduction

Microbiology refers to "the study of microorganisms, i.e., the organisms that exist as single cells or cell clusters and must be viewed individually with the aid of a microscope" (Nema 2018). These diverse communities of microbes are manipulated by humans and are used as biological warfare agents. With the increase in tools and technologies to manipulate these organisms, their use or abuse has also increased. Thus, the need to investigate such acts has also increased; herein, forensic microbiology or microbial forensics plays a vital role.

Microbial forensics is the scientific discipline that analyzes evidence related to bioterrorism and bio-crimes, hoax, or inadvertent microorganism/toxin release for attribution purposes (Budowle et al. 2005b). Microbial forensics involves the characterization of microbial evidence with the help of microbiological methods to determine the source and assist in identifying incidences of bioterrorism, bio-attack, bio-crime, outbreaks and transmission of pathogens, or accidental release of a biological agent or a toxin (Budowle et al. 2003, 2005a; Oliveira and Amorim 2018; Schmedes and Budowle 2019; Smith 2019).

Microbial forensics has its application not only to bio-crime investigation but also for clinical and toxicology purposes. The human skin is believed to have a unique microbiome that may be individualistic to a person. Therefore, microbial forensics also encompasses identifying a person from their leftover microbial traces on the materials they were in contact with. In addition to this, it also has its application in determining the cause of death, helps to investigate drowning cases, toxicological cases, estimation of postmortem interval, etc. (Oliveira and Amorim 2018).

Bioterrorism

Bioterrorism is emerging as a global threat and one of the deadliest means used to cause mass disasters. This affects not only the economic condition but also the workforce of the country. The usage of bioweapon was prevalent for thousands of years. For example, the Romans were known to contaminate the water resource by decaying animal carcasses to harm their enemies. During the war of Kaffa, Tatar soldiers threw diseased bodies on the city's walls to spread plague among enemies.

Even in many wars such as WWI, WWII, French and Indian war, biological weapons were reportedly employed (Schmedes and Budowle 2019). There are other known incidences of potential bio-attacks. One such example is the bubonic plague attack in the middle of the fourteenth century in the Siege of Caffa (Budowle et al. 2005b; Barras and Greub 2014).

Another classic example is the use of plague as a biological weapon by the Japanese during the Sino-Japanese war in the 1930s–1940s (Budowle et al. 2005b; Barras and Greub 2014; The College of Physicians of Philadelphia 2018). *Salmonella* was used to contaminate the salad bars in the Dalles (Török et al. 1997; Schmedes and Budowle 2019). In 1996, Dallas, TX, hospital technician intentionally infected the muffins with *Shigella* and placed them in the eating area. Due to this, 12 people were infected, and 4 were hospitalized (Kolavic et al. 1997; Schmedes and Budowle 2019). In 1993, anthrax produced by bacterium *B. anthracis* became the most potential bioweapon disseminated in Tokyo by the Aum Shinrikyo Japanese cult (Schmedes and Budowle 2019).

The intentional use of pathogens and microorganisms to create biowarfare, the threat of terrorism, and detection of West Nile virus in New York City in 1999 was the major concern of the United States. In 2001, a 63-year-old employee of American Media in Boca Raton, Florida, suffered from fever, deprived sleep, emesis, and confusion. Later he was diagnosed with anthrax. The detection was confirmed by Laboratory Response Network, Department of Health, Florida. Meanwhile, at the commencement of the twentieth century, most cases of inhalation of anthrax in the United States were due to occupational exposure to infected animal skins or products. Thus, an urgent need was felt to improve the capabilities to collect, examine, and investigate the scene of incidence involving potential bio-crime acts (Morse and Budowle 2006).

Letters laced with anthrax (*Bacillus anthracis* spores) were sent through United States Postal Service along the eastern seaboard, to two senators, to news anchor of NBC News, and the *New York Post*, each containing *B. anthracis* spores. These spores infected 22 persons causing 5 deaths and disruptions (Barras and Greub 2014; Schmedes and Budowle 2019). As a result, the use of the US postal system for spreading endospores of *B. anthracis* became a national threat, and the FBI began the investigation.

Over the decades, the number of epidemic outbreaks has been increasing, such as H1N1 virus-swine-origin influenza A 2009, (Smith et al. 2009); *Mycobacterium tuberculosis* (Gardy et al. 2011); *Vibrio cholerae* (Hendriksen et al. 2011); MERS coronavirus, 2012, Saudi Arabia (Assiri et al. 2013); H7N9 virus, 2013, China (Kageyama et al. 2013); *Escherichia coli* O104:H4, 2011, Europe (Grad et al. 2012); Ebola virus 2014, Sierra Leone (Cenciarelli et al. 2015); Zika virus 2015, Brazil (Faria et al. 2016; Oliveira and Amorim 2018); Nipah virus, 2018, Kerala India; Ebola, 2018–2020, the Democratic Republic of the Congo and Uganda; Coronavirus disease 2019 COVID-19, worldwide 2019 to present; and Black fungus, 2021, India (Wikipedia Contributors 2021). The devastating effect of such epidemics, pandemics, and global bio-attacks led to the official launching of microbial forensics (Budowle et al. 2005b; Barras and Greub 2014).

Individualistic Microflora for Person Identification

The massive collection of microbes that reside on or inside a human body is termed as human microbiome. Personal identification is primordially made using DNA and fingerprints with a great degree of accuracy and validity. Nevertheless, some attempts have been made to individualize person based on the unique microbiome each human is believed to have (Fierer et al. 2010; Tridico et al. 2014; Meadow et al. 2014; Lax et al. 2015; Schmedes et al. 2017; Oliveira and Amorim 2018; Robinson et al. 2020). Several different microbial colonies reside in or on different parts of our body. To identify persons, skin microbiome or hair microbiome are potentially used. Oral microflora also has great significance for identification purposes (Robinson et al. 2020). Each individual has a unique skin microbiome composition, which can be collected from the items they came in contact with. Many studies are conducted on establishing the use of skin microbiome for individualization purposes (Fierer et al. 2010; Schmedes et al. 2017). As stated by Sir Edmond Locard in principle of exchange, "when two objects or entities or surfaces come into contact, there is always a mutual exchange of traces."

Similarly, when an individual touches any surface, they transfer their unique microbiome to that surface. This can be principally utilized in criminal identification. When a suspect touches any surface, it leaves traces there. Surfaces such as mobile phone screens, tablet screens, laptop screens, keyboards, mouse, floor, doorknobs, etc., act as an excellent source for harnessing the microflora of the person using it. Similarly, suspect's unique microflora can be extracted from floors, victims clothing, or anything probable of being touched by him. Appropriate examination of the crime scene could assist in getting unique skin microbiome of the suspect (Meadow et al. 2014; Lax et al. 2015; Nema 2018; Oliveira and Amorim 2018; Robinson et al. 2020). Additionally, this can assist in finding geological locations of suspects based on the identification of microbes that are found explicitly in particular areas; however, this study is still in its stage of infancy (Oliveira and Amorim 2018; Robinson et al. 2020).

When individualizing a person using skin microbiome, one should be aware and cautious of the fact that whole human skin is not supposed to have a uniform microflora. The human body is heterogeneous and consists of different substances in or on the body. Likewise, the skin has various compositional changes moving from hair to toes every part having slightly different microflora. Thus, when using skin microflora for identification, the place where it is generated should be looked upon. Hair is a shred of persistent evidence found on several crime scenes. Some approaches also made to use scalp microflora (also found on hair) for individualization (Tridico et al. 2014; Robinson et al. 2020).

Postmortem Examination

PMI estimation: Determination of postmortem interval (PMI) is an integral part of postmortem examination. PMI estimation is predominantly done using sequential changes occurring in cadavers or entomological evidence. Many approaches are now

directed to estimate PMI using microbial techniques (Metcalf et al. 2013; Hauther et al. 2015; Javan et al. 2016; Nema 2018; Oliveira and Amorim 2018; Zhang et al. 2019; Metcalf 2019). It is well known that microbes carry out decomposition, which is one of the postmortem changes in human cadavers. Decomposition occurs in several sequential steps, affecting the colonization of microbes on or in the body. This successional colonization of microbes can be used for the determination of PMI. This can be done using an ecological succession of microbes of skin or in internal organs of the body defined as thanatomicrobiome. Studies have been conducted to estimate PMI based on changing microbial community on the skin (Metcalf et al. 2013). Meanwhile, other approaches are based on the thanatomicrobiome, which harnesses the microbial community of internal organs and natural orifices, including gastrointestinal tracks, intestines, stomach, eyes, vagina, ears, lungs, etc. (Hauther et al. 2015; Javan et al. 2016; Metcalf 2019).

Estimation of PMI using microbiology has great advantages and can also be potentially applied to calculate submersion interval by studying successional colonization of marine microbes with great accuracy. Further, based on statistical regression models, there are also approaches to estimate PMI (Zhang et al. 2019). There are external factors affecting sequential colonization of microbes on the corpse, which includes that till what time after death PMI can be accurately determined, the weather conditions (sunlight, wind, humidity, etc.) the body is present in, whether the body is covered or open or is buried in soil (Metcalf 2019). These factors also need consideration and also standardization through further research.

Other postmortem examination: It includes determination of cause and manner of death. The cause and manner of death can also be potentially determined using microbial community. There are five manners of death, i.e., natural, accidental, suicidal, homicidal, and undetermined. Some studies show a predominance of a specific community of microbes for a particular manner of death, also some studies have shown that the abundance of specific taxa in different internal organs is also affected by the manner of death, which could be positively used to infer the manner of death (Oliveira and Amorim 2018; Robinson et al. 2020). However, the gender, age, and sex can also interfere in the interpretation of results. Thus, this cannot be used as a sole method of determination of cause and manner of death. It would require proper validation to be admissible in the court of law.

There are numerous causes of death. Death can occur due to prolonged infection, drowning, poisoning, etc. Microbial evidence cannot provide the definitive cause of death but has excellent potential in determining the cause of death. In death cases due to drowning, examination of diatoms is the gold standard in forensic investigations (Díaz-Palma et al. 2009; Oliveira and Amorim 2018). Some microbial communities can be utilized to investigate these cases (Oliveira and Amorim 2018). In this order, some studies establish the use of microflora in toxicological cases and hospital-acquired infections (Castle et al. 2017; Oliveira and Amorim 2018). In addition to determining the cause and manner of death, microbiome can also assist in determining the sex of the cadaver (Bell et al. 2018). Some approaches are being made to differentiate the postmortem changes in the body based on sex to be employed potentially for sex determination (Bell et al. 2018).

Detection of Body Fluids

Biological fluids such as saliva, semen, synovial fluid, blood, vaginal fluid, menstrual blood, urine, etc., are common evidence that forensic investigators encounter at crime scenes. Several preliminary and confirmatory tests achieve identification of these fluids. However, the microbial diversity also has its part in the identification of different body fluids (Zou et al. 2016; Hanssen et al. 2018; Oliveira and Amorim 2018; Robinson et al. 2020). Different body fluids have characteristic microbial profiles where some microbial communities are found in abundance and others in traces, thus can be used as bioindicators confirming the presence of a particular body fluid (Hanssen et al. 2018; Oliveira and Amorim 2018; Robinson et al. 2020). For instance, a study conducted on the Han Chinese population for detection of a particular microbial profile of saliva, vaginal fluid, and feces suggested that the presence of *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Bacteroides uniformis*, *Bacteroides thetaiotaomicron* can be used as identification microbes for vaginal fluid and feces (Zou et al. 2016). Similarly, many approaches are being carried out and many more are needed to establish several biomarkers specific to a particular body fluid that can be positively used for fluid identification.

Microflora of Soil and Water for Forensic Use

Soil and water microflora can ultimately be used to identify geological locations as the microbial profile in soil and water changes over a few meters. Soil has its unique microflora, which can act as bioindicators to confirm two soil samples originating from the same site. Microflora of soil traces found on a crime scene, compared with the suspected soil site, can confirm the location (Oliveira and Amorim 2018; Robinson et al. 2020). In the same way, water microflora can also assist in determining geographical locations. Water microflora has its prime role in investigating drowning cases where diatoms are primarily employed for the purpose. The species of diatoms present in water and the composition of diatoms present in the lungs and internal organs of drowned bodies can confirm whether the person was dead or alive at the time of the drowning. Also, if some different diatom species or different microflora are seen in the body's internal organs, it generates the line of doubt that the person was initially killed in some different environment. Other microbes can also be used as biomarker while investigating drowning cases (Levkov et al. 2017). Besides, there are studies conducted to examine the microflora of freshwater and seawater (Kakizaki et al. 2009). This application also needs some advancements and validation so that it can be potentially used to prove or disprove legal questions. Apart from diatoms, some other microbial markers should also be potentially included in such investigation.

Epidemiology and Microbial Forensics

Epidemiology is defined as the manifestation, topographies, and causes of disease among populations. Epidemiologic methods to investigate the infectious outbreak by examining the possible evidence of intentional and criminal behavior as

contributing factors are termed forensic epidemiology (Goodman et al. 2003; Morse et al. 2019). Similar principles of epidemiology were being utilized to investigate the bio-crime, which involved the bioagent in creating a threat (Flowers et al. 2002).

Microbial forensic epidemiological investigation deals with the legal system that involves examining a crime scene, sustaining the chain of custody, validating the methods, and understanding of results and evolution of new methods to investigate bio-crimes. The investigator must use techniques that match the standard legal system, such as Daubert Standard, which will positively defend the cross-examination (Morse et al. 2019). There are important factors that need to be considered while investigating the outburst of contagious diseases, such as:

- Occurrence of the outbreak
- Identifying the population and community at risk
- Mode of transmission and medium used for transmission
- Illustrating the agent(s) responsible for the dissemination of infectious disease

An epidemiologic investigation would try to recognize the contributing agent and its source of disease outbreaks. Various molecular techniques are used to examine, identify, and characterize pathogens involved in deadly bio-crimes. However, in microbial forensics, the investigation is used for legal purposes (Morse et al. 2019). Numerous indicators of outbursts can be identified in the epidemiological investigation of infectious diseases. The below-listed factors are the potential clue to indicate the signs of an epidemic (Morse et al. 2019).

- Disease caused by unknown agents and without explanation of epidemic.
- The presence of uncommon strain and its antibiotic resistance pattern.
- Higher rate of illness and death due to uncommon disease and patients showing failure in responding to the treatment.
- The distribution of uncommon diseases affected by the season and geographical conditions like influenza spreading in the Northern Hemisphere in the summer season.
- Transmission of disease through different mediums such as air, food, water, and aerosol, etc.
- The presence of one or more strains of the disease in one patient and its reason remains unknown.
- The transmission of the disease affecting a significant heterogeneous population.
- The unfamiliar pattern of morbidity among animals is caused by the unexplained agent responsible for causing the same effect in humans.
- The unexplained and uncommon illness and death occurring in humans by the agent responsible for causing illness and deaths in animals.
- Origin of agents of illness from the source having the same genotype.
- Dissemination of uncommon illness to the noninfectious area either, domestic and foreign.
- A large number of senseless deaths and diseases.

The two essential aspects of microbial forensics are determining the reason for the release of pathogens, whether caused intentionally or due to some negligence and studying the application of protocols for monitoring the pathogens to distinguish between the unprompted and destructive spread of pathogens microorganisms.

The bio-agents are classified based on the hazards on humans. This classification was given by the National Institute of Health (NIH) in 2019 (Department of Health and Human Services, National Institutes of Health 2019). In 2013, they classified these microbes into three categories, but in 2019, these were divided into four main categories below in Table 1. In addition, the Centre for Disease Control and Prevention

Table 1 NIH classification of bio-agents (2019)

Risk Group 1 (RG1)	These are the agents that do not have any impact on humans	*Bacillus subtilis, Bacillus licheniformis, Escherichia coli*, etc.
Risk Group 2 (RG2)	These are the agents associated with humans but are not severe and preventive, and therapeutic interventions are available	*Actinobacillus, Bacillus anthracis, Clostridium botulinum, Dermatophilus congolensis, Erysipelothrix rhusiopathiae, Francisella tularensis, Haemophilus ducreyi, H. influenzae, Leptospira interrogans, Mycobacterium, Pseudomonas aeruginosa, Staphylococcus aureus* Fungal agents such as *Blastomyces dermatitidis, Cladosporium bantianum, C. (Xylohypha) trichoides, Cryptococcus neoformans, Dactylaria galopava*, etc. Parasitic agents such as *Entamoeba histolytica, Enterobius, Fasciola* including *F. gigantica, F. hepatica, Giardia* including *G. lamblia, Heterophyes, Hymenolepis*, etc. Viruses such as *Alphavirus, Adenovirus, Coronavirus, Arenoviruses, Hepatitis A, B, C, D, and E*, etc.
Risk Group 3 (RG3)	These are associated with serious or lethal diseases for which preventive or therapeutic interventions may be available	*Bartonella, Yersinia pestis, Coccidioides immitis, Orientia tsutsugamushi*, Bunyaviruses, *Corona virus, Rhabdovirus*, Retroviruses, Orthomyxoviruses, etc.
Risk Group 4 (RG4)	These are the reagents associated with serious or lethal diseases in humans for which the preventive or therapeutic interventions are usually not available	*Arenavirus, Bunyaviruses, Ebola virus, Herpesviruses, Equine Morbillivirus, Herpesvirus simiae*, etc.

Table 2 CDC classification of biological agents (2018)

Category A	These are the highest priority agents Can be quickly disseminated or transmitted Results in a high mortality rate and have the potential for public health impact May cause public panic and social disruption and may require special action for public health preparedness	*Bacillus anthracis, Clostridium botulinum toxin, Yersinia pestis, Variola major, Francisella tularensis, Filoviruses, Arenaviruses*, etc.
Category B	These are the second highest priority agents Can be moderately easily disseminated or transmitted Result in moderate morbidity rate and low mortality rate Requires specific enhanced surveillance	*Brucella, Clostridium perfringens, Escherichia coli, Burkholderia mallei, Burkholderia pseudomallei, Chlamydia psittaci, Coxiella burnetiid, Ricinus communis, Staphylococcal enterotoxin B, Rickettsia prowazekii, alphaviruses, Vibrio cholerae, Cryptosporidium parvum*, etc.
Category C	Third highest priority agents Availability Less easily disseminated as category A and B Potential for high morbidity and mortality rates and significant health impact	Emerging Nipah virus and hantavirus

(CDC) classifies these biological agents into three categories: Category A, B, and C, given in Table 2 (Centres for Disease Control and Prevention 2018).

The Elements

Microbial forensics adopt genomic, microbiological, and epidemiological methods to characterize and determine biowarfare weapons and ascertain the intentional or unintentional release of destructive pathogens and toxins (Morse et al. 2019). Microbial forensics is an emerging field dedicated to the depiction, examination, and elucidation of evidence found during the act of biological terrorism, biowarfare, and involuntary release of endospores and microorganisms for the attribution purpose (Murch 2003; Morse and Budowle 2006).

The significant elements include:

(i) **Detection and identification:** This is the first and the foremost step that incorporates detection and identification of the attack and the causing microbe behind it. For this, there should be proper collection and preservation of samples for analysis. Furthermore, for accurate analysis, powerful tools and techniques are needed to improve the sensitivity, accuracy, and specificity of

results. These techniques can be categorized into three broad groups that are: (a) molecular techniques, (b) analytical techniques, and (c) physical analysis techniques (Budowle et al. 2005b; Varun et al. 2012).

(ii) **Information and database:** The availability of information and databases, of course, increases the accuracy of results. Thus, the available database should be enhanced and expanded such that they contain the bioagent genomic sequence data, the whole genome of the agent used in the past, etc. This can be achieved by establishing the record systems at the national and international levels (Budowle et al. 2005b). In addition, interagency sharing of information and database must be encouraged.

(iii) **Development of strain repository:** It is imperative to house these pathogens and near neighboring microorganisms in a strain repository. This information may prove essential in determining and identifying these near neighbors' broad and narrow classes using different robust and sophisticated techniques (Budowle et al. 2005b). However, the security of such a repository would remain an active concern as any potential lapse may have devastating outcomes.

(iv) **Need for validation:** This emphasizes the need for proper validation of new and existing techniques used in microbial forensics. All the methods used to analyze microbial samples must be accepted and validated (Budowle et al. 2005b; Varun et al. 2012). Furthermore, they must be robust enough to provide accurate results even in mutations and changes in the known strain.

(v) **Quality assurance guidelines:** Safety and quality assurance must be practiced using the appropriate guidelines and norms. These guidelines are important to be followed by microbial forensic laboratories to guarantee reliable results and maintain safety during and after the analysis (Budowle et al. 2005b; Varun et al. 2012).

Sample Collection

Given the exchange principle, every contact leaves a trace; an investigator needs to collect evidence adequately; otherwise, it will lose its evidential value. Therefore, to avoid human error, the National Institute of Justice issued specific guidelines to minimize the chance of error in collecting microbial evidence. The guidelines mainly ensure the safety, collection, and preservation of microbial forensic samples. The main postulates of guidelines are mentioned below:

- Valuation of the actual situation at a crime scene.
- Planning related to sample collection, which includes assessing safety protocols for personnel, acquiescence with all guidelines and legal requirements, and discussing the prioritization of samples. It also mentioned selecting personnel and equipment utilized to collect and preserve samples with a proper time frame.

- Documentation of place, area, subjects, whether human and animal. The possible and source need to find out and maintain the proper chain of custody of possession of evidence.
- Mention the proper method and equipment that will utilize to collect the microbial forensic evidence.
- Method of proper preservation of samples (Smith 2019).

The techniques for collecting microbial forensic evidence involves strategic planning, logistic support, and statistic data in collecting microbes that will not produce a toxic effect on other microorganisms. Some principles and guidelines need to be followed to properly handle, collect, and preserve microbial forensic evidence. Tools used to collect the microbial evidence should be validated first and should not react with the sample of interest (Schutzer et al. 2011). The preservation procedure is also based upon the same principle that it will not affect the targeted sample. After the preservation process, samples are packaged and sent for analysis to the laboratory (Schutzer et al. 2011). This is the primary step of the analysis as the quality of the result may vary if the sample is not collected and preserved well. In addition, contamination and cross-contamination should be avoided. The sampling process includes collection and preservation, but several other steps need to be followed given in Fig. 1 (National Research Council 2014; Smith 2020).

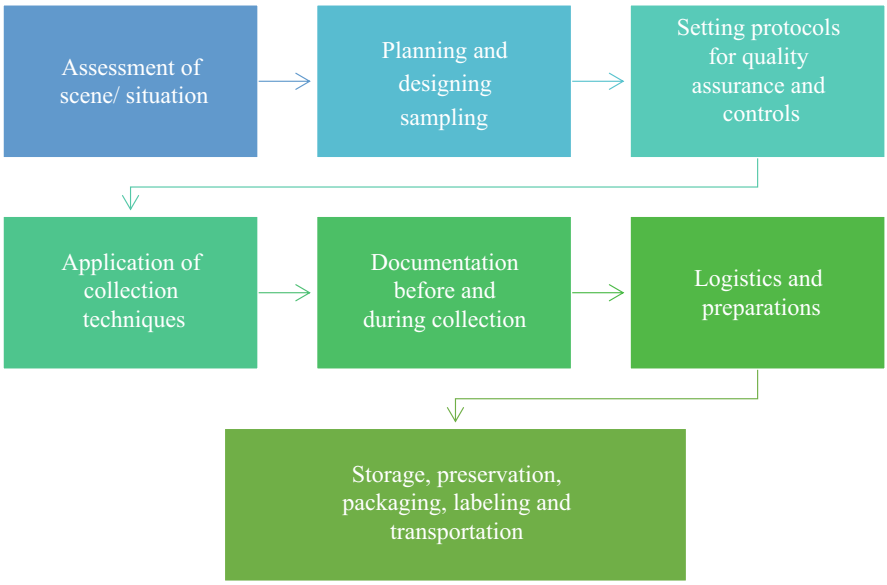


Fig. 1 Steps in collection and preservation of sample

Sampling can be achieved by one of the two strategies mentioned below (National Research Council 2014; Smith 2020). Choosing the sampling strategy varies depending on the location, type, and the extent of symptoms in clinical and agricultural settings (Budowle et al. 2006)

(i) *Targeted sampling:* This is the sampling procedure where a sample is collected from a targeted area. In this process, the sample is collected from an area that is believed to be contaminated based on past knowledge. It is also known as judgmental sampling (Budowle et al. 2006; Sego et al. 2007; Smith 2020).
(ii) *Random sampling:* This is the type of sampling in which the sample is collected from random areas without any prior knowledge (Budowle et al. 2006; Sego et al. 2007; Smith 2020).

Actual sample collection is the process of taking the sample for analysis. Sample collection can be carried out using three approaches that are:

(i) *Collecting the whole item (is also termed as a bulk collection)*: In this, the whole item is collected and transported to the laboratory for analysis. This approach reduces the extra time required for the collection, but this can only be used if the object or the evidence can be easily removed from the scene (Budowle et al. 2006; Smith 2020).
(ii) *Collecting a portion of an item*: This applies to immovable objects that cannot be transported to the laboratory. This includes methods such as vacuuming using high-efficiency particulate air vacuums, filtration, etc.
(iii) *Swabbing or wiping surfaces*: This approach is best suited for trace pieces of evidence. It can be achieved using relevant sample collection devices. For this purpose, dry swabs, premoistened swabs or wipes can be employed (Budowle et al. 2006; Smith 2020).

There are three main issues with these collection methods: (a) many of these collection methods and devices are not rigorously validated; (b) some of the methods are validated, but some security restrictions hamper sharing of this validation data to authorities who need them; (c) the collector must be well acquainted with the analyte or target signatures that are to be analyzed and accordingly the method should be chosen.

Once the evidence is collected, they are packed appropriately and labelled along with the tags indicating the biohazard materials. The preservation of these samples can be achieved using preservative or transport media such as buffered tryptose broth, buffered glycerine, phosphate-buffered sucrose, etc. (Budowle et al. 2006). These media should be chosen based on the type of pathogen to not interfere with the analysis. In some cases, postmortem sampling is also required. Table 3 shows the viscera sample and quantity to be collected related to some common pathogens (Fernández-Rodríguez et al. 2019).

Table 3 Postmortem microbial samples to be collected for forensic analysis

Pathogen suspected	Sample	Quantity
Mandatory specimen	Blood Serum Spleen	5–10 ml 3–5 ml >1–2 cm^3
Pneumonia and other respiratory infections	A portion of the affected lung part Urine Pleural exudate Swabbing the affected mucosa	>1–2 cm^3 3–5 ml >1–2 cm^3 Two swabs
Flu/viral respiratory infection	Nasopharyngeal swab	Two swabs
Invasive fungal infection	Lung, heart, brain, kidneys, large intestine, cerebrospinal fluid (CSF)	
Malaria and other parasitoses	Brain, liver, lung, myocardium, blood	>1–2 cm^3
Botulism	Feces Exudates Tissues	>/= 1 ml Two swabs >1–2 cm^3

Detection Methods

Microbial forensic evidence encompasses various samples such as food, water, air, swab, soil, animal food, tissue, and clinical samples like blood, urine, stool, tissue, sputum and saliva, etc. The microbial samples are analyzed using various methods such as culture, microscopy, immunoassays, mass spectrometry, real-time PCR, microarray, genetic typing, whole-genome sequencing, and targeted sequencing. Meanwhile, the culturing method is considered the gold standard, but sometimes there is a delay in growth, and it may compromise the safety of an individual. The culture method also suffers problems when dealing with novel and uncharacterized microorganisms (Schmedes and Budowle 2019). Microbial analysis can be carried out using three methodologies: (a) molecular methods, (b) analytical methods, and (c) physical analysis (Budowle et al. 2005b; Varun et al. 2012).

Molecular techniques can be either protein-based, such as microarray assays, immunoassays, etc., or DNA-based techniques (Fig. 2) such as PCR, SNP, etc. (Budowle et al. 2005a; National Research Council 2014; Nema 2018; Blondeau et al. 2019; Kieser and Budowle 2020). **Analytical techniques** include the use of various instrumental techniques such as matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF), gas chromatography-mass spectroscopy (GC-MS), and liquid chromatography-mass spectroscopy (LC-MS) (National Research Council 2014; Nema 2018; Blondeau et al. 2019). **The physical analysis** includes microscopic analysis of materials such as soil, water, etc. (National Research Council 2014; Nema 2018; Schmedes and Budowle 2019).

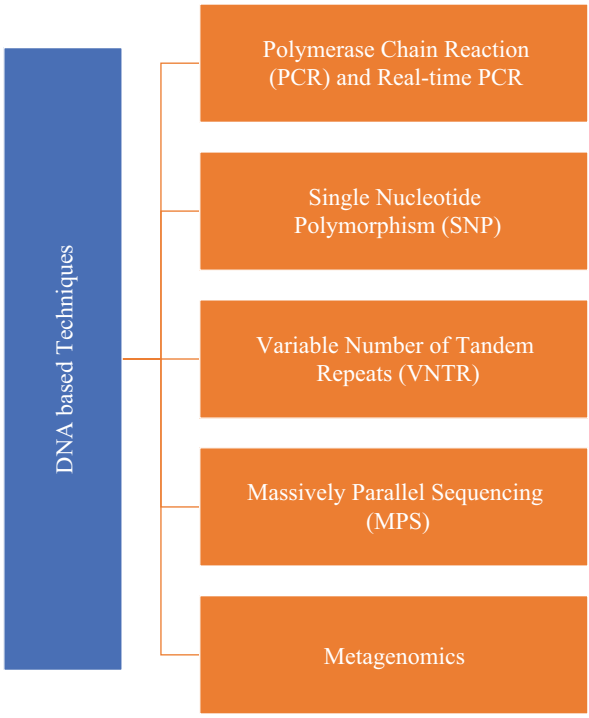


Fig. 2 DNA-based techniques used in microbial forensics

Polymerase chain reaction (PCR) and real-time PCR: PCR-based techniques are the easiest to perform and require a minimal sample quantity. This includes in vitro amplification of DNA carried out in a specified instrument (Budowle et al. 2005a). The instrument performs programable cycling at controlled temperatures to generate millions of copies of targeted DNA that can be detected using different techniques such as hybridization or electrophoresis (Budowle et al. 2005a). These assays can screen many samples with a specific target DNA sequence (Budowle et al. 2005a). Real-time PCR (RT-PCR) is employed for those microbes having RNA as genetic material; this is because PCR can only be proceeded on DNA, and therefore in RT-PCR, reverse transcription is carried out to generate DNA from RNA to procced with PCR. With real-time PCR, amplification and detection of a specific variant of a microbe can be done simultaneously. This is generally done using fluorescent chemistries (Budowle et al. 2005a). Pyrosequencing PCR can also be helpful in many ways. This process is based on detecting luminescence by releasing pyrophosphate on nucleotide addition into the strand (Kieser and Budowle 2020).

PCR and RT-PCR show tremendous significance in forensic microbiology (Bauer et al. 1999; Power et al. 2010; Wang et al. 2013; Park et al. 2014; Rajalakshmi 2017; Aqeel and Omran 2018; Jung et al. 2018). There are numerous species for which PCR markers are readily available, i.e., *Lactobacillus*, *Gardnerella vaginalis*,

Mycoplasma hominis, etc. (Rajalakshmi 2017). Some studies have positively used PCR markers for the identification of fluids such as blood, saliva, menstrual blood, other vaginal secretions, etc. (Bauer et al. 1999; Wang et al. 2013). Studies have been carried out on messenger RNA (mRNA) profiling using PCR and RTPCR. SPTB, PBGD, HBB, HBA, ALAS2, CD3G, ANK1, PBGD, SPTB, AQP9 can be potentially used as RT-PCR mRNA markers for identification of blood form dried as well as wet stains (Bauer et al. 1999; Wang et al. 2013). Some other attempts are also made using streptococcal bacteria present in saliva as a marker to identify expirated blood (Power et al. 2010). Similarly, matrix metalloproteinase (MMP) and protamine mRNA markers can be used to detect menstrual blood (Wang et al. 2013). One study used oral bacteria *Streptococcus salivarius*, *Streptococcus sanguinis*, and *Neisseria subflava* to identify saliva (Jung et al. 2018). Other fluids such as semen, vaginal fluid, saliva also have such identification mRNA biomarkers (Wang et al. 2013). PCR and RT-PCR markers are also available for the skin microbiome (Hanson et al. 2012; Aqeel and Omran 2018). These are some standard PCR and RT-PCR markers that can be potentially used in microbial forensics. Other than those mentioned here, there are several other PCR and RT-PCR markers available for different species as well.

Single nucleotide polymorphism (SNP) array: This technique detects the variation at a single DNA site. It is a multiplex analysis of SNP (Budowle et al. 2005a; Schmedes and Budowle 2019; Kieser and Budowle 2020). After the PCR product is achieved, it is subjected to SNP primer and fluorescently labelled terminator nucleotide. The polymerase reaction is ceased by terminator nucleotide. This product is then separated by capillary or slab electrophoresis (Budowle et al. 2005a). This microarray can be a highly efficient screening technique with species to strain-level detection (Schmedes and Budowle 2019).

SNPs (single nucleotide polymorphisms) and other genetic signatures, which can be highly efficient screening and characterization tools, have been used precisely for bacterial and viral detection and can attain species to strain-level identification. Whole-genome shotgun sequencing (WGSS) is based on the sequencing method, requiring any previous sequence information to be determined. It can analyze any number of genetic markers such as SNPs, insertion, duplication, deletion, rearrangement, genetically and engineered genomes. Previously the WGSS used Sanger sequencing but is required to use cloning vector, was time-consuming, had low output, and was more expensive (Sanger et al. 1977). PCR-based assays are one of the less expensive and easy methods to perform being utilized to analyze the strain. Although real-time PCR allows analysis of the microbial sample, this technique is limited to detecting few microbes variants (Schmedes and Budowle 2019).

SNP has the potentials of providing strain-specific information about the bacterial or any microbial genome being examined. A study conducted on *Bacillus anthracis Ames* strain for its detection by SNP concluded that among 88 other *B. anthracis* strains, they successfully found 6 different SNP characteristics to Ames strain, and 5 of them were even capable of differentiating Ames strain from its close isolates. Thus, this shows that SNP can be utilized to gain strain-level information about the microbial community (Van Ert et al. 2007). Similarly, another study on

Mycobacterium tuberculosis using IS*6110* SNP and some other analyses showed that SNP assays have also concluded its use for establishing ancestry in the microbial community (Faksri et al. 2011). While SNP assays having application in providing information about these disease-causing microbes, it also has potential in providing species or strain-specific information in 16s rRNA gene which is the most used gene for forensic analysis (Gu et al. 2017). One other such approach was also made to discover potential SNP of *Cutibacterium acnes* (*Propionibacterium acnes*) 16S rRNA, which is the most used microbial marker from the human skin microbiome to establish a contact of a person. It would assist in identifying different strains of C. acnes and could be incorporated for determining ownership. This resulted in discovering many such SNPs that can be utilized for the defined purpose (Yang et al. 2019). SNPs can be used as informative markers in Microbial forensic.

Variable number of tandem repeats (VNTR): Multilocus VNTR can be performed to analyze polymorphism at the minisatellite region of DNA (Budowle et al. 2005a; National Research Council 2014; Schmedes and Budowle 2019). This polymorphism is unique to a species and can be used to screen samples for a particular species or strain. This technique encompasses amplification of fragments of DNA that differ in size by the number of repeat units present within the sample (Budowle et al. 2005a). These are then separated using electrophoresis and are viewed under fluorescently labelled primer incorporated during PCR. This ensures the separation of pathogen DNA but has limitations in phylogenic isolation (Budowle et al. 2005a).

MLVA (multilocus variable number tandem repeat analysis) is another method to detect polymorphisms found in minisatellite regions found in bacterial genomes. It is found to affect the discrimination of strains of highly monomorphic species such as *B. anthracis* (Schmedes and Budowle 2019). MLVA also has a great species to strain identification capacity as SNP and thus can differentiate distantly related isolates and closely related isolates (Klevytska et al. 2001; Noller et al. 2003; Keim et al. 2008; Thierry et al. 2014). Varied approaches are being made to generate different VNTR markers that can be used to identify microbes. MLVA markers for microbes causing disease outbreaks such as *Bacillus anthracis* and for *Yersinia pestis* are studied. A study on the genome of *B. anthracis* using 31 VNTR loci showed that this technique, combined others, can potentially differentiate different strains of *B. anthracis* (Thierry et al. 2014).

Similarly, many such studies have shown positive discrimination of *B. anthracis* using multiple locus VNTR (Keim et al. 2008). Another study on *Yersinia pestis* using 42 VNTR loci analysis from one chromosomal and two plasmid pMT1 and pCD1 DNA sequences also suggested the potential use of MLVA for differentiation of closely and distantly related isolates (Klevytska et al. 2001). Another similar study on *Escherichia coli* O157:H7 suggested that MLVA is as potent a method as pulsed-field gel electrophoresis (PFGE) (Noller et al. 2003). The studies mentioned here shows that multilocus VNTR analysis can also be positively utilized for microbial analysis.

Massively parallel sequencing (MPS): Massively parallel sequencing (MPS) is another genetic tool available in the hand of a scientist. MPS is the technique used

to analyze gigabase sequences of data in a short period. MPS is the technique in which millions of sequencing reactions can be carried out in a massively parallel way in a single run (National Research Council 2014; Kieser and Budowle 2020). This technique offers a complete characterization of the viral or bacterial genome. Much deeper genetic information can be achieved using this technique (Schmedes and Budowle 2019). It provides a high outturn, culture-independent method for whole genome sequencing (WGS). MPS has been a potent tool to detect and identify various disease outbreaks (Schmedes and Budowle 2019). Also, it provides an advantage that it has no requirement for enriched DNA for sequencing. MPS technique detects SNP and other genetic variants among the selected agents such as *B. anthracis* and *Y. pestis*. This technique can detect and differentiate among four different variants of both the microbial agent in a single sequencing run (Schmedes and Budowle 2019; Wetterstrand 2020).

Metagenomics: Metagenomics involves the application of sequencing genetic material collected from an environmentally generated source such as water (Biers et al. 2009), soil (Mocali and Benedetti 2010), and human linked samples (Huttenhower et al. 2012; Schmedes and Budowle 2019). Metagenomic samples are applied in the determination of various forensic aspects such as to cause of death (Kakizaki et al. 2012), identification of human (Fierer et al. 2010), time since death (Hyde et al. 2013), characterization of biological fluids (Benschop et al. 2012), and pathogenic outburst investigation (Loman et al. 2013). In the forensic investigation, the microbial samples encountered are variable, mixed profile with other organisms, and low quantity samples. Forensic metagenomics is used to analyzed target micro-organisms from complex matrices (Schmedes and Budowle 2019).

These were the commonly used DNA-based techniques that can assist in the investigation of microbial attacks. As a result, these techniques are widely used in microbial forensics.

Interpretation of Results

Interpretation is a crucial step that validates the findings and provides confidence to withstand trial and scrutiny. Forensic science is based upon the comparative study between the questioned and specimen/reference samples. Three types of interpretation are generally acceptable viz., inclusion, exclusion, and inconclusive result. Inclusion is the similarity between the compared samples and shows the exact antecedent beyond a reasonable doubt. Exclusion show dissimilarity between the compared samples and different origin beyond a reasonable doubt. An inconclusive result signifies the insufficient information is obtained to summarize interpretation and reach any specific conclusion. The result should be statistically strengthened by using various tools to validate and concrete scientists' observations and should be able to endure strongly before the legal system. When dealing with microbial data, interpretation of results is critical as it requires high knowledge about sequencing and other techniques. Different techniques have their interpretation guidelines. For instance, in the case of analysis using commercially available RT-PCR kits, the

presence of a particular microbe is or can be confirmed using fluorescence which confirms its presence. This is relatively easy and does not require much knowledge about sequencing data. However, when dealing with more sophisticated techniques such as SNP, VNTR, RFLP, MPS, metagenomics, etc., prior knowledge about sequencing phenomena and other basics of technique is needed to interpret results accurately. While interpreting results, the other thing that should be looked upon is reviewing sampling technique, sample condition, external factors affecting samples such as geographical location, temperature, humidity, etc., affecting microbial community colonizing at the site of colonization. These are how interpretation can be interfered by different factors needed to be kept in hand while interpreting results.

Conclusion

Microbial forensics is the emerging branch of forensic science. It plays a vital role in investigating bio-crimes. The increasing use of microorganisms as war agents has also increased the need to have a body that separately analyses such outbreaks. Microbial forensics deals with the collection, preservation, storage, transport, and analysis of microbial forensic evidence. Forensic analysis of such evidence incorporates many techniques, among which molecular techniques may be helpful for analysis. Different molecular techniques used in forensic microbiology are discussed in the above sections. Further standardization is needed to strengthen the system to make it more robust and enabling high throughput outcomes.

References

Aqeel L, Omran R (2018) Detection of staphylococcal species diversity within human skin microbiome using a simple PCR-SSCP technique. J Univ BABYLON Pure Appl Sci 26:1–6. https://doi.org/10.29196/jubpas.v26i8.1646

Assiri A, McGeer A, Perl TM et al (2013) Hospital outbreak of Middle East respiratory syndrome coronavirus. N Engl J Med 369:407–416. https://doi.org/10.1056/NEJMoa1306742

Barras V, Greub G (2014) History of biological warfare and bioterrorism. Clin Microbiol Infect 20: 497–502. https://doi.org/10.1111/1469-0691.12706

Bauer M, Kraus A, Patzelt D (1999) Detection of epithelial cells in dried blood stains by reverse transcriptase-polymerase chain reaction. J Forensic Sci 44:1232–1236

Bell CR, Wilkinson JE, Robertson BK, Javan GT (2018) Sex-related differences in the thanatomicrobiome in postmortem heart samples using bacterial gene regions V1-2 and V4. Lett Appl Microbiol 67:144–153. https://doi.org/10.1111/lam.13005

Benschop CCG, Quaak FCA, Boon ME et al (2012) Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? Int J Legal Med 126:303–310. https://doi.org/10.1007/s00414-011-0660-8

Biers EJ, Sun S, Howard EC (2009) Prokaryotic genomes and diversity in surface ocean waters: interrogating the global ocean sampling metagenome. Appl Environ Microbiol 75:2221–2229. https://doi.org/10.1128/AEM.02118-08

Blondeau LD, Rubin JE, Deneer H et al (2019) Forensic, investigative and diagnostic microbiology: similar technologies but different priorities. Future Microbiol 14:553–558. https://doi.org/10.2217/fmb-2019-0088

Budowle B, Schutzer SE, Einseln A et al (2003) Public health. Building microbial forensics as a response to bioterrorism. Science 301:1852–1853. https://doi.org/10.1126/science.1090083

Budowle B, Johnson MD, Fraser CM et al (2005a) Genetic analysis and attribution of microbial forensics evidence. Crit Rev Microbiol 31:233–254. https://doi.org/10.1080/10408410500304082

Budowle B, Schutzer S, Breeze R (eds) (2005b) Microbial forensics, 1st edn. Academic Press

Budowle B, Schutzer SE, Burans JP et al (2006) Quality sample collection, handling, and preservation for an effective microbial forensics program. Appl Environ Microbiol 72:6431–6438. https://doi.org/10.1128/AEM.01165-06

Castle JW, Butzbach DM, Walker GS et al (2017) Microbial impacts in postmortem toxicology. In: Forensic microbiology. Wiley, Chichester, pp 212–244

Cenciarelli O, Pietropaoli S, Malizia A et al (2015) Ebola virus disease 2013–2014 outbreak in West Africa: an analysis of the epidemic spread and response. Int J Microbiol 2015:1–12. https://doi.org/10.1155/2015/769121

Centres for Disease Control and Prevention (2018) Bioterrorism agents/diseases. In: Centers Dis. Control Prev. https://emergency.cdc.gov/agent/agentlist.asp. Accessed 10 Jun 2021

Department of Health and Human Services, National Institutes of Health (2019) NIH guidelines for research involving recombinant or synthetic nucleic acid molecules. NIH Guidel 142

Díaz-Palma PA, Alucema A, Hayashida G, Maidana NI (2009) Development and standardization of a microalgae test for determining deaths by drowning. Forensic Sci Int 184:37–41. https://doi.org/10.1016/j.forsciint.2008.11.015

Faksri K, Drobniewski F, Nikolayevskyy V et al (2011) Genetic diversity of the Mycobacterium tuberculosis Beijing family based on IS6110, SNP, LSP and VNTR profiles from Thailand. Infect Genet Evol 11:1142–1149. https://doi.org/10.1016/j.meegid.2011.04.007

Faria NR, Azevedo R d. S d. S, Kraemer MUG et al (2016) Zika virus in the Americas: early epidemiological and genetic findings. Science 352:345–349. https://doi.org/10.1126/science.aaf5036

Fernández-Rodríguez A, Burton JL, Andreoletti L et al (2019) Post-mortem microbiology in sudden death: sampling protocols proposed in different clinical settings. Clin Microbiol Infect 25:570–579. https://doi.org/10.1016/j.cmi.2018.08.009

Fierer N, Lauber CL, Zhou N et al (2010) Forensic identification using skin bacterial communities. Proc Natl Acad Sci U S A 107:6477–6481. https://doi.org/10.1073/pnas.1000162107

Flowers LK, Mothershead JL, Blackwell TH (2002) Bioterrorism preparedness. II: the community and emergency medical services systems. Emerg Med Clin North Am 20:457–476. https://doi.org/10.1016/s0733-8627(01)00009-8

Gardy JL, Johnston JC, Sui SJH et al (2011) Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364:730–739. https://doi.org/10.1056/NEJMoa1003176

Goodman RA, Munson JW, Dammers K et al (2003) Forensic epidemiology: law at the intersection of public health and criminal investigations. J Law Med Ethics 31:684–700. https://doi.org/10.1111/j.1748-720x.2003.tb00135.x

Grad YH, Lipsitch M, Feldgarden M et al (2012) Genomic epidemiology of the Escherichia coli O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci 109:3065–3070. https://doi.org/10.1073/pnas.1121491109

Gu Y, Zha L, Yun L (2017) Potential usefulness of SNP in the 16S rRNA gene serving as informative microbial marker for forensic attribution. Forensic Sci Int Genet Suppl Ser 6:e451–e452. https://doi.org/10.1016/j.fsigss.2017.09.176

Hanson E, Haas C, Jucker R, Ballantyne J (2012) Specific and sensitive mRNA biomarkers for the identification of skin in "touch DNA" evidence. Forensic Sci Int Genet 6:548–558. https://doi.org/10.1016/j.fsigen.2012.01.004

Hanssen EN, Liland KH, Gill P, Snipen L (2018) Optimizing body fluid recognition from microbial taxonomic profiles. Forensic Sci Int Genet 37:13–20. https://doi.org/10.1016/j.fsigen.2018.07.012

Hauther KA, Cobaugh KL, Jantz LM et al (2015) Estimating time since death from Postmortem human gut microbial communities. J Forensic Sci 60:1234–1240. https://doi.org/10.1111/1556-4029.12828

Hendriksen RS, Price LB, Schupp JM et al (2011) Population genetics of vibrio cholerae from Nepal in 2010: evidence on the origin of the Haitian outbreak. MBio 2. https://doi.org/10.1128/mBio.00157-11

Huttenhower C, Gevers D, Knight R et al (2012) Structure, function and diversity of the healthy human microbiome. Nature 486:207–214. https://doi.org/10.1038/nature11234

Hyde ER, Haarmann DP, Lynne AM et al (2013) The living dead: bacterial community structure of a cadaver at the onset and end of the bloat stage of decomposition. PLoS One 8:e77733. https://doi.org/10.1371/journal.pone.0077733

Javan GT, Finley SJ, Can I et al (2016) Human Thanatomicrobiome succession and time since death. Sci Rep 6:29598. https://doi.org/10.1038/srep29598

Jung JY, Yoon HK, An S et al (2018) Rapid oral bacteria detection based on real-time PCR for the forensic identification of saliva. Sci Rep 8:10852. https://doi.org/10.1038/s41598-018-29264-2

Kageyama T, Fujisaki S, Takashita E et al (2013) Genetic analysis of novel avian a(H7N9) influenza viruses isolated from patients in China, February to April 2013. Euro Surveill 18:20453

Kakizaki E, Kozawa S, Sakai M, Yukawa N (2009) Bioluminescent bacteria have potential as a marker of drowning in seawater: two immersed cadavers retrieved near estuaries. Legal Med 11:91–96. https://doi.org/10.1016/j.legalmed.2008.10.004

Kakizaki E, Ogura Y, Kozawa S et al (2012) Detection of diverse aquatic microbes in blood and organs of drowning victims: first metagenomic approach using high-throughput 454-pyrosequencing. Forensic Sci Int 220:135–146. https://doi.org/10.1016/j.forsciint.2012.02.010

Keim P, Pearson T, Okinaka R (2008) Microbial forensics: DNA fingerprinting of bacillus anthracis (anthrax). Anal Chem 80:4791–4799. https://doi.org/10.1021/ac086131g

Kieser RE, Budowle B (2020) Select methods for microbial forensic nucleic acid analysis of trace and uncultivable specimens. In: Microbial forensics. Elsevier, pp 195–205

Klevytska AM, Price LB, Schupp JM et al (2001) Identification and characterization of variable-number tandem repeats in the Yersinia pestis genome. J Clin Microbiol 39:3179–3185. https://doi.org/10.1128/JCM.39.9.3179-3185.2001

Kolavic SA, Kimura A, Simons SL et al (1997) An outbreak of Shigella dysenteriae type 2 among laboratory workers due to intentional food contamination. JAMA 278:396–398

Lax S, Hampton-Marcell JT, Gibbons SM et al (2015) Forensic analysis of the microbiome of phones and shoes. Microbiome 3:21. https://doi.org/10.1186/s40168-015-0082-9

Levkov Z, Williams DM, Nikolovska D et al (2017) The use of diatoms in forensic science: advantages and limitations of the diatom test in cases of drowning. In: The archaeological and forensic applications of microfossils: a deeper understanding of human history. The Geological Society of London on behalf of The Micropalaeontological Society, pp 261–277

Loman NJ, Constantinidou C, Christner M et al (2013) A culture-independent sequence-based Metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104:H4. JAMA 309:1502. https://doi.org/10.1001/jama.2013.3231

Meadow JF, Altrichter AE, Green JL (2014) Mobile phones carry the personal microbiome of their owners. Peer J 2:e447. https://doi.org/10.7717/peerj.447

Metcalf JL (2019) Estimating the postmortem interval using microbes: knowledge gaps and a path to technology adoption. Forensic Sci Int Genet 38:211–218. https://doi.org/10.1016/j.fsigen.2018.11.004

Metcalf JL, Wegener Parfrey L, Gonzalez A et al (2013) A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. Elife 2. https://doi.org/10.7554/eLife.01104

Mocali S, Benedetti A (2010) Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. Res Microbiol 161:497–505. https://doi.org/10.1016/j.resmic.2010.04.010

Morse SA, Budowle B (2006) Microbial forensics: application to bioterrorism preparedness and response. Infect Dis Clin N Am 20:455–473. https://doi.org/10.1016/j.idc.2006.03.004

Morse SA, Budowle B, Schutzer SE (2019) Microbial forensics: what next? In: Microbial forensics

Murch RS (2003) Microbial forensics: building a national capacity to investigate bioterrorism. Biosecur Bioterror Biodef Strateg Pract Sci 1:117–122. https://doi.org/10.1089/153871303766275781
National Research Council (2014) Science needs for microbial forensics: developing initial international research priorities. National Academies Press, Washington, DC
Nema V (2018) Microbial forensics: beyond a fascination. In: DNA fingerprinting: advancements and future endeavors. Springer Singapore, Singapore, pp 295–306
Noller AC, McEllistrem MC, Pacheco AGF et al (2003) Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic Escherichia coli O157:H7 isolates. J Clin Microbiol 41:5389–5397. https://doi.org/10.1128/JCM.41.12.5389-5397.2003
Oliveira M, Amorim A (2018) Microbial forensics: new breakthroughs and future prospects. Appl Microbiol Biotechnol 102:10377–10391. https://doi.org/10.1007/s00253-018-9414-6
Park J-L, Park S-M, Kwon O-H et al (2014) Microarray screening and qRT-PCR evaluation of microRNA markers for forensic body fluid identification. Electrophoresis 35:3062–3068. https://doi.org/10.1002/elps.201400075
Power DA, Cordiner SJ, Kieser JA et al (2010) PCR-based detection of salivary bacteria as a marker of expirated blood. Sci Justice 50:59–63. https://doi.org/10.1016/j.scijus.2009.04.006
Rajalakshmi S (2017) Different types of PCR techniques and its applications. Int J Pharm Chem Biol Sci 7:285–292
Robinson JM, Pasternak Z, Mason CE, Elhaik E (2020) Forensic applications of microbiomics: a review. Front Microbiol 11:608101. https://doi.org/10.3389/fmicb.2020.608101
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467. https://doi.org/10.1073/pnas.74.12.5463
Schmedes S, Budowle B (2019) Microbial forensics☆. In: Reference module in biomedical sciences: encyclopedia of microbiology, 4th edn. Elsevier, pp 134–145
Schmedes SE, Woerner AE, Budowle B (2017) Forensic human identification using skin microbiomes. Appl Environ Microbiol 83. https://doi.org/10.1128/AEM.01672-17
Schutzer SE, Budowle B, Keim PS (2011) Microbial forensics: educating the workforce and the community. In: Microbial forensics. Elsevier, pp 667–680
Sego LH, Anderson KK, Matzke BD et al (2007) An environmental sampling model for combining judgment and randomly placed samples, Richland
Smith JAL (2019) Collection and preservation of microbial forensic samples. In: Microbial forensics
Smith JAL (2020) Collection and preservation of microbial forensic samples. In: Microbial forensics. Elsevier, pp 313–322
Smith GJD, Vijaykrishna D, Bahl J et al (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459:1122–1125. https://doi.org/10.1038/nature08182
The College of Physicians of Philadelphia (2018) Biological weapons, bioterrorism, and vaccines. Hist. Vaccines An Educ. Resour. by Coll. Physicians Philadelphia
Thierry S, Tourterel C, Le Flèche P et al (2014) Genotyping of French Bacillus anthracis strains based on 31-loci multi locus VNTR analysis: epidemiology, marker evaluation, and update of the internet genotype database. PLoS One 9:e95131. https://doi.org/10.1371/journal.pone.0095131
Török TJ, Tauxe RV, Wise RP et al (1997) A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. JAMA 278:389–395. https://doi.org/10.1001/jama.1997.03550050051033
Tridico SR, Murray DC, Addison J et al (2014) Metagenomic analyses of bacteria on human hairs: a qualitative assessment for applications in forensic science. Investig Genet 5:16. https://doi.org/10.1186/s13323-014-0016-5
Van Ert MN, Easterday WR, Simonson TS et al (2007) Strain-specific single-nucleotide polymorphism assays for the Bacillus anthracis Ames strain. J Clin Microbiol 45:47–53. https://doi.org/10.1128/JCM.01233-06

Varun CN, Kuruvilla Thomas S, Zevita F (2012) Microbial forensics- past, present and future. Int J Biol Med Res 3:1546–1549

Wang Z, Zhang S, Di Z et al (2013) Messenger RNA profiling for forensic body fluid identification: research and applications. Fa Yi Xue Za Zhi 29:368–374

Wetterstrand KA (2020) DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). http://www.genome.gov/sequencingcostsdata. Accessed 10 Jun 2021

Wikipedia Contributors (2021) List of epidemics. In: Wikipedia free encyclopedia. https://en.wikipedia.org/w/index.php?title=List_of_epidemics&oldid=1028613106. Accessed 15 Jun 2021

Yang J, Tsukimi T, Yoshikawa M et al (2019) Cutibacterium acnes (Propionibacterium acnes) 16S rRNA genotyping of microbial samples from possessions contributes to owner identification. mSystems 4. https://doi.org/10.1128/mSystems.00594-19

Zhang Y, Pechal JL, Schmidt CJ et al (2019) Machine learning performance in a microbial molecular autopsy context: a cross-sectional postmortem human population study. PLoS One 14:e0213829. https://doi.org/10.1371/journal.pone.0213829

Zou K-N, Ren L-J, Ping Y et al (2016) Identification of vaginal fluid, saliva, and feces using microbial signatures in a Han Chinese population. J Forensic Legal Med 43:126–131. https://doi.org/10.1016/j.jflm.2016.08.003

Part VII

Quality Control and Challenges in Forensic DNA Analysis

Touch DNA Analysis

43

Sourabh Kumar Singh, Amarnath Mishra, and Akanksha Behl

Contents

Abstract

Touch DNA examination is regularly used to pick up data from natural materials to help examinations related with criminal offenses, catastrophe casualty recognizable proof and missing individual's examinations. As the noteworthiness estimation of DNA profiling to scientific examinations has expanded, so too wants to create this data from more modest measures of DNA. Touch DNA tests might be characterized as any example which falls beneath suggested edges at any phase of the investigation, from test discovery through to profile translation, and cannot be characterized by an exact picogram sum. The assortment, DNA extraction, intensification,

S. K. Singh
Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, India

A. Mishra (✉) · A. Behl
Amity Institute of Forensic Sciences, Amity University Uttar Pradesh, Noida, Uttar Pradesh, India
e-mail: amishra5@amity.edu

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_45

profiling, and translation of touch DNA tests, tainting, and move issues are additionally quickly examined inside the setting of touch DNA examination.

Keywords

DNA · Helix · Swabbing · Scaping · Tape lifting · PCR · Gene analyzer

Introduction

Touch DNA usually denotes the DNA particulates left behind by a carrier while coming in the physical contact with any surface. It is encountered and collected from any biological matter and is transferred from a carrier to any surface or an individual during a contact. This evidence acts very importantly in any case seeking for a forensic lab examination and is also taken as the important tool for proceeding with the investigation. Till date there have been a lot of studies reported but only a few suggests and reports the investigation and retrieval of touch DNA from any garment. These are also referred as invisible DNA as it is not visible to the naked eye and are usually present in a very small quantity as compared to the amount of DNA present in any blood sample, and so are also very difficult to encounter because of its invisibility (Burrill et al. 2019).

This technique increased dramatically the amount of evidence which might be used for the detection of DNA. As these are invisible to the naked eyes, so one does not need to look for anything specific for DNA. It just needs a few cells (7–8) from epidermal layer of our skin (Mishra et al. 2015). STR profiles extracted from touch DNA can be analyzed for the purpose of solving crimes and the extraction process can determine the recovery of genetic material obtained from different origins (Mishra et al. 2020).

Individualization of people has for some time been a challenge confronting law authorization. An ideal framework incorporates distinguishing attributes novel to every person, with highlights that do not change over the long haul, which can be recorded with the end goal that suspects examples can be looked at against a bunch of realized reference test. Taken with regards to current legal science, in 1910 Edmund Locard set up the main crime lab as an educator of legal medication at University of Lyons, France, also is most popular for his proof exchange hypothesis in criminological science, "The Locard's Exchange Standard." In 1918, he additionally first recommended the 12-point coordinating framework for positive unique mark distinguishing proof. With the innovation of today, the straightforward demonstration of getting an item or touching a surface can prompt the recognizable proof and anxiety of a crook. Touch DNA is too called as contact trace DNA. It alludes to the DNA that is recuperated from skin (epithelial) cells that is abandoned when an individual touches or meets things, for example, garments, a weapon, or different articles. As a criminal touch the weapon or then again any article at the crime scene, skin cells left behind on surface prompts individual recognizable proof of the lawbreaker. An individual sheds around 400,000 skin cells every day, except it is

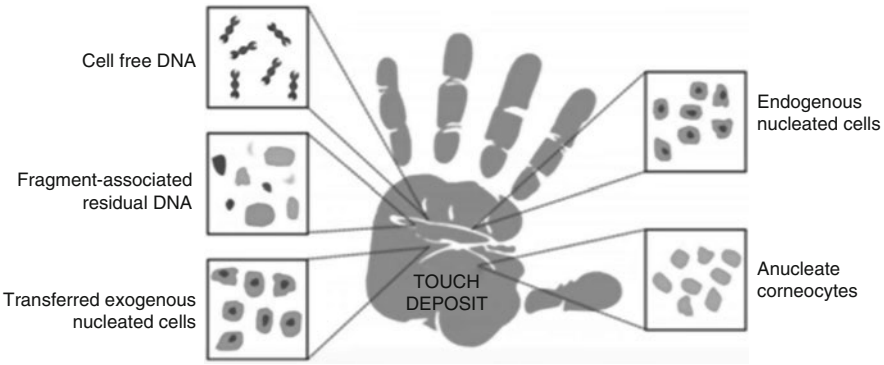


Fig. 1 Potential source of touch DNA

the lower skin cells that will give the best DNA profile. These cells are normally recuperated when power is utilized, for example, on the casualty's garments or at a crime scene after a battle has happened. "Touch DNA" is not to be mistaken for "Low Copy Number" DNA, or improved PCR techniques. Different touch DNA testing procedures have been utilized at the crime scene and in criminological labs worldwide for longer than 10 years. The quantity of cells moved to touched items is exceptionally factor, and regularly results in under 300 picograms of DNA. Touch DNA otherwise called contact trace DNA is recuperated from epithelial cells of skin that are moved to surfaces when an individual interacts with any surface. Human creatures shed a few skin cells consistently and these cells might be moved to surfaces of our skin co-crime is carried out; the culprit of the crime may store adequate number of skin cells on things found at crime scene like homicide weapon, dress, casualty, addressed archive, and so on. These confirmations whenever gathered and DNA removed may help in connecting the culprit to the crime scene (Daly et al. 2012) (Fig. 1).

Ordinarily, the lower skin cells are expected sources of DNA bringing about a decent DNA profile. These cells are commonly recuperated when power is utilized, for example, on casualty's garments or at a crime scene after a battle has happened. This opened conceivable outcomes and prompted the assortment of DNA from a more extensive scope of displays (counting: instruments, attire blades, vehicles, guns, food, bedding, condoms, lip makeup, wallets, adornments, glass, skin, paper, links, windows, entryways, and stones). Broad endeavors have been made to gather touch DNA from various substrates, for example, human skin, insides of latex gloves, lip prints, controlling wheels, entryway handles, and device holds and shafts. Among potential substrates, dress is regularly thought to be a wellspring of significant proof for criminological DNA examination in criminal examinations, for example, in rape cases. It is realized that touch DNA can be recuperated from the bodies and apparel of assault casualties, however helpful profiles are hard to obtain since touch DNA is typically saved in limited quantities contrasted with bloodstains or other organic liquids (Table 1).

Table 1 Potential evidence and types of cases

Potential evidence	Type of case
Firearm, knife handles, and weapon handles	Any
Fired cartridges	Any
Steering wheel and other vehicles	Carjacking etc.
Fingerprints on victim	Any
Ligatures, hand cuffs, and shoestrings	Strangulation, kidnapping, rape, etc.
Hand swabs from suspects	Strangulation, rape
Face swabs from child victims (slapping, hitting)	Child abuse
Swabs from limbs removed from animal carcasses	Poaching
Cell phones swabs	Robbery, etc.
Victim neck swabs	Strangulation
Swabs from torn or forcibly removed clothing	Rape, assault
Airbags	DUI cases and others
Tools	Burglary
Baggies	Drug possession
Paper demand notes	Bank robberies
Clothing items, hats, masks, gloves, and glasses	Any

History of Touch DNA

In 1997, Van Oorschot and Jones observed and reported possibility of recovering the touch DNA through the epithelial cells, found on an exhibit. This method turned into developed early inside the 2000s and "lets in evaluation of simply 'seven or eight' cells from the outermost layer of pores and skin." Fifty-one touch DNA is likewise referred to as epithelial DNA. It makes use of the equal approaches to study bodily fluids as conventional DNA makes use of, however the trying out is on those closing epithelial cells. When a person touches an item, epithelial cells are frequently left at the back of. The quantity left in the back of is frequently less than a hundred picograms and is also called low copy DNA. This is proof with "no visible staining" that might possibly incorporate DNA because of the switch of epithelial cells from the pores and skin to an item.

Since then, there are numerous research studies to be had inside the literature in which DNA has been expounded to be recovered from treated objects. Such objects include handbags, garb, rings, guns, and automobile steering wheels. While this would appear the herbal development of DNA and pores and skin hint analysis, it is far vital to understand the distinction among the abilities of touch DNA and chemical evaluation of skin traces and conventional DNA. They are twofold and raise widespread problems as to whether or not the criminal framework surrounding conventional DNA analysis is sufficient for this new shape. Due to this development, decrease quantities of human DNA can be detected and, probably, a complete or partial STR profile may be generated.

Techniques for Collection of Touch DNA

Numerous crime scene examiners and labs test for touch DNA utilizing either the wet/dry cleaning or cutting techniques. At the point when the cleaning strategy is used, the outside of the thing is typically scoured with a wet q-tip, followed by a dry q-tip with an end goal to gather conceivable skin cells. The wet/dry cleaning strategy is suggested for hard, nonpermeable things, for example, metal, glass, or plastic, and can without much of a stretch be performed at the crime scene with restricted danger of defilement with exogenous DNA (for example, from the individual gathering the example, or from close by surfaces/objects). The cutting strategy might be utilized for delicate things, for example, dress, in which texture from zones of revenue is sliced to gather potential cells (Kirgiz and Calloway 2017).

These two methodologies can be effective on numerous things of proof; nonetheless, the two of them have the constraint of putting superfluous substrate (the q-tip itself or the texture cuttings) into the little DNA preparing tube. There is a restricted measure of substrate that can be set in a cylinder, and the substrate itself may "trap" a few cells during handling, diminishing the probability of acquiring results. Notwithstanding the ordinarily utilized cleaning and cutting techniques, a few labs additionally utilize the scraping also, tape lift strategies, in which the outside of delicate/permeable things are either scratched with a sterile surgical tool sharp edge, or tested with a little bit of scotch tape, or the glue part of a post-it note, to gather conceivable skin cells. The examiner using the scratching or tape lift technique will center their testing to a region of harm, or the zone where the culprit is accepted to have had the most contact. Using these testing techniques, a territory roughly the size of a grown-up's hand can be examined (Hanson and Ballantyne 2013). The scrapings/tape/post-it notes are then positioned straightforwardly in the extraction tube. Labs utilize pretreated tape (generally presented to an UV cross-linker) and will likewise deal with a clear bit of tape close by the proof example to guarantee that no DNA has been presented by means of unusual pollution from the producer. It should be noticed that glues can be hazardous during the DNA extraction method and all things considered, the examiner should guarantee that their lab of decision has an approved extraction system that can effectively eliminate the glue without influencing the DNA yield. The scratching and tape lift strategies permit a bigger surface territory to be tested rather than the cutting strategy. An expansion in surface zone builds the odds of recuperating more skin cells, which expands the odds of getting a DNA profile (Templeton et al. 2013). As referenced, the scratching/tape lift strategies are ideal in circumstances where the researcher can find zones on the things which are well on the way to contain the culprit's skin cells. On the off chance that dress was left at the crime scene by the culprit, pressure focuses on the garments, for example, the inside neck of a shirt or within headband zone of a cap are fantastic possibility for these inspecting techniques. In a rape case in which the casualty's attire had been taken out by the culprit, regions, for example, the belt may contain adequate cells having a place with the culprit to create a profile.

The initial phase in gathering follow tests is to distinguish which regions to target. All things considered, follow tests on surfaces are not promptly recognizable. While finger printing specialists are utilized to distinguish touched regions on certain displays, numerous shows are cleaned, or tape lifted dependent on suppositions about where the DNA-containing material is found. The utilization of nonobtrusive discovery frameworks would be ideal. One such framework is the utilization of light sources, for example, the Poliliaght. Be that as it may, the utilization of these is not as far and wide as it very well may be. This might be because of an absence of consciousness of their handiness, or their apparent difficulty, or their nonideal presentation for explicit assignments or that further insightful and approval work is needed to characterize their degree and impediments (Singh Sankhla and Kumar 2017). Touched surfaces that have been uncovered utilizing fingerprinting techniques are generally those surfaces on which fingerprints are looked for as the need, instead of surfaces where DNA will be examined. Albeit many fingerprinting techniques do not antagonistically influence the nature of recovered DNA, a few approaches may do as such. Furthermore, others may diminish the amount of recovered DNA.

While thinking about the downstream utilization of the unique mark, more accentuation should be given to the effect of the fingerprinting procedure utilized on ensuing DNA recovery and quality. Upgrades in the techniques for distinguishing the natural wellspring of follow tests (not simply fingerprints) on show surfaces, and their application throughout criminological examinations, should assist with improving example assortment.

Cleaning an accepted follow test region that is more modest than the real affidavit territory will imply that a portion of the example goes uncollected. Cleaning a territory more noteworthy than the genuine region of store may imply that example is spread over a more extensive region and that less is gathered. The two methodologies additionally can possibly give a wrong perspective on where the genuine example was found. It is, thusly, not just important to know about the exact area of the material being focused on yet additionally to gather from the zone fittingly (Aditya et al. 2011).

An important thing that when a person collects the touch DNA is that everything or every individual present at the crime scene his/her DNA could also be contributed while the collection, because touch DNA is something that is invisible to our naked eye. So, an appointed personal for the collection of touch DNA is recommended to where appropriate personal protective equipment.

Swab Technique

For the collection of touch DNA from fingerprints or other skin cells a swab is often taken. While collection of touch DNA one must always choose the specifically mentioned DNA-free or controlled swabs. This means the swabs are free of human touch and any contaminations like DNase, RNase, and DNA, which possibly may contaminate a sample.

While swabbing, the swab first needs to get wet using only a single drop of distilled water and not by dipping the swab into water. When the swabbing method is utilized the surface is usually rubbed with the wet cotton swab and then is followed by a dry cotton swab in an effort to collect the possible skin cells. And the wet of the dry swabbing method is recommended for the hard nonporous items such as metal, glass plastic, etc. and it can be easily performed at the scene of crime with a limited risk of contamination with exogenous DNA (Sessa et al. 2019).

Apply the tip of the swab to the sample area and using gentle pressure rub the same while making sure to make the contact of all the surface of swab with the sample surface by rotating it. Do not rotate the swab more than once to avoid the redeposition of sample on the surface. Air dry the swab and place it into the designated vial. After labeling the vial needs to be packed and sent to the lab for the analysis. To improve the quality of the resulting DNA profiles a double swabbing technique is usually been applied.

Cutting Out Technique

It is the frequently used method in many forensic laboratories and the area of interest is usually been the soft items which have been left over in the crime scene. In this technique the most common sample are cloth samples. If a forensic expert considers the cloth samples, they are collected from the clothing leftover at the scene of crime by the perpetrators. The pressure points on the clothing such as the interior neck of a shirt or the inside headband area of a hat are excellent area for the sampling (Sessa et al. 2019). In the sexual assault case in which the victim's clothing has been removed by the perpetrator, areas such as the waistbands may contain sufficient cells belonging to the perpetrator to produce a particular profile.

Scraping Technique

In the scraping technique the surface of the porous items is either scraped with a sterile scalpel or blade to get the touch DNA. The scraping technique allows a larger surface area to be sampled as compared to the cutting technique or the swabbing technique. An increase in the surface area increases the chances of recovering more skin cells, which increases the chances of obtaining a DNA profile.

Tape Lift Technique

In this technique adhesive tapes are used to carry out the deposited touch DNA on the suspected area. The analyst will mainly focus their sampling to the area of damage or the area where the perpetrator is believed to have had the most contact and using this sampling technique an area approximately the size of an adult's hand can be sampled. So, the analyst may tape lift the sample from almost an area from the

adult's hand. The sampled tape may directly be placed for the extraction procedure. Forensic labs use pretreated tape (usually exposed to a UV cross-linker) and will also process a blank piece of tape alongside the evidence sample to ensure that no DNA has been introduces via adventitious contamination from the manufacturer (Bonsu et al. 2020).

Techniques of Touch DNA Examination

The analysis of touch DNA evidences follows the same general practices of forensic DNA except an utmost care to avoid contamination in the sample. Most of the touch DNA samples tend to contain trace amount of DNA, facilitating the contamination from the exogenous DNA in them. Hence, it is advisable to process such evidences separately without mixing them with the high DNA content samples. The routine forensic DNA analysis is a multistep process, starting with the collection of DNA evidences such as saliva, blood, semen, hair, skin cells, or any biological material containing cells. The next step is extracting the DNA, where the cells are broken open using a combination of chemical and heat to release the contents of the nucleus where the DNA is stored. The material then goes through a cleanup process, so the sample is pure DNA (Verdon et al. 2014).

The next step is the process to determine how much DNA was retrieved through a process called quantitation. Quantitation is important because the next step requires a specific amount of DNA to achieve optimum results. Too much or too little requires the concentration to be adjusted before the next step. After quantification comes the amplification process where specific regions inside the DNA molecule are copied. During the amplification process fluorescent tags are placed on regions of the DNA which will be used in the last step. The last phase is the detection process where the copied fragments of DNA are separated by size and are passed through a laser, when this laser hits the fluorescent tags a camera detector determines what color the tag is and how much there is (Comte et al. 2019). A DNA profile is a set of numbers at the different regions tested called locus. One of the numbers come from the mother and another from father linking the subject to each one.

Extraction and Isolation of Touch DNA

To extract the DNA from surfaces such as metals, different extraction buffers were used. These buffers utilize protease which accounts for DNA binding by the process of opening of cells, which is essential for touch DNA recovery. The procedure of isolation and extraction of DNA can be conducted through various methods such organic extraction, Chelex extraction, and silica-based extraction. The former two methods can cause a loss of a part of DNA while the extraction process. The silica-based extraction includes silica-coated magnetic beads to seize DNA from the disintegrated cell. DNA is generally obtained from epithelial pores and skin cells; these strains are assessed for the best quantity. Some techniques such as quantitative

PCR assay, dot blot technique, capillary electrophoresis, and fluorescent dye assay are utilized for isolation of DNA inside the touch samples.

As contact DNA is to be had in trace amounts, the DNA is amplified to provide numerous copies which can be then assessed for duration polymorphism and sequence polymorphism. Following the PCR amplification, out of 13 STR loci as depicted in Fig. 8 are identified which showcase polymorphism among one-of-a-kind individuals. In addition, the DNeasy® plant mini package (QIAGEN®) while in comparison with the QIAamp® mini kit turned into found to decorate DNA restoration from paper with the aid of over a 150%. The nature of substrate from which the DNA has been recovered may want to have an influence on DNA extraction. However, the DNA extraction manner can result in a loss of about 20% to 90% of the initial template quantity depending at the extraction method used, as well as the accuracy of the quantification approach.

Evidentiary Value

When pondering testing for DNA, the examiner needs to think about the likely evidentiary estimation of the DNA. The investigator must consider the connection between the person in question and the suspect (if one exists), and any chance of "innocent transfer" of DNA that may have happened before the alleged crime. For instance, if the suspect is a relative, and either lived with, or had ongoing contact with the person in question, at that point finding the speculator's DNA on the proof might be of restricted probative worth. Touch DNA can undoubtedly be transferred all through the family by means of everyday connections, contact with furniture things/bedding, or through the clothing (Hefetz et al. 2019).

Touch or trace samples usually have only a certain quantity of DNA, which can get affected while its collection procedure and examination. The position where such DNA is lost cannot be determined due to the fact that there is not an availability of authentic positive sample which can lead to track the quantity of DNA by the examination process. When collecting touch DNA the different methods of collection can have a direct effect on the amount of DNA that has been collected from a crime exhibit. It is recommended for the crime scene investigators to choose accurate collection methods so as to procure the samples in the best possible manner. Usually in the recent time, trace DNA profiling was also conducted to extract DNA profiles from touched objects. Touch DNA can also be called as low copy number DNA or low template DNA. However, due to its amount, trace or touch DNA would be an appropriate term. It comprises of the minute level of DNA present in the small quantity of biological exhibits, and the process of extracting it. During the process of amplification, the term low template is used, where the use of low content of DNA will not generate proper results. Trace DNA can also be referred to the exhibit that consists of low levels of DNA at any stage of the examination process such as detection, collection, extraction, amplification, and interpretation. Touch DNA usually includes scanty biological exhibits and less than 100 pg, meagre content of DNA (Figs. 2 and 3).

Fig. 2 Touch DNA sources

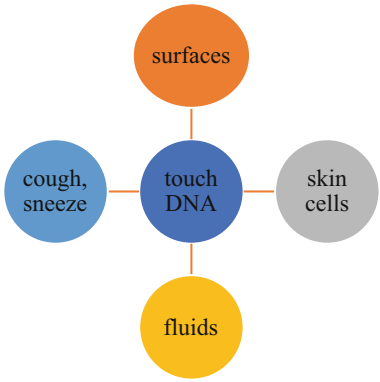
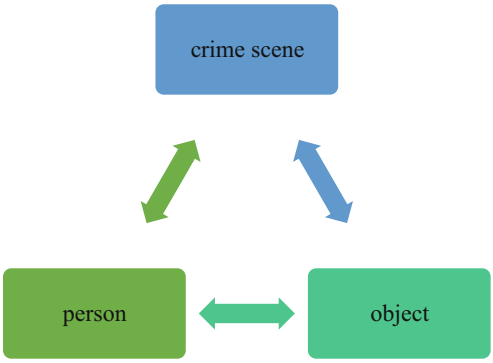


Fig. 3 Touch DNA transfer



According to one such case, touch DNA is referred to as a "kind of pseudoscience." It takes samples that incorporate a mixture of DNA from surfaces touched via an unknown wide variety of human beings and makes a wager as to the chance that a criminal suspect's DNA is a number of the haystack of unknowns the lab is inspecting. The standards and suggestions for "touch DNA" forensics are also a multitude. There's no unified set of requirements governing use, so analysts are loose to create their very own baselines. Guesswork continues to be guesswork, even though it involves technical software program and educated analysts. Guesswork is not evidence.

Considering what's at stake, touch DNA evaluation is not always specific enough to fulfill the evidentiary fashionable needed to lock human beings up. Journal articles regarding the switch of DNA have proven that DNA is not always transferred via contact on its own. In addition to the complications which can rise up from contact DNA on the technical side, there are many prison issues which can be implicated due to the wrong use and application of DNA. The maximum enormous is that it is able to lead to wrongful incarceration. The purpose touch DNA is the least favored DNA

source is because the man or woman has to touch the object for a prolonged period of time and/or time and again touch the object several times to go away a sufficient quantity of skin cells to generate a usable DNA profile. Touch-switch DNA "should falsely link someone to against the law" and forensic scientists counting on current excessive-sensitivity gadget ought to "falsely conclude that DNA left on an object is a result of direct contact." Their findings found out that it is impossible for scientists to determine whether or not the tiny bits of DNA got here.

Touch DNA can get transferred from skin to object or from skin to skin and then to the sample. Secondary transfer of DNA may be from skin to skin to object or from skin to object to skin. A systematic analysis of secondary DNA transfer observed that underneath their experimental conditions secondary switch of complete profiles changed into no longer located, however those occasionally minor peaks in a few samples were located. Secondary switch of DNA became determined in some research and additionally located in DNA recovered from bedding. Research suggests a person may also slough off thousands of lifeless pores and skin cells in line with day. Studies have shown that the type of floor the DNA is being deposited on matters substantially. The extra abrasive the floor, the higher the threat extra skin cells may be sloughed off in the course of touch.

The terms like direct and indirect transfer pertains to the routes by way of which DNA can be passed on. Some other terms such as primary, secondary, and tertiary are also used while referring to the transfer of DNA. DNA can be passed on multiple time and can be called as multistep transfer pathway; one needs to be clear on what is meant with the aid of secondary switch within the context of the state of affairs at hand. For a few, secondary switch method is any switch event after the number one switch; for others it simply refers back to the singular transfer step after the preliminary deposit. In addition, while relating to a selected touch event inside a long collection of a couple of contacts, one may also talk over with the number one and secondary substrates worried in a specific contact even though they will now not be the primary or second substrates inside the collection of contacts. As in most case situations, when taking into account the possibility of direct versus indirect transfer, the range of oblique steps are unknown, consequently forensic expert decides on the use of the term "indirect transfer" in place of "secondary switch," until the situations put forward by means of prosecution or defence, or known facts within the case set up that the indirect transfer is primarily based on best a single step after preliminary deposit.

However, on account of a rape by an odder, finding the suspect's DNA anywhere on the victim's dress might be of evidentiary worth. In these circumstances it is critical to assemble however much data from the victim as could be expected (if living), or to endeavor to reproduce the occasions if the victim is expired. For example, in the event that the victim's pants were pulled somewhere around the culprit, at that point the examiner and criminological researcher ought to consider testing regions for touch DNA where one would imagine that the suspect would have snatched during the attack. Finding the suspect's DNA on the victim's dress, and in specific regions of the apparel, may help validate the victim's rendition of occasions

and help address the charges being referred to. It is additionally significant that the examiner endeavors to gather the apparel of expired people or gather tests from the garments before the deceased being taken out from the scene. Assortment of the apparel at the scene and ideal safeguarding considers the chance of acquiring touch DNA sometime in the future, regardless of whether it is not at first shown to be available at the crime scene.

It is important that the investigator furnish the forensic scientist with some case foundation data to get the best guidance on the possible estimation of DNA evidence, just as proposals for testing. The arrangement of crime scene photographs can regularly be very valuable. Thus, it is similarly significant that the forensic scientist is outfitted with appropriate inquiries for the investigator to reply or consider (Oldoni et al. 2017).

Limitation

Sampling techniques of touch DNA and the procedures of processing the examination for the DNA are very sensitive. So, it has a huge chance of getting contaminated by the investigating officers even while they are covered in a PPE kit. It very well might be important to acquire end tests from key individuals in the situation where unfamiliar DNA profiles are acquired that cannot be credited to a suspect or the victim. There is a huge chance of encountering a mixed DNA profile as it might be possible that a few individuals have come in contact with the victim or the surface of evidence; this mixed DNA might contain the DNA of victim, suspect, or any other individual who has been in the crime scene at any time. The investigator may likely to be confronted with the examination of what it implies if unexplainable DNA is encountered (Nunn 2013).

For instance, an unfamiliar male profile from a touch DNA test might be acquired from proof relating to a female victim. On the off chance that the male DNA profile does not coordinate the suspect being referred to the specialist needs to think about its importance to the case. The unfamiliar profile could be from the genuine culprit and the first suspect could be blameless. Or on the other hand maybe the DNA profile is from unusual exchange from crime scene staff, specialists on call, lab experts, or crime scene equipment, for example, brushes for fingerprint collection. These are largely conceivable outcomes that law requirement may need to assess and deliver to push ahead with the examination (Tobias et al. 2017).

Some evidence things are likewise not suggested for the examination of touch DNA tests. Such things incorporate those that are seriously debased (for example, rotten apparel) have been presented to extraordinary ecological conditions, (for example, weapons left outside for quite a long time or years), have been washed, or are vigorously absorbed the victim's body fluids. Additionally, things that are probably going to have been touched by numerous individuals, for example, a public compensation telephone or store counter are generally bad hotspots for probative or interpretable touch DNA profiles. Most forensic scientists will not agree with the proposal of the inspection of these things (Hanson and Ballantyne 2013).

Case Study

In Missouri, the Kansas City Police Department's (KCPD's) cold case crew and crime research facility, and the Jackson County Prosecutor's Office structure a triumphant group. The sum of what three have been granted National Institute of Justice (NIJ) DNA awards. Since January 2008 (the first of the 3 years that the KCPD Sex Crimes Cold Case Squad got NIJ Solving Cold Cases with DNA award subsidizing), the group has gotten more than 150 CODIS hits, outstandingly cleared in excess of 100 cases and gave in excess of 50 charges.

Capt. Imprint Folsom, authority of the KCPD Special Victims Unit, says the Sex Crimes Cold Case Squad, which was rebuilt under the Violent Crimes Cold Case Squad in March, has a commendable working relationship with the crime research center and the Jackson County Prosecutor's Office. Returning to 2000, around 115 virus cases, including manslaughters and rapes, have since been indicted to conviction and speak to a 96% conviction rate, says Ted Hunt, boss preliminary lawyer in the Jackson County Prosecutor's Office.

A Valid Example: Charging the Waldo Attacker

In May 2010, collaboration and DNA evidence helped Jackson County accuse Bernard Jackson of 15 lawful offenses from four assaults in 1983 and 1984, in the Waldo and Armor Hills zones of Kansas City. At that point, KCPD was researching Jackson as an individual of interest in five comparable rape cases occurring in the Waldo zone in 2009 and 2010. Primer reports from the 1980s were pulled, alongside evidence from the property room and the lab's drawn out capacity cooler. In the stretch of time of a Friday evening to Monday morning, criminalists found a coordinating profile.

"That sort of correspondence and fast activity, and the eagerness of the lab to drop everything and work overtime to settle the case upgraded our capacity to charge the suspect rapidly and get him off the road," Hunt says. The supposed chronic attacker had been captured and indicted after a 1984 robbery and assault, invested energy in jail, was delivered in 2008, and purportedly again began an example of crime. Folsom, who drove the Waldo attacker examination team, says the crime research center group was a gigantic assistance, especially those working in DNA and trace evidence who needed to organize which of thousands of things would be broke down first. In February 2011, the Jackson County Prosecutor's Office accused Jackson of 22 lawful offenses for two of the five cases from 2009 to 2010.

In one case, a girl become sexually assaulted and strangled to demise. She has been certain with a couple of ligatures, which include leather straps. Initial checking out found out that DNA from a semen stain at the victim's nightgown matched a convicted perpetrator. However, this became not sufficient evidence for a conviction due to the fact the suspect had recognized the victim and claimed they had had a consensual relation. The scraping method become performed on the ligatures to gather possible skin cells and the suspect's DNA turned into recognized on one of

the leather-based straps, offering compelling proof that the sexual come across changed into now not consensual. The suspect pled guilty following the presentation of the DNA proof at trial and is presently serving a life sentence.

Conclusion

The objective of forensic science is to boost the estimation of evidence. Distributed information and reports from operational scientists have demonstrated that up to 90% of the DNA in an example can be lost during assortment and examination (Cavanaugh and Bathrick 2018), anyway different investigations normal this misfortune at around 39% (Kemp et al. 2015). An expansion in comprehension of the difficulties in effectiveness of each progression in the DNA test work process has significant repercussions regarding likely improvement for all organic example types, including semen, blood, and salivation. Enhancements in proficiency in acquiring tests from substrates and ensuing extraction will expand the quantity of tests that will yield effective profiles, especially when the beginning amounts are low or ecologically tested, like the case with touch tests. With better comprehension of DNA yields and the components of misfortune, directed cycle enhancements will bring touch DNA tests into significantly more normal use with normalized improved strategies. Contingent upon substrate type and porosity, it is a test to acquire all the accessible DNA.

References

Aditya S et al (2011) Generating STR profile from "touch DNA". J Forensic Legal Med 18(7):295–298. https://doi.org/10.1016/j.jflm.2011.05.007

Bonsu DOM, Higgins D, Austin JJ (2020) Forensic touch DNA recovery from metal surfaces – a review. Sci Justice 60(3):206–215. https://doi.org/10.1016/j.scijus.2020.01.002

Burrill J, Daniel B, Frascione N (2019) A review of trace "Touch DNA" deposits: variability factors and an exploration of cellular composition. Forensic Sci Int Genet 39:8–18. https://doi.org/10.1016/j.fsigen.2018.11.019

Cavanaugh SE, Bathrick AS (2018) Direct PCR amplification of forensic touch and other challenging DNA samples: a review. Forensic Sci Int Genet 32:40–49. https://doi.org/10.1016/j.fsigen.2017.10.005

Comte J et al (2019) Touch DNA collection – performance of four different swabs. Forensic Sci Int Genet 43:102113. https://doi.org/10.1016/j.fsigen.2019.06.014

Daly DJ, Murphy C, McDermott SD (2012) The transfer of touch DNA from hands to glass, fabric and wood. Forensic Sci Int Genet 6(1):41–46. https://doi.org/10.1016/j.fsigen.2010.12.016

Hanson EK, Ballantyne J (2013) "Getting blood from a stone": ultrasensitive forensic DNA profiling of microscopic bio-particles recovered from "touch DNA" evidence. Methods Mol Biol 1039:3–17. https://doi.org/10.1007/978-1-62703-535-4_1

Hefetz I et al (2019) Touch DNA: the effect of the deposition pressure on the quality of latent fingermarks and STR profiles. Forensic Sci Int Genet 38:105–112. https://doi.org/10.1016/j.fsigen.2018.10.016

Kemp BM et al (2015) How much DNA is lost? Measuring DNA loss of short-tandem-repeat length fragments targeted by the powerplex 16® system using the Qiagen Minelute purification kit. Hum Biol 86(4):313–329. https://doi.org/10.13110/humanbiology.86.4.0313

Kirgiz IA, Calloway C (2017) Increased recovery of touch DNA evidence using FTA paper compared to conventional collection methods. J Forensic Legal Med 47:9–15. https://doi.org/10.1016/j.jflm.2017.01.007

Mishra A, Sathyan S, Shukla SK (2015) Application of DNA Fingerprinting in an Alleged Case of Paternity. Biochem Anal Biochem 4:165. https://doi.org/10.4172/2161-1009.1000165

Mishra IK, Singh B, Mishra A, Mohapatra BK, Kaushik R, Behera C (2020) Touch DNA as forensic aid: a review. Indian J Forensic Med Toxicol 14(2)

Nunn S (2013) Touch DNA collection versus firearm fingerprinting: comparing evidence production and identification outcomes. J Forensic Sci 58(3):601–608. https://doi.org/10.1111/1556-4029.12119

Oldoni F et al (2017) Sensitive DIP-STR markers for the analysis of unbalanced mixtures from "touch" DNA samples. Forensic Sci Int Genet 28:111–117. https://doi.org/10.1016/j.fsigen.2017.02.004

Sessa F et al (2019) Touch DNA: impact of handling time on touch deposit and evaluation of different recovery techniques: an experimental study. Sci Rep 9(1):1–9. https://doi.org/10.1038/s41598-019-46051-9

Singh Sankhla M, Kumar R (2017) Identification of criminal by using touch DNA: a new tool for investigation in forensic science. Imp J Interdiscip Res 3(5):2454–1362. Available at: http://www.onlinejournal.in

Templeton J et al (2013) Genetic profiling from challenging samples: direct PCR of touch DNA. Forensic Sci Int Genet Suppl Series 4(1):224–225. https://doi.org/10.1016/j.fsigss.2013.10.115

Tobias SHA et al (2017) The effect of pressure on DNA deposition by touch. Forensic Sci Int Genet Suppl Series 6(September):e12–e14. https://doi.org/10.1016/j.fsigss.2017.09.020

Verdon TJ, Mitchell RJ, Van Oorschot RAH (2014) Evaluation of tapelifting as a collection method for touch DNA. Forensic Sci Int Genet 8(1):179–186. https://doi.org/10.1016/j.fsigen.2013.09.005

The Use of Rapid DNA Technology in Forensic Science

44

Robert O'Brien

Contents

R. O'Brien (✉)
National Forensic Science Technology Center, Largo, FL, USA
e-mail: robrien@fiu.edu

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_46

Abstract

Forensic rapid DNA is a technology that develops a DNA profile from a sample in under 2 h, without human intervention. The entire process is performed on a single automated instrument and is much faster than conventional DNA methods. In addition to its speed benefits, rapid DNA supports processing outside of the laboratory and can be used at booking stations, crime scenes, border checkpoints, and disaster sites to aid in victim identification. This chapter briefly examines the history of rapid DNA, as well as currently available rapid DNA instrumentation and the characteristics of rapid DNA technology. It compares rapid DNA to traditional DNA methods, including the process, data interpretation, trouble-shooting, and ability to perform swab retesting. It also explores the possible uses of rapid DNA in various applications and environments. Finally, it examines areas for improvement and potential future development.

Keywords

Rapid DNA · Forensic DNA · Disaster Victim Identification · DVI · mobile DNA

Introduction

Forensic DNA typing was first used as an investigative tool in 1986. There have been significant improvements to the speed and sensitivity of forensic DNA testing, and it has grown to become the gold standard in forensic testing (Butler 2011). Decades ago, forensic DNA processing took weeks. With rapid DNA technology, testing can now be completed in under 2 h.

Rapid DNA analysis is commonly defined as the fully automated process that develops a forensic DNA profile from a sample in under 2 h, without human intervention (U.S. Federal Bureau of Investigation "Rapid DNA" 2021). While the improved processing time is critical for certain use cases of this technology, the implications of rapid DNA are much more significant. By reducing or eliminating the need for forensic analyst review, the technology expands the possibilities of who can process DNA samples. This technology has demonstrated the potential to move DNA processing from specialized laboratories into booking stations, crime scenes, and even border checkpoints and disaster scenes.

A powerful and versatile new tool in human identification, rapid DNA is poised to revolutionize the field of forensic DNA testing. This chapter will explore the world of rapid DNA testing, including its history, its technology, its current uses, and what the future might hold for this branch of forensic DNA testing. The information outlined in this chapter is intended to serve as an introductory overview based on current developments. Additional updates will surely be required with time as rapid DNA testing becomes as commonplace in the forensic field as traditional DNA testing is today.

History of Forensic Rapid DNA

Traditional forensic short tandem repeat (STR) testing methods require multiple workflow steps, including DNA extraction, quantification, PCR amplification, STR fragment analysis via capillary electrophoresis, data review, and interpretation. The process also requires specialized equipment located in a carefully controlled laboratory environment, run by trained forensic DNA analysts. Given this complexity and other factors, including case volume relative to capacity and available resources, it should come as no surprise that it can require multiple days or even sometimes months to obtain reported results.

Rapid DNA testing technology required significant advances in both amplification chemistry and instrumentation to fully integrate and accelerate processing time and enable analysis outside of the laboratory environment.

The first step toward rapid DNA technology was taken in the late 2000s, when researchers developed a faster polymerase chain reaction (PCR) amplification method (Vallone et al. 2008; Giese et al. 2009). The total time required for PCR amplification decreased to about 30 min. There were also significant advancements at that time in microfluidic efforts for PCR, separation, and detection (Horsman et al. 2007).

The initial development of the current rapid DNA systems was spurred in the late 2000s by funding released by a consortium of US federal agencies, including the Federal Bureau of Investigation (FBI), Department of Defense (DOD), Department of Homeland Security (DHS), and Department of Justice (DOJ). These agencies recognized the need for automated, integrated DNA analysis platforms with a simplified "swab in – profile out" process. Ideally, such platforms could be used for identification at locations ranging from local booking stations to mass disaster sites and operated by nontechnical personnel. By 2012, these funded development efforts had led to the release of two rapid DNA systems: the NetBio ANDE®/DNAscan™ and the IntegenX RapidHIT™ 200.

With the advent of rapid DNA systems capable of producing DNA profiles in less than 2 h, law enforcement agencies began to adopt the technology. A commercially available rapid DNA system was installed at the Palm Bay Police Department in 2012, leading to reports of successful use in active investigations not long after ("Rapid DNA Used by Prosecution," 2014). In 2014, rapid DNA programs were launched by the Arizona Department of Public Safety and Orange County District Attorney's Office in California. As of May 2020, the Orange County District Attorney's Office reported they had processed 427 cases using rapid DNA technology resulting in 138 investigative leads (Dadhania et al. 2020). In June 2020, the Arizona Department of Public Safety reported that rapid DNA had been used in over 530 cases and generated over 170 investigative leads (Arizona DPS SAB 2021).

A major development occurred in the USA with the passage of the Rapid DNA Act in 2017, which amended the DNA Identification Act of 1994 in order to permit upload to the National DNA Index System (NDIS) of DNA profiles generated by criminal justice agencies using rapid DNA instruments approved by the FBI Director

in compliance with FBI issued standards and procedures (U.S. Congress 2017). This enabled the FBI to work with the Criminal Justice Information Services (CJIS) Division and the CJIS Advisory Policy Board (CJIS APB) Rapid DNA Task Force to plan the effective integration of rapid DNA into the booking station process (U.S. Federal Bureau of Investigation "Rapid DNA" 2021).

September 1, 2020, marked the FBI release of the *Standards for the Operation of Rapid DNA Booking Systems by Law Enforcement Booking Agencies* (U.S. Federal Bureau of Investigation "Standards" 2020b) and the *National Rapid DNA Booking Operational Procedures Manual* (U.S. Federal Bureau of Investigation "National" 2020a), thus creating the operational framework for mainstream deployment of rapid DNA systems in booking station environments.

A number of law enforcement agencies outside the USA are also implementing or evaluating rapid DNA for custody suites/booking stations, crime scenes, and other scenarios. For example, the Italian Carabinieri Scientific Investigation Group (Raggruppamento Carabinieri Investigazioni Scientifiche (RACIS)) received ISO/-IEC 17025 accreditation for rapid DNA in 2019 and thereafter started utilizing the technology to successfully resolve multiple forensic cases, including a homicide, and upload profiles to the nation's genetic database, the Banca Dati Nazionale del DNA (Thermo Fisher Scientific "Italian" 2020b).

In Singapore, a comparison study was conducted on the performance of rapid DNA workflow and conventional forensic DNA workflow (Thong et. al. November 2015a). The study supported the use of rapid DNA for providing quick intelligence through the concomitant use of offender database searches. The authors subsequently published another article that reviewed several actual crime cases, which were resolved using rapid DNA technology (Thong et. al. December 2015b).

Due to international differences in legislation and approaches to forensic DNA in general, implementation and potential final solutions using rapid DNA could vary. However, the growing global acceptance and support for DNA databases as an effective means to reduce and deter crime (Anker et al. 2021) provides a strong incentive for many jurisdictions (DNA Resource) to adopt rapid DNA in order to accelerate and increase the scalability of these benefits.

Instrumentation and Technology

All current rapid DNA instruments perform the complete forensic DNA workflow using disposable consumables which include all reagents, materials, and waste containment required for STR analysis within a single instrument. This compact size allows users to deploy the instruments at a variety of locations, including laboratories, police booking areas, disaster sites, and border crossing stations. In addition, the most recently introduced rapid DNA instruments have been further optimized for use outside the laboratory in decentralized environments to enable operation by nonscientists (Salceda et al. 2017).

In 2018, the three current rapid DNA instruments were analyzed as part of a multiagency technology maturity assessment (Romsos et al. 2020). This included the

ANDE 6C System, developed by ANDE (Longmont, CO), and the RapidHIT 200 and RapidHIT ID instruments, developed IntegenX (Pleasanton, CA), which was acquired by Thermo Fisher Scientific (Waltham, MA) in 2018. The assessment was funded by the FBI Laboratory and the Biometrics Center of Excellence, and results were analyzed by the National Institute of Standards and Technology. Rapid DNA analysis for the systems tested resulted in an overall 85% success rate for generating 20 CODIS core loci profiles with automated analysis and an overall success rate of 90% (with 240 tested swabs) in generating data.

In addition to the technology maturity assessment, numerous studies have evaluated and/or validated the performance of individual instruments to ensure consistent, balanced, and precise results. The RapidHIT 200 instrument was evaluated in several studies (Hennessy et al. 2014, Jovanovich et al. 2015). The original DNAscan instrument was evaluated in 2016 (Della Manna et al. 2016; Moreno et al. 2017). The RapidHIT ID instrument was also evaluated (Salceda et al. 2017; Wiley et al. 2017), as was the ANDE 6C (Carney et al. 2019; Ragazzo et al. 2020; Turingan et al. 2020).

Ease of Use

The traditional DNA analysis methods require specialized education and training. Ease of use is a major factor in the adoption of rapid DNA technology, which is intended to be a fully automated (hands free) process requiring minimal operator training. First responders and individuals trained to deal with hazardous situations can be easily trained to set up and operate these instruments. Standard instrument operation could be taught in less than an hour. The instrument guides the user through sample processing with on-screen prompts. This simplicity makes it possible for the instruments to be readily and successfully deployed across a variety of agencies with users of many different skill levels. This expands the use of rapid DNA to include:

- Police officers at booking stations: DNA is collected at arrest and processed and searched before the individual is released.
- Crime scene personnel at crime scenes: DNA profiles can be developed from biological evidence and known reference samples to more quickly include or exclude potential suspects or develop investigative leads.
- Medical examiners and coroners: to confirm identity of victims of accidents or mass disasters.
- Military personnel: to identify human remains and enemy combatants in theater.
- Border patrol and immigration personnel at border crossings: to verify familial relationships and prevent human trafficking.

The straightforward operation of these instruments is as important to the rapid DNA process as the technology contained within. Enabling nontechnical personnel to generate DNA data can allow forensic scientists to focus on more complex

analyses while also preventing them from being deployed into potentially hazardous conditions. Data can be transmitted to a scientist located miles away for review before it is acted upon.

The "sample in, answer out" means there is very little preprocessing of samples and no manipulation of the sample once the DNA process is started. Most samples can be collected by simply swabbing and can go directly into the instrument without any further manipulation. However, this also means that any problems with the samples or the process cannot be detected or remedied during testing. If there is a problem, it is only detected at the end of the process based on the quality of the profile developed. Therefore, additional training is recommended to educate the user on what sample types are most likely to yield usable results (i.e., single-source buccal, blood, and saliva as opposed to touch/trace samples with potential mixtures) and the optimal methods for collection and loading into the rapid DNA instrument.

Portability

Traditional DNA laboratories require sufficient space to house the instrumentation, reagents, and consumables and a staff of trained forensic scientists. Certain processes also require a physical barrier between workspaces. Rooms must be routinely cleaned to avoid introducing contamination from the environment. Certain instruments are also susceptible to environmental factors, so controls are required for instruments to perform optimally.

Given these limitations, it is not a simple process to quickly establish an impromptu DNA laboratory at a mass disaster or a crime scene. Shipping samples to an established laboratory for analysis delays processing. Samples are generally shipped in batches, introducing a delay while the items are gathered. Transportation time can be a significant factor, especially to distant laboratories.

The portability of rapid DNA instruments brings the power of DNA processing to the front lines. The instruments are more rugged than traditional instrumentation used in DNA laboratories, allowing them to be moved between locations. The size of these instruments allows for easy shipping in a protective case. The instrument setup is as easy as plugging it in and turning it on. In less than an hour, a location can have DNA processing capability. Unlike traditional instruments, rapid DNA instruments do not require calibrations after shipping, a controlled laboratory environment, or highly trained scientists for operation.

Instrument portability means DNA samples can be processed on site. Only data files would need to be transmitted, rather than DNA samples. This means within 2 h a sample can be processed, the data sent to a laboratory for comparison, and a result reported back to personnel on location for immediate action. A process that historically takes weeks or even months can now be accomplished in hours. This is the true power of rapid DNA.

Uses and Applications

Booking Station

Booking is the procedure of collecting and recording information about an arrested person and the charge against that person into police records. In many jurisdictions, arrestee DNA is collected for certain crimes, i.e., felonies or violent offenses. The rationale behind this collection is to test the arrestee DNA sample and compare it to a local, state, or national DNA database, to determine if the individual, currently in custody, matches or "hits" to any DNA left behind at unsolved crimes.

For many agencies, it can take months to process arrestee samples. Once collected, the samples must travel to a forensic laboratory for DNA testing and are uploaded and searched against the relevant database. If a match occurs, this investigative lead is provided back to the law enforcement agency to pursue. This often requires locating and rearresting the person of interest who was previously in custody, which consumes valuable time of law enforcement officers and potentially exposes them to danger when reapprehending violent offenders. In addition, if a suspect was released from custody in the time it takes to collect, process, upload, and search the profile, that person could potentially flee the jurisdiction or commit further crimes.

If booking stations implement rapid DNA processing and standard operating procedures and DNA networks are modified to accommodate rapid DNA booking searches, crimes can be solved more expeditiously and suspects can be identified while still in custody, reducing the need to rearrest them. This also reduces the likelihood of improperly released individuals committing more crimes and impacting more victims.

Crime Scene

The typical procedure when dealing with biological evidence at a crime scene focuses on identifying the materials and then collecting, preserving, and preparing them for shipment to the laboratory. After the entire scene is processed, which could take hours to days depending on the type of scene, then those items collected are transported back to the agency. At this point, the agency then must log these items and perform the necessary paperwork to get the case entered in their system. Following this, if the agency that did the collection has a DNA laboratory, the items need to be put into the evidence vault and wait for the case to be assigned, before DNA testing can start. If the agency does not have its own DNA laboratory, then the items collected must then be transported to the laboratory that offers services to that agency. That process could take days, and once it arrives at the new laboratory, it may once again sit waiting for the case to be assigned.

Once the case is assigned, the analyst often has limited information on what pieces of evidence would give the best result. If multiple samples were collected from the same area, an analyst should ask themselves: Should all of those be processed, or only one or several? Which are the best ones? He or she must rely on whatever information is relayed in the case file, along with their training and experience. If case file information is limited, the analyst must contact the investigator for clarification. The next potential issue could be finding the people who collected the reference samples. It is no wonder cases can sometimes take weeks to months or even years to complete!

How can rapid DNA help to improve this situation? First, known reference samples from individuals at the crime scene can have their DNA processed at the scene. These profiles can be electronically sent to the DNA laboratory, rather than consuming valuable processing time for those typically routine samples. In that way, the laboratory can concentrate on processing the more complex or challenging samples that are more likely to yield reportable data with traditional methods.

With actual crime scene samples (i.e., biological evidence), care must be taken to triage what will be processed using rapid DNA and what will be sent back to the laboratory. However, if there are many samples present, they can be run at the scene and, using the instrument software, matched to any known reference samples. Rapid DNA could be used to determine which samples contain DNA foreign to the victim and those can be prioritized for processing at the laboratory.

Could this go one step further and lead to an arrest based on DNA results processed at the scene with rapid DNA technology? This has already occurred in Miami, Florida, on more than one occasion (Plasencia 2017). In one such instance, with a fatal "hit-and-run," DNA from the gear shift in the automobile was matched to a suspect in a crime where he was detained (Plasencia 2017). Samples from the vehicles were taken both for rapid DNA testing and for testing in a traditional laboratory. However, once the match was made to the individual, detaining officers had definitive proof to arrest the individual. The other samples sent to the laboratory would only need to be analyzed if the case went to court.

It could be argued that the result may have eventually been the same if rapid DNA was not involved. However, factoring the time it would have taken to get these samples processed, barring a confession from the individual, the suspect would most likely have been released while waiting for these results from the laboratory. If the individual chose to flee the area, this case could possibly take months to be resolved, as opposed to the resolution in only several hours enabled by rapid DNA. This case and others like it show demonstrate the utility of rapid DNA technology.

Satellite DNA Laboratories

In the city of Boca Raton, Florida, a small ancillary laboratory was brought online to screen items of evidence. The items that are positive for pertinent biological material are then submitted to the Palm Beach County Crime Laboratory for DNA testing. Screening of evidence can be very time-consuming and does not lend itself to

automation. Having a smaller screening laboratory with well-trained personnel takes that burden from the main DNA laboratory.

These screening laboratories can be smaller in size and can operate with less staff than a traditional DNA laboratory. They can screen a high volume of cases because that is their primary focus. The main DNA laboratory benefits because they no longer receive large items for testing. They will instead receive a cutting from an item, which contains the biological material needed for DNA processing.

Rapid DNA can give a screening laboratory additional capability. It can be used to help triage the evidentiary items and run reference samples, allowing the main DNA laboratory to receive items ready for DNA processing along with known reference sample profiles for comparison.

In remote locations, such as islands or rural areas, rapid DNA can also be used in satellite locations to generate profiles for analysis on-site or via secure electronic data exchange with a centralized laboratory, thus providing efficient DNA analysis capabilities for communities that are typically underserved by forensic services.

Disaster Victim Identification

Disaster victim identification (DVI) is typically performed under difficult conditions and, at times, in remote locations. Typical scenarios not only include natural disasters, such as hurricanes, tsunamis, wildfires, and earthquakes, but also include aircraft crashes, capsized ships, mass graves, and military incidents. Any mass fatality situation can quickly overcome local resources. Given the additional potential compounding infrastructure factors such as power and water supply disruption, setting up another standard laboratory workflow can be a prohibitive challenge.

The International Criminal Police Organization (Interpol) is a well-known intergovernmental organization that investigates many DVI incidents, since not all countries have a fully resourced and trained DVI response team. Interpol methodologies recommend that victims are first identified using dental and fingerprint records. However, depending on the situation and the condition of the recovered body, DNA analysis may be required to identify victims.

Disasters can occur anywhere and include victims from many different origins. Multinational cooperation is sometimes necessary when providing support but also when repatriating human remains. Having a DNA solution that can be deployed and provide answers quickly, without the need for a complex setup, can make a significant difference in these efforts.

As part of the postmortem phase of a mass disaster, rapid DNA provides a portable solution that can be taken to these scenarios and set up quickly with minimal requirements for additional instruments and reagents (Murakimi et al. 2020). In some situations, body preservation may also be difficult, and a quick response is necessary to ensure a DNA profile can be retrieved.

Extraction of DNA from bone using traditional laboratory methods is typically a laborious process involving physical and chemical techniques that can, depending on the method used, take over a day of hands-on time. Rapid DNA protocols have

been developed that streamline this process and allow investigators to obtain interpretable data from medium- to high-quality bone samples in significantly less time (Thermo Fisher Scientific "Bone Sample" 2020a).

Rapid DNA can also be used to identify and match the relatives of the victims of DVI incidents in the ante mortem phase, where other means of identification have failed or are not available. This can be an extremely time-consuming process, especially when families are not aware that their relative is a victim in the disaster. A mobile rapid DNA solution can be used to transport the instrument to the suspected families of the victims, providing a quick conclusion to an open case.

Rapid transport vehicles and transport boxes have been developed to support the mobility of the instrument. Some transport solutions include battery power, enabling instruments to be used in extreme locations where traditional workflows would otherwise be impossible. Given the mobility of the instrumentation, decisions in DVI situations can be made in the field.

The goal of DVI is to identify and provide closure for families of victims as quickly as possible, and rapid DNA can be a valuable part of that solution.

Match Confirmation

Offender DNA database systems such as the CODIS (Combined DNA Index System) are a very powerful tool. They are especially powerful in cases where no suspect is named in the case, but DNA has been found that could potentially identify the perpetrator. The DNA profile is entered and searched against the database, and any possible database "hit" must be confirmed. Under normal procedures in the USA, the hit confirmation must be done with a separate sample collected following the rules of evidence collection and preservation. Evidence chain of custody must also be maintained to be used for court purposes. In court, the analyst will not be allowed to say the profile generated from the crime sample matched against a profile in CODIS. This would potentially raise questions as to why this person's sample was in CODIS, which could introduce evidence of prior bad acts, which could in turn cause the jury to be biased. Therefore, the analyst must only testify to the work conducted to compare the DNA profile developed from the evidence against the DNA profile developed from a fresh buccal swab taken from the person of interest.

Often, when these new reference samples are submitted for CODIS confirmation, the analyst has already moved on to another case. They must then stop what they are doing to work on this single buccal swab. This causes a disruption in the case they are currently working on, and if they do not process the reference sample right away, this can delay the confirmation of the CODIS hit. It is important to realize that the confirmation of the CODIS hit requires putting this one reference sample through the entire DNA process from extraction to capillary electrophoresis and data analysis. This consumes time and can seriously impact the workflow of the laboratory. It would be ideal to batch confirmations, but that would mean waiting and delaying the closing of the cases to which they are related.

Rapid DNA can play a major role in these CODIS hit confirmations. Rapid DNA instruments can be run on-demand when the buccal swab is submitted to the laboratory, especially the instruments that can process one sample at a time. No waiting or batching is required. Since the instruments are so easy to use, the actual running of the sample can be performed by a less trained technician, rather than an analyst, and the resulting DNA profile sent to the analyst for comparison to the profile developed from the evidence in the case. This saves time on two fronts. Firstly, the CODIS confirmation sample is run as soon as it is submitted to the laboratory, accelerating closure of the case; and secondly, ongoing casework is not delayed when this sample must be run through the entire DNA process.

Removing Bias in Mixture Interpretation

Mixture interpretation guidelines from the Scientific Working Group on DNA Analysis Methods (SWGDAM) indicate that determinations concerning the possible contributors to a mixture should be made before performing comparisons to reference samples (SWGDAM 2017). Utilizing rapid DNA for known reference sample processing, conducted in a separate workflow or facility by different personnel than the forensic DNA analysts who analyze evidentiary samples, can help remove bias in mixture interpretation. After the forensic analyst has performed deconvolution of the mixed profile in questions, developing profiles for possible contributors and documenting these results, only then would they formally request the profiles developed from the reference samples for comparison purposes. In court, the analyst could potentially testify that they never had custody of the reference samples until such time as it was necessary.

Retesting of Swabs

One of the primary concerns with rapid DNA testing has been sample consumption. Sample consumption often requires permission from attorneys in order to prevent court challenges by defense. As a result, it is important that whatever sample is used to generate an investigative lead from a DNA sample does not consume all of the sample. Rapid DNA can be used to generate important investigative leads in a short period of time. However, if the DNA is consumed by the rapid DNA instrument, it prohibits additional DNA testing using conventional forensic laboratory methods.

This is a major reason why the *Non-CODIS Rapid DNA Considerations and Best Practices for Law Enforcement Use* from the FBI (U.S. Federal Bureau of Investigation "Non-CODIS" 2019) advises that two swabs should be collected when using rapid DNA instrumentation – ensuring that a swab can be sent to the laboratory for traditional DNA testing and court purposes once an investigative lead is created on a rapid platform. These two swabs can be either collected at the same time using a "bouquet method" or one swab at a time ("A-swab, B-swab"), providing one swab for laboratory testing followed by one swab for rapid DNA testing.

This is a good approach; however, one can still envision scenarios, where it would be beneficial to retest the swab run on the rapid instrument. To date, only one rapid DNA system has been shown to enable sample retesting using either rapid or traditional DNA methods as reported below (O'Brien and Barnhart 2021).

The RapidHIT ID System (Thermo Fisher Scientific) allows a processed swab to be removed after testing. The extraction method used on the RapidHIT ID is not as rigorous as conventional DNA methods, potentially leaving sufficient DNA on a swab for reanalysis. If a sample contains enough DNA, it is possible to reanalyze the swab using conventional DNA methods after it has already provided a profile via the RapidHIT ID. If so, a single swab may provide both quick investigative information and a confirmatory result.

Studies were performed to determine if the same swab run on the rapid system can be removed and used again with conventional DNA techniques to achieve similar quality profiles. After the rapid DNA instrument finished running, the swabs were removed and placed into a 1.5 ml test tube to air-dry. Some swabs were air-dried for a week, while others air-dried for a month or more. The swabs were then put through the entire traditional DNA testing process from extraction through capillary electrophoresis. The DNA profiles developed from these swabs were compared to DNA profiles developed from the rapid DNA instrument to check for the accuracy and overall quality of the profile. In all cases, if a high-quality profile was developed by the rapid DNA instrument, the swab was also able to generate a high-quality profile upon reprocessing, using traditional DNA testing methods. Furthermore, 100% concordance was obtained between the duplicate samples and those samples run with the RapidHIT ID and reanalyzed using conventional analysis. Similarly, with crime scene-type samples, like cigarette butts and drinking containers, enough DNA was present after running the swab through the RapidHIT ID to produce, in most cases, a full profile using conventional DNA analysis methods (O'Brien and Barnhart 2021).

These data exhibited that rapid DNA processing does not consume the entire sample on certain rapid DNA instruments. Swabs that are removed properly, taking the necessary precautions to avoid contamination, can be reprocessed with traditional DNA methods to generate a profile consistent with rapid DNA results.

Although retesting swabs would not be a first choice, if an unexpected problem occurs with the swab sent off to the laboratory, then it is important to note that the swab used for rapid DNA testing can be retested. For this to work, however, the swabs used on the rapid DNA instrument must be preserved until the laboratory has successfully processed the evidence sent to them. The frontline operators of the rapid DNA instrument will have to develop a process for preserving and maintaining the chain of custody for swabs after they are processed and communicating with the laboratory to track their progress. This process will require coordination between agencies as the time between samples run on rapid DNA instruments and in the laboratory can be months apart.

Comparison to Traditional DNA Methods

Rapid DNA systems are smaller, faster, and less complicated to operate than traditional DNA testing methods. Therefore, why hasn't traditional DNA testing been replaced by rapid DNA thus far?

There are still limitations on what rapid DNA instrumentation can do when compared to traditional DNA testing methods. Traditional DNA testing methods still tend to be more affordable and sensitive than rapid DNA and able to handle a wider range of sample types. The sections below provide a brief comparison of the traditional forensic DNA workflow with rapid DNA, to help elucidate differences and future areas for improvement.

Sample Collection

Both traditional and rapid DNA methods require collection of the samples so that they can be run through the extraction process. While some rapid systems use conventional swabs commonly used in traditional DNA processing, a study indicates that one rapid DNA instrument performs demonstrably better when using specialized swabs that are designed for use with that system (Manzella and Moreno 2020). These particular swabs use RFID-tagged caps that lock inside the biochip, meant to avoid the potential for switching samples during a run. While these RFID swabs provide an advantage to the rapid DNA instrument user, the need to use such swabs increases cost and reduces the flexibility to submit samples to a traditional or rapid DNA process as needed.

Extraction

In both traditional and rapid DNA methods, the first step is lysis, which breaks open cells, releasing DNA to be transferred for further processing. During most traditional DNA extraction methods, lysis is just the first step of a longer process called extraction. The lysis buffer is often coupled with mechanical processes like heating and agitation for a period which can be adjusted based on the sample type. High content DNA samples can be exposed to these conditions for a shorter period, whereas low content DNA samples can be incubated in these conditions for up to 24 h, if necessary. The flexibility to add these extra mechanical processes and increased incubation times make the lysis step in traditional DNA processing more effective at breaking open most, if not all, of the cells contained within the sample.

In traditional DNA extraction, lysis is followed by purification. The purification process removes cellular components that are released from the cell during lysis. In the rapid DNA systems, there is currently no mechanism to perform a thorough purification process. This means that cellular components other than DNA may pass

through the system after the lysis step. As a result of these differences in the lysis and purification processes, more DNA may be obtained using traditional extraction methods compared to lysis on the rapid DNA instrument.

Quantification

Traditional DNA testing uses a quantification step to estimate the amount of DNA present in samples prior to PCR amplification. Using the quantification results, the scientist can adjust the sample concentration to optimize the DNA input that works best for the amplification kit used.

The current rapid DNA instruments do not perform quantification of DNA. Eliminating this step saves time, because the sample is transferred directly to the amplification step following lysis. Because there is no quantification step, the DNA input for amplification cannot be adjusted, which may affect the results. To optimize DNA recovery, rapid DNA instruments currently employ two different sample cartridges: one designed for high-quantity samples and one for lower-quantity samples. The protocol for each cartridge has been optimized to maximize DNA recovery and interpretable results. If too little sample is placed into the cartridge, partial or incomplete profiles may be generated.

PCR Amplification

Traditional DNA and rapid DNA methods use similar or the same PCR amplification chemistry, depending on the platform, with a common set of standardized loci. The amplification kits in use today target the 20 CODIS core loci established by the FBI as well as additional, kit-specific loci. Therefore, profiles resulting from either rapid or traditional DNA processing can be compared. The profiles can be uploaded into local or national databases and compared to other profiles generated worldwide. The amplification chemistries, used in both traditional and rapid DNA instruments, have received FBI NDIS approval in the USA, so reference samples generated using these chemistries can be uploaded into the national level of the CODIS database, so long as all FBI standards and procedures are followed (U.S. Federal Bureau of Investigation "Rapid DNA" 2021).

Electrophoresis

Electrophoretic separation is the process used to separate and detect DNA in both rapid DNA and traditional DNA methods. There are differences between the electrophoretic separation processes on different rapid DNA instruments.

The RapidHIT ID instrument is very similar to traditional DNA methods in that it employs capillary electrophoresis. The cartridge on this instrument contains one capillary that can run at least 100 runs and as many as 150 runs with consistent use.

The instrument will alert the user when the cartridge containing the capillary needs to be replaced. Each new capillary placed on the instrument can be tested by running a ladder and positive and negative controls to ensure that the capillary and reagents are functioning properly and are not contaminated.

The RapidHIT 200 instrument utilizes an eight-capillary system that is embedded inside the instrument, which is guaranteed for 100 runs. A single run includes a native ladder and up to seven samples utilizing two sample cartridges. Positive and negative controls can be run at the user's discretion.

The ANDE 6C system utilizes single-use electrophoretic separation channels that are injection molded into the cassette. A new capillary is used for every run. The operator will rely on the manufacturer's quality control measures as there is no way to verify the quality of the capillaries before use. This system requires very tight quality control employed at the production level to ensure every capillary in each cartridge functions to the same specifications. Any differences can impact the result.

Data Interpretation

In 2004, the FBI incorporated guidelines and acceptance standards, enabling laboratories to utilize expert system software tools to automatically evaluate the quality of a single-source DNA profile. This allows for the upload of profiles to a DNA database without requiring manual review.

Rapid DNA systems employ expert system software to perform data interpretation according to defined rules, like those used in traditional DNA analysis methods. Expert system software automates data analysis and allows the rapid DNA system to be operated by nonscientists who are not trained in advanced data interpretation. Manufacturers set static interpretation thresholds based on the type of cartridge used (high- or low-level DNA). Interpretation thresholds and quality flags are used to identify profiles requiring review, such as mixture profiles or profiles resulting from poor-quality samples. If the sample passes the interpretation thresholds, the DNA profile is considered acceptable and manual review is not required.

The difference between the data interpretation systems becomes more evident with low-level samples or samples involving a mixture of DNA contributors (Hares et al. 2020). For these types of samples, trained scientists must perform the interpretation. Scientists use previously validated thresholds and interpretation protocols to evaluate these challenging profiles. Additional software tools can perform mixture deconvolution; however, extensive DNA analysis training is required to correctly identify and enter profiles into these tools.

Typically, mixture interpretation should not be performed on profiles generated from rapid DNA systems. The reduced sensitivity and peak height ratio balance of rapid DNA systems can make detection of minor contributors more difficult than traditional methods. Rapid DNA samples that yield complex profiles will require interpretation and will be flagged by the quality system. If a more in-depth look at the data generated from a rapid DNA system is desired, a scientist trained in such interpretation methods should be consulted.

Troubleshooting

Traditional DNA processing is a step-by-step process that allows for detailed troubleshooting to identify if the problem is sample specific or can be attributed to a specific process, instrument, or reagent used in the workflow.

Rapid DNA, in contrast, is a fully automated process from extraction through to capillary electrophoresis and data analysis, which can make it more difficult to target the problematic step. There are only three main areas to troubleshoot: the instrument, consumables, and the sample. Issues are often identified by the instrument itself or by the outcome after a sample is processed. Certain instruments have internal diagnostics that alert the user to potential errors. When a specific code is displayed, the manufacturer should be consulted to assist with specific troubleshooting steps.

If there is no sample profile at the end of a run, the presence of a size standard in the resulting data would lead the user to believe that the run successfully completed, and the instrument is still likely to be functioning properly. In this case, it is possible the sample was at too low level (contained too little DNA) to produce a result. If the sample was expected to contain a high level of DNA, like a reference buccal sample or blood sample, then the problem could have occurred either in the transfer of sample through the cartridge or the instrument, or during a part of the process such as lysis or amplification. Depending on where these processes occur, i.e., whether it is on the cartridge or instrument, then the problem can be corrected by using a new cartridge or it may require an instrument repair.

Typically, troubleshooting problems that occur on rapid DNA instruments require the user to send log or data files to the manufacturer. Based on the review of log and data files, the manufacturer can potentially determine the problem and advise the user on the next steps.

Areas for Improvement and Future Development

The current rapid DNA systems have already made a significant impact on both the forensic and law enforcement communities and are poised to make an even greater impact in the future. However, there is still a need for further development to address the remaining performance gaps and barriers to adoption.

In July 2020, a letter to the editor of Forensic Science International: Genetics was published entitled "Rapid DNA for crime scene use: Enhancements and data needed to consider use on forensic evidence for State and National DNA Databasing – An agreed position statement by ENFSI, SWGDAM and the Rapid DNA Crime Scene Technology Advancement Task Group" (Hares et al. 2020). This document describes five major areas that should be addressed by manufacturers to improve rapid DNA analysis of forensic evidence samples:

1. An integrated method of human specific internal positive controls must be incorporated to identify low quantity, DNA degradation, and inhibition. This is

intended to address the human quantification requirements shared by SWGDAM and many of the member countries of the European Network of Forensic Science Institutes (ENFSI).
2. It must be possible to export analyzable raw (optical preprocessed) data. This ensures the availability of data for analysis or reanalysis by qualified forensic DNA scientists if needed for mixture interpretation and court purposes.
3. An onboard fully automated expert system must be incorporated and programmed with rules to accurately flag allele calls in both single-source and mixture data that require analyst evaluation.
4. Improved peak height ratio balance (per locus and across loci) for low-quantity and mixture samples (high and low quantity) must be achieved to facilitate interpretation of DNA mixtures and low-level samples by qualified forensic DNA scientists.
5. Rapid DNA manufacturers must perform a well-defined publicly available developmental validation on a wide variety of forensic evidence-type samples commonly encountered in the forensic DNA laboratory.

It should be noted that several of these areas have already been addressed. For example, it is already possible to export raw data files (#2) for several on-market rapid DNA systems, and extensive developmental validations have been published by manufacturers (#5). In addition, expert system software (#3) is already incorporated into rapid DNA systems as described earlier in this chapter; however, it is recognized that further optimization of these software systems could increase automation in order to significantly reduce the amount of manual data review required by trained forensic scientists.

Additional development is clearly required to address area #1 (internal positive controls) and #4 (improved peak height ratio balance). It is interesting to note that while there is wide recognition of the need for increased sensitivity with rapid DNA systems in order to facilitate improved detection of low-level DNA samples, the authors of this article specifically focused on improved peak height ratio balance. Thus, a preference for enabling enhanced mixture interpretation as opposed to solely focusing on increased sensitivity can be inferred.

Once the areas above have been developed and validated, rapid DNA systems will in many respects possess performance capabilities that are on par with the current traditional laboratory methods. Still, there are several additional areas for potential improvement, as outlined below.

Cost

Presently, it is more expensive to process a sample using a rapid DNA system versus traditional bench processing, when only considering the cost of reagents. For some applications, this is not prohibitive, as rapid DNA processing presents cost savings when considering all factors. For example, the Kauai Police Department reported that implementation of rapid DNA technology saved the agency significant time and

several thousand dollars in one case, projecting that it will save over ninety thousand dollars annually (Thermo Fisher Scientific "Kaua'i" 2020c).

For broader adoption, however, especially within the laboratory environment, rapid DNA processing is costly for routine STR analysis. Decreasing the cost of instrumentation and the cost per sample will likely be an important factor in wide-scale adoption.

Flexible Throughput

The current rapid DNA instruments can run between one and seven samples at a time. For the instruments that enable multi-sample runs, the user may feel obliged to wait until they have gathered sufficient samples to run the maximum, or risk wasting costly consumables. This can defeat the purpose of possessing "rapid" analysis capabilities and is in stark contrast to traditional DNA laboratories, which can batch samples to increase throughput and reduce costs simultaneously. Future rapid DNA systems will need to evolve to enable more flexible throughput capabilities.

Size

While current systems are portable as discussed above, they range from about 50 pounds to over 100 pounds, with the larger systems requiring more than one person to move. If the devices were smaller, lighter, and internally battery-powered, this would further enhance portability and opportunities for field deployment.

Speed

Although generating an STR profile in under 2 h is fast, producing results even faster does present advantages. Systems currently process between one and seven samples per run and decreasing the time to generate a result would increase the potential throughput per instrument. Increasing the number of samples per run would also accomplish this, but the advantage gained may be mitigated by decreased portability and increased cost, as discussed above.

Expanded Testing Capabilities

To date, rapid DNA instruments have focused on producing autosomal STR profiles that contain the same genetic loci that are used in CODIS and similar databases, to facilitate upload and searching. There is, however, benefit in expanding the scope of markers tested with rapid DNA.

For example, Y-STR profiles can be used to identify the presence of male DNA even when there is an overwhelming amount of female DNA present. Many laboratories use Y-chromosome-based screening on sexual assault samples, as it is a quicker way to determine the presence of male DNA than chemical identification and differential extraction. If no male DNA is detected with a "Y-screen" method, the case is deemed negative, and no further testing is performed.

If rapid DNA technology could also generate Y-STR profiles, this could greatly aid in the screening of sexual assault kits. Instead of sending sexual assault kits directly to the laboratory, they could be screened by the submitting agency or even at the hospital where the sexual assault kit was collected, thus increasing the efficiency of downstream testing performed by DNA laboratories. However, an increase in sensitivity would also be necessary for rapid DNA systems to utilize this method and enable the operator to identify true negatives confidently.

Conclusion

Rapid DNA testing is still in the early stages of adoption, yet it has already made significant impacts in the identification of disaster victims and proven the ability to solve crimes in hours rather than days, weeks, months, or even years. As rapid DNA testing becomes more widespread, it is destined to become a dominant force in forensic science. Some are hesitant to put this power into the hands of nonscientists, and others are concerned that rapid DNA testing may replace traditional DNA testing methods. However, the most likely near-term progression will be that an increasing percentage of simpler cases, as well as less violent offenses such as property crimes, will be solved quickly using this technology, while more complex cases will continue to be the focus of forensic scientists using conventional methods.

Advancements in rapid DNA technology will continue to contribute to the field of forensic science, enhancing the power of forensic DNA testing and databases. Looking further ahead, it may one day replace the current laboratory methods as the primary routine workflow.

References

Anker AST et al (2021) The effects of DNA databases on the deterrence and detection of offenders. Am Econ J Appl Econ. https://www.aeaweb.org/articles?id=10.1257/app.20190207&&from=f. Accessed 15 Mar 2021

Arizona Department of Public Safety Scientific Analysis Bureau. Rapid DNA Law Enforcement Program. https://www.azdps.gov/sites/default/files/media/Rapid%20DNA%20Law%20Enforcement%20Program%20Flyer_1.pdf. Accessed 15 Mar 2021

Butler JM (2011) Advanced topics in forensic DNA typing: methodology. Academic Press: San Diego

Carney C et al (2019) Developmental validation of the ANDE™ rapid DNA system with FlexPlex™ assay for arrestee and reference buccal swab processing and database searching. Forensic Sci Int Genet 40:120–130. https://doi.org/10.1016/j.fsigen.2019.02.016

Dadhania A et al (2020) The power of rapid DNA results. Forensic Magazine. https://www.forensicmag.com/564692-The-Power-of-Rapid-DNA-Results/. Accessed 15 Mar 2021

Della Manna A et al (2016) Developmental validation of the DNAscan™ Rapid DNA Analysis™ instrument and expert system for reference sample processing. Forensic Sci Int Genet 25: 145–156. https://doi.org/10.1016/j.fsigen.2016.08.008

DNA Resource. Global resources. http://www.dnaresource.com/resources.html. Accessed 15 Mar 2021

Giese H et al (2009) Fast multiplexed polymerase chain reaction for conventional and microfluidic short tandem repeat analysis. J Forensic Sci 54:1287–1296

Hares DR et al (2020) Rapid DNA for crime scene use: enhancements and data needed to consider use on forensic evidence for State and National DNA Databasing – an agreed position statement by ENFSI, SWGDAM and the Rapid DNA Crime Scene Technology Advancement Task Group. Forensic Sci Int Genet 48:102349. https://doi.org/10.1016/j.fsigen.2020.102349

Hennessy LK et al (2014) Developmental validation of the GlobalFiler® express kit, a 24-marker STR assay, on the RapidHIT® System. Forensic Sci Int Genet 13:247–258. https://doi.org/10.1016/j.fsigen.2014.08.011

Horsman KM, Bienvenue JM, Blasier KR, Landers JP (2007) Forensic DNA analysis on microfluidic devices: a review. J Forensic Sci 52(4):784–799

Jovanovich S et al (2015) Developmental validation of a fully integrated sample-to-profile rapid human identification system for processing single-source reference buccal samples. Forensic Sci Int Genet 16:181–194. https://doi.org/10.1016/j.fsigen.2014.12.004

Manzella AM, Moreno LI (2020) Assessing the impact of using conventional swabs on the ANDE 6C arrestee biochip. Forensic Sci Int Genet 48:102358. https://doi.org/10.1016/j.fsigen.2020.102358

Moreno LI, Brown AL, Callaghan TF (2017) Internal validation of the DNAscan/ANDE™ Rapid DNA Analysis™ platform and its associated PowerPlex® 16 high content DNA biochip cassette for use as an expert system with reference buccal swabs. Forensic Sci Int Genet 29:100–108. https://doi.org/10.1016/j.fsigen.2017.03.022

Murakimi C et al (2020) Individual identification using the RapidHIT™ ID system for forensic samples. Legal Med 47:101776. https://doi.org/10.1016/j.legalmed.2020.101776

O'Brien R, Barnhart T (2021) One Swab, two uses: rapid and conventional DNA testing. Poster presentation at the 2021 meeting of the American Academy of Forensic Sciences, National Forensic Science Technology Center. https://www.aafs.org/common/Uploaded%20files/Meetings/2021%20Meeting/AllPosters.pdf

Plasencia A (2017) Miami Beach Police using rapid DNA Technology to help solve crimes faster. NBC Miami. https://www.nbcmiami.com/news/local/miami-beach-police-using-rapid-dna-technology-to-help-solve-crimes-faster/21798/. Accessed 15 Mar 2021

Ragazzo M et al (2020) Comparative analysis of ANDE 6C rapid DNA analysis system and traditional methods. Genes 11(5):582. https://doi.org/10.3390/genes11050582

Rapid DNA Used by Prosecution (2014) Evidence technology magazine. https://www.evidencemagazine.com/index.php?option=com_content&task=view&id=1581. Accessed 15 Mar 2021

Romsos EL et al (2020) Results of the 2018 rapid DNA maturity assessment. J Forensic Sci 65: 953–959. https://doi.org/10.1111/1556-4029.14267

Salceda S et al (2017) Validation of a rapid DNA process with the RapidHIT® ID system using GlobalFiler® express chemistry, a platform optimized for decentralized testing environments. Forensic Sci Int Genet 28:21–34. https://doi.org/10.1016/j.fsigen.2017.01.005

Scientific Working Group on DNA Analysis Methods (SWGDAM) (2017) Interpretation guidelines for autosomal STR Typing by forensic DNA testing laboratories. https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_50e2749756a242528e6285a5bb478f4c.pdf. Accessed 15 Mar 2021

Thermo Fisher Scientific (2020a) Bone sample processing on the RapidHIT ID System with RapidINTEL cartridges. https://assets.thermofisher.com/TFS-Assets/GSD/Application-Notes/

Bone-sample-processing-RapidHIT-ID-system-RapidINTEL-cartridges-application-note.pdf. Accessed 15 Mar 2021

Thermo Fisher Scientific (2020b) The Italian Carabinieri achieves success using rapid DNA technology. http://assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/italian-carabinieri-rapid-dna-customer-profile.pdf. Accessed 15 Mar 2021

Thermo Fisher Scientific (2020c) The Kaua'i Police Department uses rapid DNA technology to save time and cost in disaster victim identification. http://assets.thermofisher.com/TFS-Assets/GSD/Reference-Materials/Kauai-police-dna-technology-disaster-victim-identification.pdf. Accessed 10 Mar 2021

Thong Z et al (2015a) Evaluation of the RapidHIT™ 200 System: A comparative study of its performance with Maxwell® DNA IQ™/Identifiler® Plus/ABI 3500xL workflow. Forensic Sci Int Genet 19:22–27. https://doi.org/10.1016/j.fsigen.2015.05.006

Thong Z et al (2015b) Investigative leads from DNA: Casework experience from the IntegenX RapidHIT™ 200 System. Forensic Sci Int Genet Suppl Ser 5:e69–e70. https://doi.org/10.1016/j.fsigss.2015.09.028

Turingan RS et al (2020) Developmental validation of the ANDE 6C system for rapid DNA analysis of forensic casework and DVI samples. J Forensic Sci 65(2020):1056–1071. https://doi.org/10.1111/1556-4029.14286

U.S. Congress (2017) H.R.510 – Rapid DNA Act of 2017. https://www.congress.gov/bill/115th-congress/house-bill/510/all-info. Accessed 9 Mar 2021

U.S. Federal Bureau of Investigation (2019) Non-CODIS rapid DNA Considerations and Best Practices for Law Enforcement Use. https://www.fbi.gov/file-repository/non-codis-rapid-dna-best-practices-092419.pdf/view. Accessed 15 Mar 2021

U.S. Federal Bureau of Investigation (2020a) National Rapid DNA Booking Operational Procedures Manual. https://www.fbi.gov/file-repository/national-rapid-dna-booking-operational-procedures-manual-eff-090120.pdf/view. Accessed 15 Mar 2021

U.S. Federal Bureau of Investigation (2020b) Standards for the operation of rapid DNA booking systems by Law Enforcement Booking Agencies. https://www.fbi.gov/file-repository/standards-for-operation-of-rapid-dna-booking-systems-by-law-enforcement-booking-agencies-eff-090120.pdf/view. Accessed 15 Mar 2021

U.S. Federal Bureau of Investigation (2021) Rapid DNA. https://www.fbi.gov/services/laboratory/biometric-analysis/codis/rapid-dna. Accessed 9 Mar 2021

Vallone PM, Hill CR, Butler JM (2008) Demonstration of rapid multiplex PCR amplification involving 16 genetic loci. FSI Genetics 3(1):42–45

Wiley R et al (2017) Internal validation of the RapidHIT® ID system. Forensic Sci Int Genet 31:180–188. https://doi.org/10.1016/j.fsigen.2017.09.011

The Interpretation of Mixed DNA Samples 45

Historical Perspective and Current Developments

Francesco Sessa, Monica Salerno, and Cristoforo Pomara

Contents

Abstract

The interpretation of mixed DNA samples represents a very compelling research field for the forensic community. Modern short tandem repeat (STR)-based kits can sequence several markers with a high degree of diversity; in the same way, the analysis of STR sequence variation in sample population groups highlighted that several current STR loci lack the necessary sequence diversity. Moreover, in the past few years, new forensic DNA profiling methodologies have been developed based on the use of massively parallel sequencing (MPS), also referred to as next-generation sequencing (NGS).

The current human STR kits use markers with different heterozygosities, showing some highly heterozygous loci, while others are only marginally heterozygous. New methods offer the opportunity to enhance human identification for forensic purposes in DNA mixtures: using new multiplex kits with highly

F. Sessa (✉)
Department of Clinical and Experimental Medicine, University of Foggia, Foggia, Italy
e-mail: francesco.sessa@unifg.it

M. Salerno · C. Pomara (✉)
Department of Medical, Surgical and Advanced Technologies "G.F. Ingrassia", University of Catania, Catania, Italy
e-mail: monica.salerno@unict.it; cristoforo.pomara@unict.it

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_47

heterozygous STRs, it is possible to solve more effectively the interpretation of many complex mixture samples. Regarding these concerns, MPS methods have improved the discrimination power by investigating genetic diversity through the analysis of both STRs and single-nucleotide polymorphisms (SNPs) located near the flanking region.

Nevertheless, to date, the mixtures generated by more than two contributors (three or more persons) remain frequently unsolved because of the difficulties in the evaluation of stochastic effects and/or other technical problems, such as the differences between true alleles and stutter products. For these reasons, future studies are needed in order to overcome both these technical and interpretation difficulties.

Finally, it is important to remark on the importance of statistical methods to establish the "weight of the evidence," particularly for the interpretation of MPS multiplex kits for forensic DNA analysis. The interpretation of MPS results remains another interesting research field for the forensic community, particularly in the application of statistical methodologies.

Keywords

Mixed DNA samples · Short tandem repeats · Mixture and complex profile · Probabilistic genotyping · Massive parallel sequencing · Next-generation sequencing

Introduction

In the first half of the 1980s, Jeffreys used restriction enzymes that cleaved DNA while investigating hereditary diseases, pointing out the possibility of fragmenting DNA samples. Applying this experimental technique, small and large DNA sections were obtained. Using a radioactive probe at low stringency combined with electrophoresis, exposing a film directly on the gel, an image similar to a barcode was obtained (Jeffreys et al. 1985). Based on this evidence, Jeffreys understood the potential use of identifying a subject responsible for a crime: "a single barcode for a single individual." A few years later, this profiling method was applied for the first time to solve two homicide cases in the forensic field; particularly, it was used to exonerate an innocent person and, only after a mass screening, to identify the guilty subject (Hecht 1989). This case may be considered the start of the forensic application of DNA analysis. A few months later, this technique was purposed for forensic applications. Demonstrating somatic and germline stability combined with individual specificity, the DNA fingerprints produced by Southern blot hybridization have suggested the use of this application as a key element to revolutionize forensic biology, particularly with regard to the identification of suspects in different crimes, such as rape, robbery, and homicide (Gill et al. 1985). DNA fingerprinting, at that time known as the MultiLocus Probe (MLP) technique, demonstrated its great power to both implicate and exonerate individuals as offenders in criminal cases; MLPs

were compared visually, rather than relying on any agreed statistical methods. Matching patterns were compelling, and if sufficient numbers of bands were present, most practitioners reported that the evidence conclusively identified the samples as coming from the same individual (Aronson 2007).

In the early 1990s, several criticisms and disagreements were moved to this technique by the forensic community, generating the so-called DNA wars. For example, problems were found with cases where only some of the MLP bands matched (such as cases involving relatives) and the application of the method to mixed samples or paternity work. There was also the issue of whether individual MLP bands were present in the profile or not. Another disadvantage of the MLP technique was that relatively large biological samples were required to give a reliable result. Subsequently, the single locus probe (SLP) technique was introduced in the UK in around 1988. In forensic science, DNA analysis of restriction fragment length polymorphisms (RFLP) has become the most powerful method. Almost all laboratories working in the field of stain analysis prefer SLPs, because these probes have a better sensitivity than MLPs and offer the possibility to build an allele frequency database. By using specific probes to target different regions of DNA, different chromosomes were typed. Each probe produced one or two bands, called homozygous and heterozygous, respectively, depending on whether the versions of the DNA sequence that an individual inherited from his two parents. Based on this theory, different methods were described (Gill and Werrett 1990; Budowle et al. 1991).

In the second phase of the analysis, a database was required to estimate the relative frequency in the population of each of the corresponding SLP bands. Based on the assumption that the population was approximately in a Hardy-Weinberg equilibrium (HWE), which means that the bands are inherited independently within the loci, and using about 4 different SLP probes, the method was able to produce very strong evidence with matching probabilities in the range of 1 in several millions. The probability of matching is an estimate that non-correlated members of the population would have the same profile as an SLP. Although different statistical tests were described (Evett and Gill 1991; Evett and Pinchin 1991), several critical issues were raised: the dimension and representativity of the databases, predetermined band frequency, independent data in the different loci, measurement and laboratory errors, and management of relatives and mixed DNA samples from more than one person.

Berry et al. (1992) and Evett et al. (1993) described a more impartial and continuous approach to the evaluation of SLP evidence. This avoided the necessity of a two-step comparison; however, the evidence was reported as a likelihood ratio (LR) rather than a random match probability (RPM). The appropriateness of the likelihood of concordance and LR values resulting from the above methods has been proven by a large-scale experiment performed by Risch and Devlin (1992) and Lambert et al. (1995), involving comparisons among millions of people. In the 1990s, the importance of statistical analysis was stressed, describing different methods on which to base the interpretation of DNA profiles (Morton 1997; Thompson 1997).

Technological changes since those times have been considerable. DNA profiling systems were further improved by the development of the short tandem repeat (STR) method in 1993. This followed the pioneering work of Mullis who reported a new technique (called PCR) that uses small amounts of DNA to be typed. STR regions typically involve a special sequence of nucleotides that are repeated about 5–30 times. There are many hundreds of these STR regions in the human genome, but several sequences are sufficiently efficient for providing very high levels of human discrimination.

The loci targeted for forensic analysis are referred to as short tandem repeats, or STRs. STRs are discrete sequences of DNA that are repeated end to end. They can be repeated in tandem up to 100 times within the genome. The sequence of an STR consists of two to five nucleotides, and these are termed di-, tri-, tetra-, and penta-nucleotide repeats. Tetranucleotide repeats are the most common type used in forensic DNA profiling (Fig. 1).

The obtained STR profiles look like a series of peaks on a graph; each peak is labeled with a special number according to its position on the graph. This number represents the number of times that the core sequence is repeated. These numerical markers are known as alleles. The STR-PCR technique has provided some distinct advantages over the SLP system: the size of STR peaks can be accurately and unequivocally determined. Therefore, the STR peaks are simply shown as a number series ("alleles"), two for every locus inherited from each parent. This facilitated the creation of a complete intelligence database containing the profiles of all typed subjects; moreover, it amplified several STR regions, analyzing smaller DNA samples compared to SLPs; finally, in the human genome there is a large number of STR regions; therefore, there is no theoretical limitation to the power of the technique.

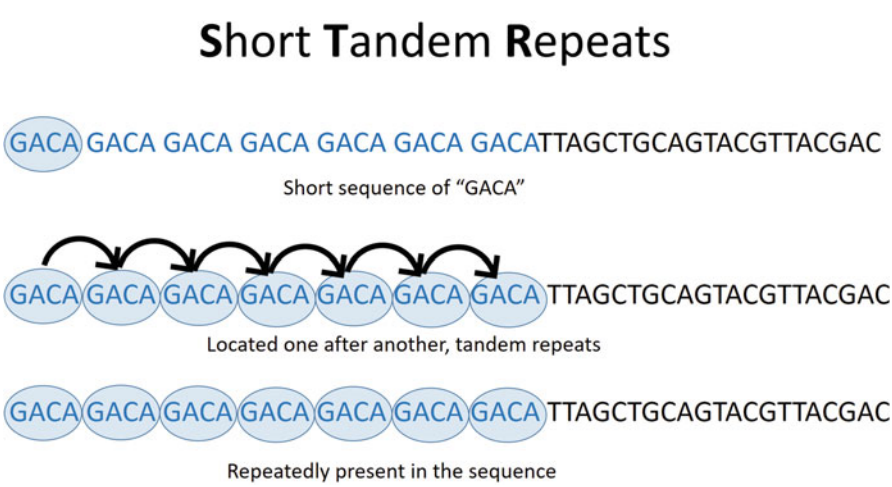


Fig. 1 Summary of the main STR characteristics

The first STR kit was developed using 4 STRs, followed by a growing system that could analyze up to 26 STRs (Gill and Evett 1995; Vraneš et al. 2019). Modern techniques have also greatly improved the speed at which a DNA profiling result can be obtained. High-throughput laboratories now have the capacity to generate a plethora of profiling results in just a few hours. More data have meant increased pressure for efficient and effective interpretation methods. Multiplexing is a rapid and convenient way to type many STRs and generate a DNA profile, also maximizing the amount of information obtained from forensic DNA samples, which can be limited in quantity and/or compromised in quality. The STR technique has not led to new statistical questions in forensic cases. However, the question persisted as to which databases were used to estimate the proportions of the alleles. Different authors argued about the bias involved in these methods. Basically, the process is flawed because it calls for the idea of proving a null hypothesis of perfect independence (both within and between loci) that cannot be true in a real population (Buckleton et al. 2001). Therefore, this is obtained via a Bayesian estimation of allele proportions using a binomial or trinomial probability and uniform priors (Foreman et al. 1997).

The rapid and progressive evolution of analytical techniques and the advent of next-generation sequencing (NGS) have completely revolutionized the DNA sequencing approach. Other markers were suggested as an alternative to STRs, such as single-nucleotide polymorphisms (SNPs) and insertion/elimination polymorphisms, for the identification of individuals. These types of markers have some advantages compared to STRs, such as their smaller size, high discriminatory power, and the fact that they are efficient for typing degraded samples (Oldoni and Podini 2019). Even though this method is widely used in different diagnostic fields, to date, there is no general agreement on the interpretation rules of forensic data produced via NGS technology. Moreover, they are not commonly used by forensic labs, which rely on historical databases of STR profiles, often containing thousands of known individuals to identify criminals.

While the gold standard of capillary electrophoresis (CE) in forensic DNA typing is robust and reliable, the ability to determine the number and appropriation of contributors in DNA mixture profiles remains a challenge.

Mixture and Complex DNA Profiles

In the 1990s and the first decades of the 2000s, new techniques revolutionized the progress made in the fields of processing and interpreting of data. More sensitive DNA profiling methods generate a profile even if the quality and quantity of DNA are poor. STR typing relies primarily on polymerase chain reaction (PCR) coupled with capillary electrophoresis (CE), where STRs are amplified using extensively validated primer-based multiplexes and separated by size, resulting in precise allele designations on the basis of DNA fragment length (Lohmueller et al. 2014; Ludeman et al. 2018). In forensic genetics, current STR genotyping kits are able to amplify starting from 9 to 27 forensic relevant STRs (Ludeman et al. 2018).

To date, several STRs are currently used worldwide in forensic cases and are chosen, considering their characteristics and their human genome distribution (Moretti et al. 2016). The current CODIS core loci are composed of 20 STRs: D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, CSF1PO, FGA, TH01, TPOX, vWA, D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, and D22S1045. Moreover, the STRs are chosen considering their inherited characteristics: specifically, they should be independent. It is possible to obtain a profile starting from different sample types, increasing both DNA detection from multiple contributors (called mixed DNA profiles) and the determination of incomplete (partial) and degraded profiles (Fig. 2). For all of these reasons, profile interpretation became very complex and challenging.

As shown in Fig. 3, more than two allelic peaks are found for each locus representing a mixed profile. In general, a mixed profile appears to be degraded with the height of the peak decreasing through the profile as molecular weight increases.

Moreover, as shown in Fig. 3, one allele is detected at the higher-molecular-weight loci (such as D2S1338): no other unlabeled peaks are visible, even though this could be related to the quantity and quality of the DNA. Furthermore, in the D16S539 locus, it may be difficult to distinguish peaks originating solely from stutter from composite peaks, with both allelic and stutter contributions, or from

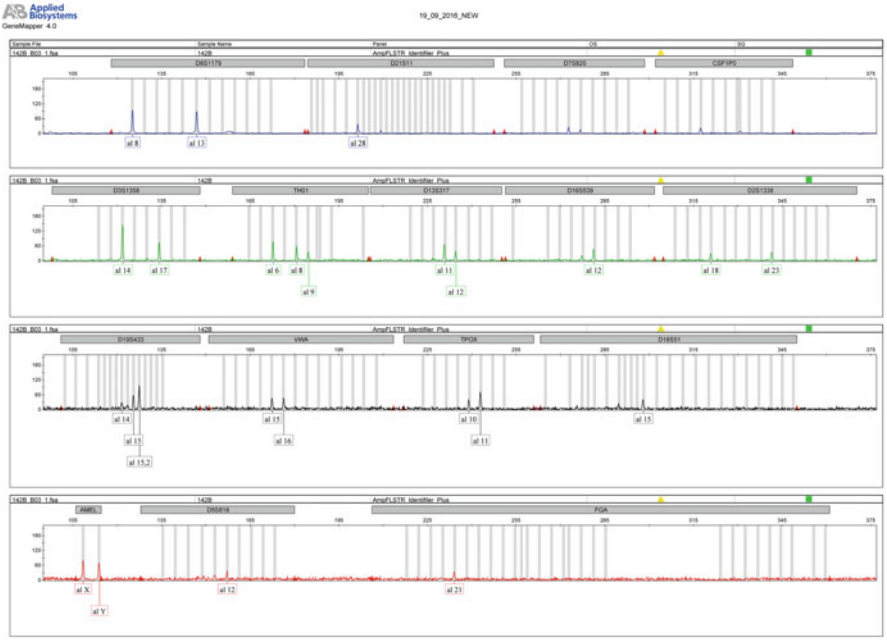


Fig. 2 The interpretation of incomplete (partial) profiles is challenging for the forensic examiner

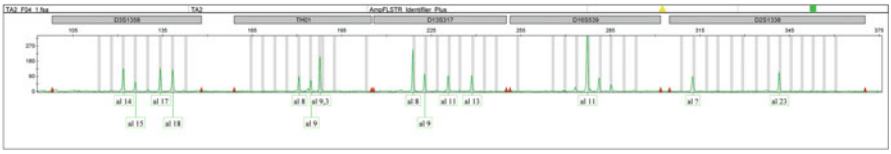


Fig. 3 Mixed profile: considering that no more than four alleles were found for each locus, it could be thought that this is composed of two contributors

allelic peaks originating from a minor contributor. As previously described, at the high molecular weight, it is possible that there is the "dropout" phenomenon.

A further difficulty that can occur, especially when using increased sensitivity techniques, is the presence of peaks that are extraneous to the profile. These usually occur in the profile as single peaks (or rarely as more than one peak), characterized by low peak heights, which are referred to as drop-ins. A drop-in is believed to originate when small fragments of human DNA ubiquitously found in the environment are introduced into the DNA sample during PCR setup. Drop-in is not replicable (except by accident), and it is impossible to establish its sources. On the contrary, the contamination phenomenon is typically manifested as reproducible foreign DNA within the worked profile and can often be attributed to a particular source (i.e., the analyst who processed the sample in the laboratory).

When a biological trace is recovered from a crime scene, one of three general scenarios can be reconstructed: the comparison is inconclusive; the person of interest (POI) can be excluded as a contributor to the DNA profile obtained from the crime scene evidence; the POI cannot be excluded as a possible contributor to the complex profile obtained from the evidence.

This last scenario is very interesting and it appears necessary to supply the weight of this evidence. For this purpose, statistical analysis should be performed: the weight of the evidence is strictly related to the number of detected and included loci in the evidence profile and to the rarity of the observed alleles.

Several important organizations, such as the Scientific Working Group on DNA Analysis Methods (SWGDAM), remark in their guidelines on the importance to perform statistical analysis to support inclusion evidence. Moreover, another important aspect is related to the procedures that should be followed for qualifying the significance of associations in the report, whether by a statistic or a qualitative statement (SWGDAM 2010; Moretti et al. 2017).

The most important methods used to evaluate the weight of evidence are the combined probability of inclusion (CPI) and the random match probability (RMP). CPI represents the probability that a person could be included as a DNA profile contributor; the RMP is the probability that a random person could have the same genotypes found in the DNA profile. Obviously, these tests are perceived as simple and easily explained during a juridical process. In cases of mixture samples generated by multiple contributors, the DNA profiles may contain several alleles in each locus and in turn the DNA profiles are made more complex. As a result, the combinatory power of the mixed DNA profile using the product rule becomes

compromised as the complexity of the mixed DNA profile increases. In particular, RMP works best for single source DNA profiles and DNA mixtures that can be resolved into unique components (i.e., mixed DNA profiles with clear major and minor contributors that can be interpreted independently) (Bille et al. 2013).

Another commonly used method is the likelihood ratio (LR): this ascertains the probability of two mutually exclusive propositions; in particular, it represents the ratio between the prosecution hypothesis (Hp) and the defense hypothesis (Hd): this method is frequently applied in real forensic cases and it is well-supported by the forensic community. As previously described, considering that recent forensic kits have a high limit of detection, the interpretation of a DNA profile could be very complex. Particularly, several forensic laboratories continued to use inadequate methods, both procedurally and statistically, and these situations generated several failures at important trial procedures.

Several significant papers were published to describe inter- and intra-laboratory inconsistency, remarking on the importance to follow well-established guidelines in order to avoid or reduce procedural bias. In 2011, Dror and Hampikian reported an exercise studying the variability in DNA profile interpretation (Dror and Hampikian 2011). In their study, the authors underlined intra-laboratory variability: this conclusion pointed out the so-called "fallibility of DNA evidence," which could influence the decision about suspected people (Geddes 2010).

In a recent publication, interlaboratory DNA methods were also discussed (Butler et al. 2018). Even though the study presented several criticisms, it could be considered an important example to highlight the variability in DNA interpretation. Due to subjectivity in the interpretation of the guidelines, several different interpretations were made by different scientists involved in the evaluation of the evidence (Butler et al. 2018). In order to better define the weight of forensic evidence, the forensic scientist should apply a probabilistic expert system particularly in the case of DNA mixtures originating from more than one person. DNA mixtures are currently the most common challenge in forensic genetics casework. However, mixed DNA profiles involving more than two contributors are more complex than two-person mixtures.

Figure 4 shows that the presence of more than two alleles in one or more loci represents the so-called mixed profile. When using the STR technique, it is very important to evaluate several characteristics of the peak, such as the peak areas: for example, when the mixture is composed of two contributors, it could be possible to distinguish the two alleles of one contributor from the peak height, expecting that they are similar in both height and area; in contrast, excessive differences could show different contributors.

On the contrary, when the mixture is composed of more than two contributors, the probability of solving the mystery behind the profile decreases. Obviously, several parameters are more important in order to solve the mixture: first, the quality and the quantity of DNA and, second, the choice of the DNA typing kit and the instrument for the capillary electrophoresis with the relative parameters are both of utmost relevance (Gill et al. 2015; Butler et al. 2018). These kinds of profiles show multiple alleles for each locus, and in several cases, it may be possible to find the profile of an unknown person who has only handled the evidence before or after the crime (Fig. 5).

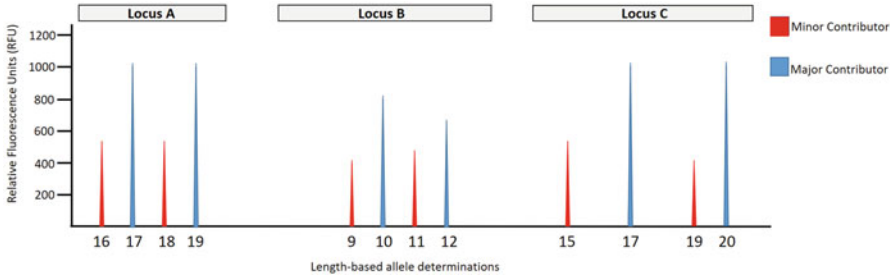


Fig. 4 Simulation of a two-person resolvable mixture: the peaks might may be useful to distinguish major and minor contributors

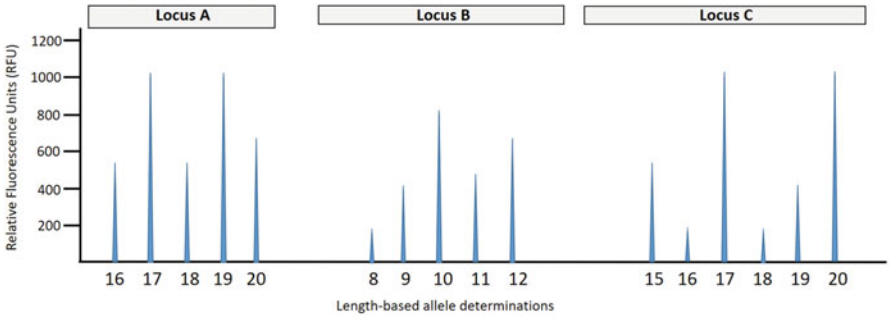


Fig. 5 When the contributors to the mixture are more than two persons, it is hard to distinguish the real profile. Moreover, it is very difficult to understand if several peaks are realistic or artifacts such as stutter (e.g., in locus B, allele 8)

These factors in tandem may generate profiles hard to be deconvoluted in the different contributors' genotype, reducing the weight of DNA typing and leaving cases unsolved.

Issues Surrounding the LCN Technique

As previously stated, the current forensic techniques are able to detect a complete profile working with a low copy number (LCN) DNA. LCN typing, particularly for current short tandem repeat (STR) typing, refers to the analysis of any sample that contains less than 200 pg of template DNA (Budowle et al. 2009). These techniques are extremely sensitive and enable profiles to be obtained from only a few cells of starting DNA.

In this context, it is important to highlight the possibility to transfer skin cells with the relative profile under different circumstances (Phipps and Petricevic 2007; Sessa et al. 2019); the adequate interpretation of LCN profiles should take into account modeling contamination, the possibility to obtain the so-called drop-in of spurious alleles, and also the reproducibility of DNA profiles. Concerning this last issue, LCN

analysis is by nature not reproducible: for this reason, it cannot be considered as robust as that associated with conventional DNA typing. Nevertheless, LCN typing is commonly used for the identification of missing persons and human remains and for developing investigative leads: in this scenario, caution should be taken with its use in other endeavors until developments are made that overcome the vagaries of LCN typing.

The most important limitation in the case of LCN analysis is related to a great degradation of DNA, both because of the limited number of cells and because of the storage conditions of the source. Therefore, the obtained profile can be characterized by an incomplete representation of alleles (often called "allelic drop-out"). Indeed, in different loci, a single allele could be observed, but it is not possible to determine if the subject is homozygote or if it is a dropout phenomenon. Moreover, to allow for a complete interpretation, several parameters are frequently modified, such as the baseline value on the graph; contrariwise, this choice generates imbalances in the interpretation of the profile, particularly in heterozygote alleles. Working with an LCN profile, the more complex biases are related to the undetected loci (no peaks) or when the peak is very difficult to assess.

Moreover, the possibility of a stochastic contamination event should be considered: inserting a few exogenous cells into the reaction can generate the presence of other additional alleles. This contamination could be traced back to degraded human epithelial cells (i.e., dust particles), referred to as "Touch DNA," or it may itself be the product of an earlier PCR amplification. Genetic profiles generated from fingermarks were first described in 1997 (van Oorschot and Jones 1997). As regard the so-called Touch DNA, there is a growing interest in it by the forensic community. Several important aspects of "Touch DNA" have been studied such as the "handling time," analyzing the contact time needed to deposit a sufficient amount of DNA on a garment to produce an interpretable profile, or the technique that guarantees the best recovery (Sessa et al. 2019). Moreover, this kind of DNA has been investigated under different scenarios in order to better clarify several aspects linked to the deposition or transferring conditions (Gosch et al. 2020; Neckovic et al. 2020; Sakurada et al. 2020).

To reduce all of these biases, a consensus profile approach was suggested (Whitaker et al. 2001): the reproducibility of the allele is taken to be present in the profile. Obviously, the consensus profile restricts the designation of the artifacts and the potential dropout or drop-in of alleles, improving the statistical analysis of the obtained mixed profiles.

Probabilistic Genotyping

Statistical evaluation is very important in order to provide an estimate of the number of people who would match the crime profile in a specified population. In fact, the improvement of the STR methods poses a number of relevant questions to issues that may be critical for the defense and can include the following: how and when the DNA was deposited, its relevance to the alleged crime, establishing continuity of exhibits and integrity of the evidence, and, ultimately, laboratory error.

One of the first methods used to evaluate the mixture profile is related to the calculation of the probability of exclusion (PE) (Devlin 1993). This case is represented by the event of the allele not being shared in the mixture profiles. Obviously, the main limitation of this proposed approach is the presence of the dropout event, particularly in the case of degraded DNA. To improve this approach, new methods were proposed that considered the mixture profile of a subject involved in the crime as a contributor (e.g., the victim's profile): this profile could be very useful in order to ascertain the unknown profile (Devlin et al. 1990).

As previously described, the use of modern forensic kits has improved the possibility to obtain LCN profiles, which, unfortunately, has increased the background noise in a profile. Several experimental models have been proposed in order to solve these difficulties, describing different scenarios and the relative software or approaches (Weir et al. 1997; Curran et al. 1999; Fung and Hu 2002; Fukshansky and Bär 2000).

All of the LR approaches discussed above are based solely on which alleles or peaks are present in the mixture. Therefore, if we consider the case of a two-person mixture, where "a" denotes the vector of alleles observed at a locus, "G" denotes the genotypes of the two contributors, and "H" denotes the hypothesis, the LR is then composed of terms of the type:

$$p(\underline{a}|\underline{G}, H)\Pr(\underline{G}|H)$$

As previously described, to interpret an STR mixture profile, several parameters should be evaluated, such as the peak intensities or areas, because these data could supply the indication of the amount of DNA being contributed by the allele to the mixture. The areas are usually indicative as to the relative DNA contributions; to date, forensic scientists may use this information to assign the value of "0" or "1" to the possible combination of contributors. This information is summarized in several papers that describe this approach in the deconvolution of two-person mixtures (Clayton et al. 1998; Gill et al. 1998; Gill 2002).

One of the first methods to analyze the mixture profile (two-person mixed profile) was provided by Evett et al. (1998). This method considered the peak area evidence more formally by modeling ratios of peak areas using normal probability distributions. A few years later, Balding (1999) indicated a new recursive formula to interpret the evidence considering the population substructure, although this approach can only be applied to the mixture of two contributors. In 2001 (Perlin and Szabady 2001), a new deterministic approach was applied to identify the contributor in mixture profiles using an intelligence database. This is reached by a process of optimization based on a least-squares-type metric following the removal of the random PCR processes that act on STR profiles. Although providing an adequate solution for intelligence purposes, this is wholly inappropriate once the evaluative stage is reached. Modern forensic technologies improved the DNA profile results; in the same way, the complexity of profile interpretation became very difficult and demanded more complex methods to correctly validate the results. Starting in 2010, the method required to validate the results have been the so-called "probabilistic genotyping (PG)." PG evaluates several different parameters

such as peak height, and peak area, using specific software based on mathematical processes. This software increased both the objectivity and the weight of the DNA evidence, reducing the analyst's subjective interpretation as well as improving the overall interpretation of the DNA profile. The PG software is based on the LR (1), meaning the ratio of two mutually exclusive hypotheses (two conditional probabilities).

$$\mathrm{LR} = \frac{\Pr\left(O|H_p, I\right)}{\Pr\left(O|H_d, I\right)} \tag{1}$$

In this formula, "O" indicates the "observation"; in other words, it represents the DNA findings, "Hp" is one proposition (in general, it represents the prosecution theory including the POI), "Hd" is the other proposition (in general, it represents the defense theory excluding the POI), while "I" represents the other information named "framework of circumstances."

To implement the PG, another form is used (2):

$$\Pr\left(O|S_j\right) = \int_M p\left(O|S_j, M\right) \mathrm{d}M \tag{2}$$

In this equation, "Sj" is the possible genotype set that may be included in the profile, and "M" reflects the parameters used to obtain the DNA profile. These parameters are commonly defined as "weights," and different approaches may be applied to improve the value of these terms.

The most used approach is a "binary" interpretation approach, assigning the probabilities of "0" and "1" to these weights. Semicontinuous models were applied, establishing the presence or absence for each allele, calculating the probabilities for the "drop-in"/"dropout" phenomena. This approach is easier to understand than its counterparts. The main limits of this approach are the absence of the evaluation of peak height and the operator decision about the stutters. The semicontinuous models have good discrimination power in order to distinguish the true/false donors; nevertheless, these methods are unable to separate a mixture profile into individual profiles: for this reason, they cannot be used in DNA mixture interpretation when no reference profile is available for a POI.

To mark the naissance parameters indicated with "M," a set of approaches can be applied, using the maximum likelihood estimation (MLE) that does not lead to Eq. (2). To obtain functions proportional to the integral in Eq. (2), several applications can be applied, such as Markov chain Monte Carlo (MCMC). This model represents a continuous model, assigning a value between 0 and 1 to show the likelihood of a proposed genotype given the observed DNA profile.

The very important characteristic of the continuous model is related to the possibility to assign a relative probability, whereas the previous analyses are based on the determination of a possible (1) or not possible (0) genotype. In this way, the MCMC models may deconvolute a mixed profile into separate profiles, generating different combinations as well as calculating match statistics for comparison to the POI.

MCMC is applied in different fully continuous models, such as the PG software STRmix™. This software, developed by scientists with Forensic Science South Australia (FSSA) and Environmental Science and Research (ESR) in New Zealand, is one of the most used software packages for the interpretation of forensic DNA profiles today.

The Application of Probabilistic Genotyping Methods During the Process

There is a growing interest in the identification of a new method to establish an LR. The degree of mistrust or belief of any particular PG method is strictly related to the individual profile and among all participants involved in the relative scenario. Another important question is related to how to present the statistical evaluation. At first, the presentation of the match probabilities, even in order of several billions or less, was considered insufficient by the judge. For this reason, it was necessary to translate the match probabilities into frequencies in order to clarify the correspondence between the profile obtained from the evidence and the suspect (Hoffrage et al. 2000). Similarly, insufficient knowledge of statistics in general and incorrect Bayesian reasoning in particular can result in false convictions or acquittals made by juries in the court, for example, when they have to evaluate evidence based on a fragmentary DNA sample.

Another important question is related to the techniques applied in the profiling: current methods are able to obtain a complete profile starting from only few cells, referred to as LCN (Foreman and Evett 2001). This possibility should be evaluated in the two-way analysis. On the one hand, it is possible to obtain a profile in very difficult conditions; on the other hand, in other conditions, only a mixture profile or a partial profile may be obtained. This situation must be analyzed with the relative statistical tests in order to ascertain the weight of the evidence.

In the definition of a repeatable method, it is very important to ascertain if the theory or the technique used for the statistical analysis has been tested, has been subjected to peer review and publication, and has been generally accepted by the scientific community. Moreover, it is very important to define if the potential error rate has been defined and if there is a standard to check the effectiveness of the methods.

Obviously, one of the most important open questions concerns the standards that exist for PG and the applicability with the related software. Several different international organizations, such as the UK Forensic Science Regulator (UKFSR), the International Society for Forensic Genetics, and the US Scientific Working Group on DNA Analysis Methods, have published guidelines to validate PG software, focusing on the desired actions (SWGDAM 2015; Coble et al. 2016; Coble and Bright 2019; Leib 2019).

During the processes, there are two opposite statements: these are the prosecution and the defense theories. In this scenario, Bayes' theorem is very important in order to provide a logical evaluation of evidence "E" in the DNA analysis:

$$\frac{\Pr\left(H_p|E,I\right)}{\Pr\left(H_d|E,I\right)} = \frac{\Pr\left(E|H_p,I\right)\Pr\left(H_p|I\right)}{\Pr\left(E|H_d,I\right)\Pr(H_d|I)} \tag{3}$$

in this formula (3), "Hp" indicates the prosecutors' proposition, "Hd" the proposition of the defense, and "I" the circumstances (meaning the form of time, location, and witness statements describing the alleged actions). In this scenario, each value is strictly related to the reconstruction: if, during the process, the data changed, the new interpretation should be supplied.

Moreover, the evaluation of LR is very important:

$$LR = \frac{\Pr\left(E|H_p,I\right)}{\Pr\left(E|H_d,I\right)} \tag{4}$$

The evaluation of the LR represents a key point for evidence interpretation, impacting the content of the statements (Evett et al. 2000).

Particularly, its value is very important in order to define the weight of the evidence and is strictly related to the framework of circumstances. Moreover, the forensic scientist plays a pivotal role in order to define the values of each proposition (Balding and Nichols 1994; Evett 1995).

In light of this information, before any evaluation, it should be mandatory to elucidate the two alternative propositions for each criminal scenario. One of the most important concerns that several international courts have highlighted is relative to the closed source of software used in the processes, such as STRmix™. In this way, the rights of the defense could be prejudiced. The rights of the accused are exercised with the possibility to check the source code of the software. This opportunity is guaranteed with open-source software; on the other hand, closed source complicates this opportunity, and the cost, inconvenience, and the possibility to use Internet access should be evaluated. Moreover, the source should be consultable everywhere by everyone. In addition, PG software, in several cases, conceals the algorithm applied, becoming a "black box." This situation is overcome when the mathematical processes of the software are published in a peer-reviewed journal: this is the case of STRmix™. For example, in this case, numerous training workshops have been organized, giving practical training to the participants, thus allowing them to understand all of the possibilities of the software. In this way, forensic scientists can be trained, improve their knowledge, and avoid the use of unsophisticated yet obsolete methods. Moreover, in this regard, the International Society for Forensic Genetics (ISFG) has created a specific section on its website with the indication of forensic software resources (available on https://www.isfg.org/Software). Particularly, by supporting open-source software projects in forensic statistics, the ISFG has created a list of software applications as a service to the forensic genetics community, even though it is clearly written that the ISFG is not endorsing any specific software. Moreover, it is evident that the responsibility to validate and check if the selected program meets any applicable casework standards is of the end-forensic-user.

MPS Approaches to Forensic Genetics for DNA Mixture Deconvolution

As previously described, the statistical evaluation of DNA evidence weight is a very old question, particularly for mixture profiles. Indeed, for several years the procedures in the assignment of the number of contributors in a crime scene profile were made in subjective way (Cook et al. 1998a, b). Moreover, considering the sensitivity of modern forensic kits, the question about how DNA may have been transferred remains an important issue in the evaluation of evidence weight (Kokshoorn et al. 2017; Taylor et al. 2018b). Each piece of DNA evidence should be supported through the analysis of several concerns, such as transfer, persistence, and recovery of the detected profile. For these reasons, this research field should be supported by forensic scientists.

In the last decade, several different techniques have been combined with the standard STR methods to supply the PG evaluation, such as "Y" or "X" chromosome profiling (Asmundo et al. 2006; Presciuttini et al. 2011; Ventura Spagnolo et al. 2017). In this way, the weight of the evidence is improved (Taylor et al. 2018a, b; Andersen and Balding 2019). Moreover, the traditional methods based on STR typing (including Y-chromosome STRs and X-chromosome STRs) require about 500 picograms of DNA, and in several forensic cases, the collected evidence is enough for a single DNA analysis (Ludeman et al. 2018).

Another important question related to the mixture profile is the use of the new methods in forensic DNA profiling that have been developed in the past few years, using MPS, also referred to as NGS. These methods are able to detect STRs, SNPs, and in/del (insertions and deletions) simultaneously. MPS is related to a variety of high-speed sequencing platforms (sometimes called "second generation" or "next generation") that use a common technical approach to sequence a large number of fragments in parallel using spatially separated and clonally amplified DNA templates (Børsting and Morling 2015). Several phases are included in these techniques, such as the creation of a library, bridge amplification or emulsion PCR, and enrichment (Churchill et al. 2016). In particular, STRs may be amplified and subsequently sequenced: with this approach, it is possible to analyze the nucleotide differences observed in the selected STR repeat motifs and their flanking regions. Each fragment is uniformly and accurately sequenced in millions of parallel reactions. It is important to note that these new techniques generate data that are compatible with those previously archived. In particular, moving forward with MPS will give the forensic community an opportunity to expand its current use of forensic DNA typing with increased STR and SNP genetic diversity without the loss of the use of the millions of DNA profiles currently held in DNA databases (Churchill et al. 2016; Novroski et al. 2016).

To date, the commercial test used for forensic purposes can type hundreds of markers: in this way, adding the sequence information, the discriminatory power is improved (Fig. 6).

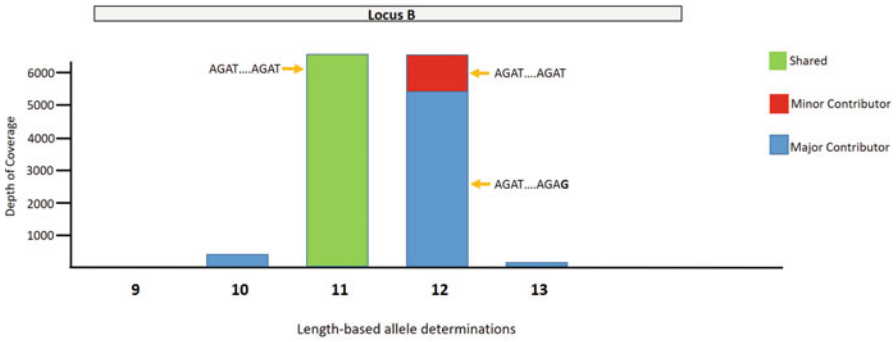


Fig. 6 Summary of the result of a single locus in a mixture profile obtained through NGS. In this case, the two people shared both alleles (11,12). The signal detected in allele 10 is a stutter from allele 11. Allele 11 is indistinguishable because no SNP difference is reported. Contrariwise, allele 12 distinguished the major and minor contributors. Using CE methods, it is impossible to obtain this result

These techniques are able to generate a large amount of DNA data using minimal starting material: NGS overcomes a major limitation of CE typing that often necessitates multiple rounds of testing, which can be impossible on limited or poor-quality material, or may not yield sufficient information for a conclusive result. Indeed, considering that in several forensic cases the initial biological sample recovered at the crime scene can be very poor, these methods minimize the overall number of samples used. For example, the ForenSeq™ DNA Signature Prep Kit (Verogen, San Diego, California, USA) contains 58 STR loci (27 autosomal, 7 X-chromosome, and 24 Y-chromosome STRs) whose amplicon sizes generally fall within the 350 base pair read length. In the same way, the Precision ID GlobalFiler™ NGS STR Panel v2 (Thermo Fisher Scientific, Waltham, MA, USA) analyzes more markers (35), including 20 CODIS STRs, 9 multiallelic STRs, 4 markers for the determination of sex, and 2 penta-STR markers with high informativity.

Moreover, MPS kits are able to supply further information such as physical characteristics, including eye color or ethnicity of the investigated subject (Wang et al. 2015).

In contrast, even though MPS may offer the greatest opportunity to resolve DNA mixtures, to date it is necessary to improve the statistical methods used to interpret these data (Coble and Bright 2019; Hwa et al. 2019). One of the most important problems is related to the use of a database to evaluate the weight of evidence statistically. Several papers worldwide have published datasets, contributing to the growth of STR sequence diversity, which is now compiled in the STRSeq database (Novroski et al. 2016; Wendt et al. 2016; Phillips 2017). Nevertheless, the difficulty of interpreting mixed DNA profiles using MPS is about the same as in traditional methods (Churchill et al. 2016). Although the quantification of each allele present in the mixture is more realistic, allele sharing is still a common phenomenon observed in a mixture. The determination of the amount of each allele allows a better

determination of allele ratios in the mixture and, in instances, of shared length-based alleles (i.e., homozygous by length) that differ by sequence (i.e., heterozygous by sequence), the contribution of each unique allele at the locus. Nevertheless, in cases where relatives may be involved, a lack of sequence variation in certain loci will limit the usefulness of MPS technologies (Novroski et al. 2016).

Conclusion

Notwithstanding the recent developments in DNA typing methods, the interpretation of a mixture remains one of the most absorbing research areas in the forensic field. Mixed DNA samples arise from the combination of two or more individual bodily fluids or secretions: this kind of evidence is common, and even expected, in many forensic investigations such as sexual crimes, large disasters, as well as in products of conception and fingernail cuttings taken by police or at autopsy. The difficulties related to DNA mixture interpretation are closely related to the different starting conditions: different types of materials, number of donors, and different proportions of each component in the mixture, as well as artifacts such as allelic dropout and allelic drop-in.

As far as the interpretation of mixed DNA samples is concerned, the objective of this analysis is the identification of major and minor contributors in the mixture profiles, following the national team and international guidelines (Gill et al. 2018).

The use of MPS methods has improved the ability to identify the contributors of a DNA mixture by identifying sequence variation through both STR typing and flanking region sequencing. Moreover, modern STR-based kits allow the sequencing of several markers with a high degree of diversity; in the same way, the analysis of STR sequence variation in sample population groups highlighted that several current STR loci lack the necessary sequence diversity.

In this game of equilibrium, the new methods offer the opportunity to enhance human identification for forensic purposes in DNA mixtures: by using new multiplex kits with highly heterozygous STRs, it is possible to have powerful conditions to solve the interpretation of many complex mixture samples more effectively. In the same way, it is possible to improve the DNA database by inserting more useful information that increases the number of forensic samples uploaded. The current human STR kits use markers with different heterozygosity, showing some highly heterozygous loci, while others are only marginally heterozygous. To improve the discrimination power in mixtures, it is important to search for new candidate loci. There are three criteria routinely used by the forensic science community to evaluate a new candidate STR: the locus should conform to the standard conditions (with particular regard to the amplicon size and composition); the locus must show great resolution power; and it should be evaluated and validated with robust studies. Recent studies have highlighted two important parameters for new forensic STR markers. First of all, they should exhibit a high degree of heterozygosity; moreover, they should show limited allele spread.

In this context, it is important to note that several highly informative STR markers are not included in the current kits because of technology constraints related to several conditions such as the chemicals used. Moreover, to date, mixtures generated by more than two contributors (three or more persons) remain frequently unsolved because of the difficulties in the evaluation of stochastic effects and/or other technical problems, such as the differences between true alleles and stutter products. For these reasons, future studies are needed to resolve both of these technical and interpretation difficulties.

To give an instance, a very recent study compared NGS and traditional methods to evaluate the ability in the identification of major and minor contributors (Ragazzo et al. 2020). Based on the results of this experimental model, NGS represents a reliable and robust method for human identification in standard conditions (starting from optimal/suboptimal amounts of DNA). It is important to note that the discrimination power (meaning the possibility to identify the major and minor contributors) in this system is obtained by setting the default analytical threshold (5% of the total coverage of alleles in each locus). As reported in this study, removing filters, the analysis fell into a range defined "inconclusive," confirming the importance of the pre-established analytical threshold for these methods.

Finally, the importance of statistical methods to establish the "weight of evidence" should be remembered, particularly for the interpretation of MPS multiplex kits for forensic DNA analysis. The interpretation of MPS results remains another interesting research field for the forensic community, particularly in the application of statistical methodologies.

References

Andersen MM, Balding DJ (2019) Y-profile evidence: close paternal relatives and mixtures. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2018.10.004

Aronson JD (2007) Genetic witness: science, law, and controversy in the making of DNA profiling. Rutgers Univ Pr. ISBN-10: 0813541883

Asmundo A, Perri F, Sapienza D (2006) Allele distribution of two X-chromosomal STR loci in a population from Sicily (Southern Italy). Int Congr Ser. https://doi.org/10.1016/j.ics.2005.11.133

Balding DJ (1999) When can a DNA profile be regarded as unique? Sci Justice. https://doi.org/10.1016/S1355-0306(99)72057-5

Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int. https://doi.org/10.1016/0379-0738(94)90222-4

Berry DA, Evett IW, Pinchin R (1992) Statistical inference in crime investigations using deoxyribonucleic acid profiling. Appl Stat. https://doi.org/10.2307/2348086

Bille T, Bright JA, Buckleton J (2013) Application of random match probability calculations to mixed str profiles. J Forensic Sci. https://doi.org/10.1111/1556-4029.12067

Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2015.02.002

Buckleton JS, Walsh S, Harbison SA (2001) The fallacy of independence testing and the use of the product rule. Sci Justice. https://doi.org/10.1016/S1355-0306(01)71858-8

Budowle B, Giusti AM, Waye JS et al (1991) Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. Am J Hum Genet 48(5):841–855

Budowle B, Eisenberg AJ, van Daal A (2009) Validity of low copy number typing and applications to forensic science. Croat Med J. https://doi.org/10.3325/cmj.2009.50.207

Butler JM, Kline MC, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2018.07.024

Churchill JD, Schmedes SE, King JL, Budowle B (2016) Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2015.09.009

Clayton TM, Whitaker JP, Sparkes R, Gill P (1998) Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Sci Int. https://doi.org/10.1016/S0379-0738(97)00175-8

Coble MD, Bright JA (2019) Probabilistic genotyping software: an overview. Forensic Sci Int Genet 38:219–224

Coble MD, Buckleton J, Butler JM et al (2016) DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2016.09.002

Cook R, Evett IW, Jackson G et al (1998a) A hierarchy of propositions: deciding which level to address in casework. Sci Justice. https://doi.org/10.1016/S1355-0306(98)72117-3

Cook R, Evett IW, Jackson G et al (1998b) A model for case assessment and interpretation. Sci Justice. https://doi.org/10.1016/S1355-0306(98)72099-4

Curran JM, Triggs CM, Buckleton J, Weir BS (1999) Interpreting DNA mixtures in structured populations. J Forensic Sci. https://doi.org/10.1520/jfs12028j

Devlin B (1993) Forensic inference from genetic markers. Stat Methods Med Res 2:241–262. https://doi.org/10.1177/096228029300200304

Devlin B, Risch N, Roeder K (1990) No excess of homozygosity at loci used for DNA fingerprinting. Science (80-). https://doi.org/10.1126/science.2205919

Dror IE, Hampikian G (2011) Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice. https://doi.org/10.1016/j.scijus.2011.08.004

Evett IW (1995) Avoiding the transposed conditional. Sci Justice. https://doi.org/10.1016/S1355-0306(95)72645-4

Evett IW, Gill P (1991) A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. Electrophoresis. https://doi.org/10.1002/elps.1150120220

Evett IW, Pinchin R (1991) DNA single locus profiles: tests for the robustness of statistical procedures within the context of forensic science. Int J Legal Med. https://doi.org/10.1007/BF01369583

Evett IW, Scranage J, Pinchin R (1993) An illustration of the advantages of efficient statistical methods for RFLP analysis in forensic science. Am J Hum Genet 52(3):498–505

Evett IW, Gill PD, Lambert JA (1998) Taking account of peak areas when interpreting mixed DNA profiles. J Forensic Sci. https://doi.org/10.1520/jfs16091j

Evett IW, Jackson G, Lambert JA, McCrossan S (2000) The impact of the principles of evidence interpretation on the structure and content of statements. Sci Justice. https://doi.org/10.1016/S1355-0306(00)71993-9

Foreman LA, Evett IW (2001) Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system. Int J Legal Med. https://doi.org/10.1007/s004140000138

Foreman LA, Smith AFM, Evett IW (1997) Bayesian analysis of DNA profiling data in forensic identification applications. J R Stat Soc Ser A. https://doi.org/10.1111/1467-985x.00074

Fukshansky N, Bär W (2000) Biostatistics for mixed stains: the case of tested relatives of a non-tested suspect. Int J Legal Med. https://doi.org/10.1007/s004140000155

Fung WK, Hu YQ (2002) The statistical evaluation of DNA mixtures with contributors from different ethnic groups. Int J Legal Med. https://doi.org/10.1007/s004140100256
Geddes L (2010) Fallible DNA evidence can mean prison or freedom. New Scientist 206 (2773):18–19
Gill P (2002) Role of short tandem repeat DNA in forensic casework in the UK – past, present, and future perspectives. BioTechniques 32:366
Gill P, Evett I (1995) Population genetics of short tandem repeat (STR) loci. Genetica 96:69–87. https://doi.org/10.1007/BF01441153
Gill P, Werrett DJ (1990) Interpretation of DNA profiles using a computerised database. Electrophoresis. https://doi.org/10.1002/elps.1150110516
Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA "fingerprints". Nature. https://doi.org/10.1038/318577a0
Gill P, Sparkes B, Buckleton JS (1998) Interpretation of simple mixtures of when artefacts such as stutters are present – with special reference to multiplex STRs used by the Forensic Science Service. Forensic Sci Int. https://doi.org/10.1016/S0379-0738(98)00094-2
Gill P, Haned H, Bleka O et al (2015) Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches – Twenty years of research and development. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2015.03.014
Gill P, Hicks T, Butler JM et al (2018) Forensic Science International : Genetics DNA commission of the International society for forensic genetics : assessing the value of forensic biological evidence – guidelines highlighting the importance of propositions Part I : evaluation of DNA pro fi li. Forensic Sci Int Genet 36:189–202. https://doi.org/10.1016/j.fsigen.2018.07.003
osch A, Euteneuer J, Preuß-Wössner J, Courts C (2020) DNA transfer to firearms in alternative realistic handling scenarios. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2020.102355
Hecht F (1989) The blooding. Am J Hum Genet 45(3):481–482
Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G (2000) Communicating statistical information. Science (80-.)
Hwa HL, Wu MY, Chung WC et al (2019) Massively parallel sequencing analysis of nondegraded and degraded DNA mixtures using the ForenSeq™ system in combination with EuroForMix software. Int J Legal Med. https://doi.org/10.1007/s00414-018-1961-y
Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific "fingerprints" of human DNA. Nature. https://doi.org/10.1038/316076a0
Kokshoorn B, Blankers BJ, de Zoete J, Berger CEH (2017) Activity level DNA evidence evaluation: on propositions addressing the actor or the activity. Forensic Sci Int. https://doi.org/10.1016/j.forsciint.2017.06.029
Lambert JA, Evett IW, Scranage JK (1995) Large scale database experiments to assess the significance of matching DNA profiles. Int J Legal Med. https://doi.org/10.1007/BF01845609
Leib H (2019) Annual report 2018. AIMS Electron Electr Eng 3:98–101. https://doi.org/10.3934/ElectrEng.2019.1.98
Lohmueller KE, Rudin N, Inman K (2014) Analysis of allelic drop-out using the Identifiler® and PowerPlex® 16 forensic STR typing systems. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2014.04.003
Ludeman MJ, Zhong C, Mulero JJ et al (2018) Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. Int J Legal Med. https://doi.org/10.1007/s00414-018-1817-5
Moretti TR, Moreno LI, Smerick JB et al (2016) Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2016.07.022
Moretti TR, Just RS, Kehl SC et al (2017) Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2017.04.004
Morton NE (1997) The forensic DNA endgame. Jurimetrics 37:477–494
Neckovic A, van Oorschot RAH, Szkuta B, Durdle A (2020) Investigation of direct and indirect transfer of microbiomes between individuals. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2019.102212

Novroski NMM, King JL, Churchill JD et al (2016) Characterization of genetic sequence variation of 58 STR loci in four major population groups. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2016.09.007

Oldoni F, Podini D (2019) Forensic molecular biomarkers for mixture analysis. Forensic Sci Int Genet 41:107–119

Perlin MW, Szabady B (2001) Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. J Forensic Sci. https://doi.org/10.1520/jfs15158j

Phillips C (2017) A genomic audit of newly-adopted autosomal STRs for forensic identification. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2017.04.011

Phipps M, Petricevic S (2007) The tendency of individuals to transfer DNA to handled items. Forensic Sci Int 168:162–168. https://doi.org/10.1016/j.forsciint.2006.07.010

Presciuttini S, Toni C, Alù M et al (2011) X-chromosome in Italy: a database of 29 STR markers. Forensic Sci Int Genet Suppl Ser. https://doi.org/10.1016/j.fsigss.2011.08.018

Ragazzo M, Carboni S, Caputo V et al (2020) Interpreting mixture profiles: comparison between precision ID globalfiler™ NGS STR panel v2 and traditional methods. Genes (Basel). https://doi.org/10.3390/genes11060591

Risch NJ, Devlin B (1992) On the probability of matching DNA fingerprints. Science (80-). https://doi.org/10.1126/science.1738844

Sakurada K, Watanabe K, Akutsu T (2020) Current methods for body fluid identification related to sexual crime: focusing on saliva, semen, and vaginal fluid. 10(9):693. https://doi.org/10.3390/diagnostics10090693

Sessa F, Salerno M, Bertozzi G et al (2019) Touch DNA: impact of handling time on touch deposit and evaluation of different recovery techniques: an experimental study. Sci Rep. https://doi.org/10.1038/s41598-019-46051-9

SWGDAM (2010) SWGDAM interpretation guidelines for autosomal STR typing SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. Website

SWGDAM (2015) Scientific working group on DNA analysis methods validation guidelines for the validation of probabilistic genotyping systems, pp 1–13. Availale at https://www.swgdam.org/publications

Taylor D, Curran J, Buckleton J (2018a) Likelihood ratio development for mixed Y-STR profiles. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2018.03.006

Taylor D, Kokshoorn B, Biedermann A (2018b) Evaluation of forensic genetics findings given activity level propositions: a review. Forensic Sci Int Genet 36:34–49. https://doi.org/10.1016/j.fsigen.2018.06.001. Epub 2018 Jun 7. PMID: 29929059

Thompson WC (1997) Accepting lower standards: the National Research Council's Second report on forensic DNA evidence. Jurimetrics 37:405–424

van Oorschot RA, Jones MK (1997) DNA fingerprints from fingerprints. Nature 387:767. https://doi.org/10.1038/42838

Ventura Spagnolo E, Mondello C, Procaccianti S et al (2017) Allele frequencies distribution of 16 forensic STR loci in a Western Sicilian population sample. Egypt J Forensic Sci. https://doi.org/10.1186/s41935-017-0013-y

Vraneš M, König M, Cornelius S et al (2019) Development of the Investigator® 26plex QS Kit: a new multiplex PCR Kit for Global STR analysis. Forensic Sci Int Genet Suppl Ser. https://doi.org/10.1016/j.fsigss.2019.10.174

Wang H, Zhu J, Zhou N et al (2015) NGS technology makes microhaplotype a potential forensic marker. Forensic Sci Int Genet Suppl Ser. https://doi.org/10.1016/j.fsigss.2015.09.093

Weir BS, Triggs CM, Starling L et al (1997) Interpreting DNA mixtures. J Forensic Sci. https://doi.org/10.1520/jfs14100j

Wendt FR, Churchill JD, Novroski NMM et al (2016) Genetic analysis of the Yavapai Native Americans from West-Central Arizona using the Illumina MiSeq FGx™ forensic genomics system. Forensic Sci Int Genet. https://doi.org/10.1016/j.fsigen.2016.05.008

Whitaker JP, Cotton EA, Gill P (2001) A comparison of the characteristics of profiles produced with the AMPFlSTR® SGM Plus™ multiplex system for both standard and low copy number (LCN) STR DNA analysis. Forensic Sci Int. https://doi.org/10.1016/S0379-0738(01)00557-6

Current Status of DNA Databases in the Forensic Field

46

Ethical and Legal Standards

Sachil Kumar, Saranya Ramesh Babu, and Shipra Rohatgi

Contents

Abstract

Recent advances in DNA technologies have made it possible to establish DNA databases for forensic investigations. Forensic DNA databases are currently in operation in about 70 countries. As expected with the tremendous success of the use of forensic DNA databases, many ethical and legal problems arise in the preparation of a DNA database, and these problems are particularly significant when one analyses the legal regulations on the subject. The continuous rise in the

S. Kumar (✉) · S. R. Babu
Department of Forensic Sciences, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia
e-mail: SKumar@nauss.edu.sa; SRamesh@nauss.edu.sa

S. Rohatgi
Former Faculty/ Research Scholar, Amity Institute of Forensic Science, Amity University Gurugram/Noida, Noida, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_48

size of forensic DNA data sets poses concerns about the requirements for inclusion and preservation and suspicions about the efficacy, affordability, and privacy violation of such vast repositories of personal data. In light of its broad reach, the database raised concerns about anonymity, government surveillance, and human rights. The aim of this chapter is to shed light on the current status of forensic DNA databases and their ethical and legal issues.

Keywords

Humans · Databases · Nucleic acid · Police · Privacy · Law enforcement · Crime · Data collection · DNA

Introduction

The establishment of DNA database began when the standard practice of forensic DNA typing of STR markers was accepted globally. The national DNA database of England and Wales (NDNAD) was the first criminal DNA database to be established in 1995 followed by New Zealand's database (Williams and Johnson 2008). In the USA, the Federal Bureau of Investigation (FBI) organized the Combined DNA Index System (CODIS) database, and it was from then, most welcomed by the criminal justice system including Parliamentarian, policy makers, police officers, and criminal detectives to solve criminal cases (Amankwaa 2018).

DNA databases were started by only collecting the profile of the individuals convicted/offended or suspected in previous crimes and in addition to the DNA profiles gathered from crime scene samples so as to find any hit or lead in other cases (Santos et al. 2013). But later, it began to acquire data from the public to maintain the references sample of the specific population be used in police investigation, and some countries continue to keep up record for specialized missing person's DNA database including the Missing Persons Relatives, Unknown Humanoid Remains, Y-STR, and mitochondrial DNA analysis (DNA 2020).

Forensic DNA databases from different countries vary in their data inclusion/retention and exclusion criteria based on laws enacted by their government policies (Montague 2011). According to Interpol global DNA profiling survey of 2016 and 2019, out of 194 member countries, only 89 countries reported acquiring DNA profile for investigative purposes of which are 11 African, 13 American, 19 Asian, and 46 European countries, while 70 countries has DNA database of which are 7 African, 10 American, 13 Asian, and 40 European (DNA 2020). The biggest DNA databases are in China with 80 million profile as of 2020 (Wee 2020), followed by the USA which is projected with 19.5 million in 2020 (CODIS – NDIS Statistics 2016) and the UK with 6.2 million profiles in 2018 (NPCC 2017–2018).

National DNA Databases

African Region

South Africa's DNA Criminal Intelligence Database (DCID) was established in 1998, and new legislation for expanding the National Forensic DNA Database of South Africa (NFDD) was passed in 2015. The profiles in NFDD are arranged by the following indexes, (a) Crime Scene, (b) Arrestee Index, (c) Convicted Offender Index, (d) Investigative Index, (f) Elimination Index, (g) Missing Persons, and (h) Unidentified Human Remains Index, and 91,240,168 DNA profiles were available by 2016 (Heathfield 2014; de Wet et al. 2011). Botswana's and Namibia DNA database contains 3300 and 1338 DNA profiles (Interpol 2011; Global Summary 2020).

Middle East

Egypt, Tunisia, Lebanon, Saudi Arabia, and Bahrain have 4162, 17,070, 23,000, 909,745, and 69,609 profiles (Interpol 2019). Israel Police DNA Index System (IPDIS) is the Israeli national DNA database established in 2007. The IPDIS includes elimination bank holding the profiles of the personnel handling the forensic sample and contains 491,380 profiles (Interpol 2019).

Iran, Kuwait, the UAE, and Qatar are reported with 10,000, 14,591, 24,370, and 2500 profiles (Interpol 2008). Jordan Database was established in 2000 and contains 14,104 profiles according to Interpol 2008. However, Algeria, Kenya, Rwanda, Uganda, Somalia, and Seychelles are planning to open up international DNA database (Global Summary 2020).

European Region

The council of European Union (EU) established DNA database in 1997. The DNA working group of the European Network of Forensic Science Institutes (ENFSI) decided that all EU member state should begin the establishment of forensic DNA database. The ENFSI has also set standards to generate DNA profiles for the DNA database, and Prüm treaty was signed in 2008 to make the database available for the exchange of forensic DNA data among the EU member states (Toom 2018). However, Prüm treaty was not operational in Greece, Ireland, and Italy (Europol 2019).

The Netherland has DNA database for criminal cases and missing persons which was by established National Forensic Institute (NFI) in 1997 and contains 351,912 Dutch people profiles as per the 2017 annual report of NFI (Ministerie van Justitie en Veiligheid 2018). The former database has index for traces, suspects, convicts, and

deceased victims and later has index for unidentified, missing persons and family members of missing persons (Ministerie van Justitie en Veiligheid 2019). MixCal6 a statistical method for assessing the evidentiary proof of mixed DNA traces was developed by NFI and included in CODIS by FBI (Ministerie van Justitie en Veiligheid 2020).

Germany's DNA database *Bundeskriminalamt* (BKA), Wiesbaden was set up the German Federal Police in 1998 and holds up to 1,213,331 profiles (Interpol 2011). The campaign "Stop the DNA Collection Frenzy!" initiated by various civil rights and data protection organizations to circumvent the increase in DNA data-collection leads to the cancellation of international exchange of DNA data in 2011. Austria's ministry of interior maintains their DNA database and is composed of 384,098 profiles (Interpol 2011).

Banca dati nazionale del DNA is the Italian national DNA database (ITNDNADB) which was formed in 2009 and uses CODIS software for comparison (Biondo and De Stefano 2011); however, concerns are raised by the member countries as the ITNDNADB are not consistent with security measures to protect the data and storage of DNA profiles and biological samples (Marchese et al. 2013).

Belgium has two national DNA database in 1999 by National Institute for Criminalistics and Criminology with 54,399 profiles as of 2019 data (Belgium 2020). The criminalistics database for the DNA profiles are derived from crime scene while convict's database for DNA profiles of persons convicted for offence crimes, and however, no database for suspects DNA profiles were created (De Moor 2018). Similarly, in Denmark's Central DNA Database was introduced in 2000 and divided into individual database where DNA profile of perpetrator charged with a crime and Trace Database for those obtained from unidentified crime scene samples and currently holds 184,913 profiles (Anker et al. 2017).

Estonia and Ireland established its DNA database in 2004 and 2013 currently holding 67,758 and 26,492 profiles, respectively (Interpol 2011). Finnish National DNA Database is formed in 1999 under the custody of National Bureau of Investigation (NBI) Crime Laboratory and contains 193,575 profiles. Both Estonian and Finnish police were authorized to forcibly collect DNA samples from adults and minors suspected or convicted for any offense (Global Summary 2020).

The French national DNA database called as the *Fichier National Automatisé des Empreintes Génétiques* (Automated National File of Genetic Prints) was formed in 1996 and holds 4,247,382 profiles (Interpol 2011) which is 5% of its population; initially profiles were reported only sexually related crimes but expanded to cover profiles from almost all violent crimes in 2003 (Global Summary 2020). Greece began its DNA database in 2008 for convicted individuals, but by 2009 it started to include suspects of adults and children and presently holds 34,647 profiles (Voultsos et al. 2011), while national database were established for Cyprus, Croatia, and Czech Republic in 1998, 2000, and 2001, and each contains 51,093,16,706, and 253,085 profiles, respectively (Interpol 2011).

Hungary's DNA database was set up in 2004 and contains 212,196 profiles, while Bulgaria's Database was formed in 1999 and holds 17,055 profiles. Portuguese DNA

databank which was formed in 2005 known as Portuguese Ethics Council (CNECV – *Conselho Nacional de Éticapara as Ciências da Vida*) has index only for criminal's DNA profile and holds 5393 profiles as per 2013 Portuguese press report (Borja-Santos 2015).

Romania established National System of Judicial Genetic Data (NSDGJ) in 2008 and is subdivided into 1) personal database containing personal data of perpetrators and instigators, 2) investigation database for derived from crime scene stains, and 3) DNA profiles database for profiles of individuals and unidentified crime scene stains (Frank and Găină 2014). According to Interpol 2011 report, 8000 reference DNA profiles are stored in NSDGJ. Slovenia, Slovakia, and Luxembourg reported to have 37,531, 96,203, and 10,358 profiles (Interpol 2011).

Sweden and Spain established their DNA database in 1998 and 2007 and have 201,900 and 627,163 profiles, respectively, while North Macedonia holds 27,545 profiles according to 2019 report of Interpol (Interpol 2011).

The other non-EU member countries including Norway, Iceland, Switzerland, and Liechtenstein also signed Prüm treaty for the exchange of DNA data (Council of the European Union 2018), while Norway and Switzerland have reported to contain 109,180 and 268,417 profiles, respectively (Interpol 2011). The National Commissioner of Police and the Ministry of Justice in Iceland established two forensic DNA databases in 2008: one contains the profiles of individuals convicted for crimes for the Identification Database and the Trace Database for unidentified genetic profiles found at crime scenes (stjórnarfrumvarp: Erfðaefnisskrá lögreglu 2020). Belarus and FYR Macedonia are currently holding up to 336,408 and 27,454 profiles, respectively (Interpol 2011).

In Russia, Federal database of genome information (FDBGI) was formed in 2009 and has voluntary and mandatory genomic registration and contains 734,373 profiles (Interpol 2011). The mandatory database includes profiles of the convicted and imprisoned persons involved in all categories of crimes and unidentified corpses/persons collected at the crime scene (Perepechina 2019).

The UK National Criminal Intelligence DNA Database (NDNAD) was foremost forensic DNA database formed in 1995. Later in 2007, an NDNAD service was reassigned to the National Policing Improvement Agency to maintain the database operations and ensure the integrity of the data. According to the Criminal Justice and Police Act 2001, DNA sampling was done only from individuals charged with any recordable offence. But later, due to the Criminal Justice Act of 2003, samples were collected only from individuals above 10 years of age on arrest for their involvement in any crimes in England and Wales. However, after the Protection of Freedoms Act in 2012, DNA data of more than 1.7 million innocent people and children who were not charged and/or not found guilty have been deleted, and more than 7.7 million DNA samples have also been destroyed.

The National DNA Database department in the Police Forensic Science Laboratory Dundee of Scotland disseminates the collected data to the NDNAD in England (Johnson and Williams 2004). Similarly, Isle of Man Constabulary's Scientific Support Department and the police of Jersey and Guernsey also store their data to NDNAD.

Australasia

New Zealand's Criminal Investigations (Blood Samples) Act of 1995 initiated the establishment of National DNA Profile Databank (NDPD) in 1995, and ESR (Environmental Science and Research) maintains the operations of the NDPD. Legislation was amended in 2004 to expand the Criminal Investigations (Bodily Samples) Act of 1995 in which samples were obtained from convicted current prison inmates prior to 1995 and mandatory sample collection as buccal swabs for the Databank (Criminal Investigations Bodily Samples Act 1995).

It has two databases, namely, DNA Profile Databank (DNAPD) and Crime Sample Database, one for the individual profiles of the registered criminals and volunteers and the other for profiles from unsolved crimes, respectively (Harbison et al. 2001). In 2010, Legislation was amended and expanded for the Enactment of the Criminal Investigations (Bodily Samples) Amendment Bill 2009, and thus Temporary Databank (TD) was created holding profiles of the person arrested or intend to charge, and if convicted, it is transferred to DNAPD (Flaus 2013).

In 2013, New Zealand signed a data sharing agreement with the USA, to allow each country legal access to the other's fingerprint database under specified conditions, for automated searching. If a match was found in either country, further information such as addresses, convictions, known associates, and aliases will be shared (Agreement between the United States of America and New Zealand 2013). From 2014, DNAPD utilizes Globalfiler® PCR amplification kit for the generation of profiles and plans to expand the database to massively paralleled sequencing technology for the analysis and integration of case samples with the DNAPD and currently maintains 237,269 DNA profiles.

In Australia, the Crimes Act 1914 enable offenders, suspects, and volunteers to be sampled for criminal investigations and also standardizes the use, storage, release, and removal criteria in the DNA database system. The Australian Criminal Intelligence Commission established three DNA databases, 1) DNA database for law enforcement purposes operated by the Australian Federal Police, 2) Disaster Victim Identification database, and 3) the National Criminal Investigation DNA database (NCIDD), in 2001. In 2015, the NCIDD started to include familial searching, kinship matching, mtDNA, and Y-STR profiling and reported to held 830,000 DNA profiles as of 2014 (The National DNA Database Is Watching You 2020). Australia has also signed agreement for supported the international exchange of DNA information with the UK, the USA, and Canada in 2014 (Smith and Mann 2015).

American Region

The Technical Working Group on DNA Analysis Methods (TWGDAM) in the USA established its first national DNA database by 1989. Later, by 1994 the Combined DNA Index System (CODIS) was created and maintained by the FBI. The CODIS is sub divided to: 1) National DNA Index system (NDIS) for the exchange and comparison of profiles among the national laboratories, 2) State DNA Index System

(SDIS) for the exchange and comparison among various states laboratories, and 3) Local DNA Index System (LDIS) collecting profile in the neighborhood and then uploaded to SDIS and NDIS. All laboratories are installed with Criminal Justice Information Systems Wide Area Network to integrate and connect all the three levels to CODIS.

For the criminal investigations purposes, three indexes are mainly used: 1) the offender index for storing convicted profiles, 2) the arrestee index for arrested criminals, and 3) the forensic index containing profiles collected from a crime scene. In addition, there are also separate indexes for staff index who are involved in handling of case samples, the multi-allelic offender index that have more than two alleles at two or more loci obtained from individual samples, and partial profiles index for degraded or mixtures of multiple individuals.

In 2000, the National Missing Person DNA Database, also known as CODIS (mp), was maintained by the FBI at the NDIS level and has unknown human remain index, missing person's index, and the biological relatives of missing person's index. The Unidentified Human remains were tested for nuclear, Y-STR (for males only), and mitochondrial analysis by the University of North Texas Center for Human Identification which is funded by the National Institute of Justice to exploit the possibilities of identifying remains (Florida International University – Digital Communications 2020).

The Department of Defense maintains a database called Department of Defense Serum Repository for US servicemen with their dependents profile to support in the identification of human remains and has more than 50 million profiles. From January 2017, CODIS adopted to include 20 STR markers in its database where previously it was only 13 markers with Amelogenin and currently contains over 14 million DNA profiles.

Since 2004, there is a DNA databank called "Caminho de Volta" for missing kids in Brazil (da Silva et al. 2009). The Brazilian DNA database the *Rede Integrada de Bancos de Perfis Genéticos* (*RIBPG*) was formed in 2013 by the Brazilian Federal Police and presently incorporates 20 laboratories from all over Brazil and further aims to connect the remaining state laboratories for utilizing CODIS to solve criminal cases, sexual assault, and missing people searches (Ferreira et al. 2013). According to the 2020 report of RIBPG, 82,000 profiles have been submitted by 18 states laboratories compared to 30,809 genetic profiles produced in 2019 report of RIBPG.

The Canadian National DNA Database (NDDB) managed by the Royal Canadian Mounted Police (RCMP) was operational from 2000 by the DNA Identification Act (Milot et al. 2013). The amendment in the Criminal Code of Canada made provision for the collection of blood, buccal, or hair samples from the convicted persons. The NDDB has four indexes: 1) the Convicted Offender Index (COI) containing DNA profiles of offenders in imprisonment and 2) the Crime Scene Index (CSI) containing DNA profiles obtained from crime scenes. The COI are operated at the national level and CSI managed by the local forensic labs: 3) the victims Index (VI) and 4) the voluntary Donors Index (VDI). DNA profiles are stored in the database based on 13 Core CODIS markers and records up to 535,236 profiles.

Additionally, National Missing Persons DNA Program (NMPDP) was established in 2018 by the RCMP and holds about 500,000 DNA profiles. NMPDP has three non-criminal/humanitarian DNA indexes: 1) Missing Persons Index (MPI) contains DNA profiles obtained from the personal objects of the missing persons, 2) Human Remains Index (HRI) contains DNA profiles obtained from the human remains, and 3) Relatives of Missing Persons Index (RMI) contains DNA profiles obtained from the close relatives of the missing person. The RMI is restricted to search only against MPI and HRI and not to any of the criminal indexes.

Chile National DNA Databases *Sistema Nacional de Registros de* (SNDD) was established in 2004. SNDD is divided into five different databases: 1) Offender Database for individuals convicted of certain serious crimes, 2) Defendant/Suspect Database for individuals charged for serious crimes and subject to the verdict either removed or moved to the Offender database, and 3) Evidence Database unidentified biological material collected at crime scenes. The profiles included in Offender, Defendant/Suspect, and Evidence databases are retained for 30 years, regardless of the offender's death, parole, or release 4) Victims Database for those confessed of serious crimes and are retained until the perpetrator(s) are identified 5) Database of the Disappeared or executed individuals and profiles are retained until they are identified. The SNDD operates using the CODIS software and has 78,733 profiles.

In 2011, new law was introduced in Argentina's Ministry of Justice to form two independent DNA databases: 1) for the profiles of the individual committed major crimes with confirmed convictions and 2) for the evidences collected from unresolved cases which also has database containing DNA donated by the family of missing children (Penacino 2008). However, in 2016, Mendoza province has created a new database the *Registro Provincial de Huellas Genéticas Digitalizadas* and contains 40,652 profiles of these 30,507 belongs to prisoners and the remaining are from the person involved in handling the case (Locarno et al. 2019).

Since 2006, the Committee of Relatives of Dead and Missing Migrants of El Salvador (COFAMIDE) with Argentine Forensic Anthropology Team (EAAF) and the Salvadoran government established archive for disappeared migrants in Mexico. Lately in 2019, Mexico planned to operate national DNA criminal database with the USA to combat the people against smuggling and human trafficking.

Bermuda established DNA database in 2005 and has stored 6620 profiles. It has criminal DNA database and general population database. The inclusion of arrested persons and retention of DNA profiles indefinitely from innocent people were implemented. Colombia's DNA database was set up in 2008, and their code of criminal procedure 2004 provides provision for DNA analysis and has 6833 profiles. DNA databases are also planned in Costa Rica, Cuba, and Ecuador.

Asia

In 2004, the Genetics Laboratory of the Institute of Forensic Sciences started DNA Database in China which is currently the largest known database with 80 million profiles as of 2020 and started 429 forensic DNA labs throughout China for the DNA

samples collection. This database also contains additional core STR loci for enhanced discriminatory capacity tailored to the ethnic make-up of China's population (Ge et al. 2013). The following are the indexes, Crime Scene Index, Convicted Offender Index, Suspect/Accused/Arrestee Index, and Unknown Deceased Index present in the database, and plans to include Missing Persons and Victim/Volunteer Index in addition to missing and migrant children database are established. China's Ministry of Public Security (MPS) established a national Y-STR database in 2017 (Ge et al. 2014). ASPI reports that the Chinese police claims that database would help to establish a link with the evidence and nothing to do with crime and estimated to contain around 35–70 million Y-STR profiles.

Hong Kong, a Special Administrative Region of China, started regional Forensic DNA Database using CODIS software in 2001. The Hong Kong Police Force Ordinance provides guidelines for collecting samples only for serious offence and the destruction of DNA samples as soon as the forensic analysis has been completed. According to 2015, there are about 49,466 profiles in the database.

Japan's National Police Agency (NPA) established DNA database in 2004 and proposed the following indexes: 1) crime scene DNA profiles, 2) unnatural death DNA profiles, and 3) suspect DNA profiles. The DNA crime database accounts to have about 1.3 million as of late 2019 which means 1 of every 100 citizen's DNA has been profiled (Cyranoski 2004). National Research Institute of Police Science plans to start up a first of its kind database to include suspect's ethnicity, blood type, metabolic enzymes, hair and skin pigment proteins, mitochondrial DNA, and signs of asymptomatic viral infections to distinguish ethnicities.

South Korea's DNA database of missing children were started since 2002; however by the Use and Protection of DNA Identification Information Act in 2010, Forensic DNA Division build up a national DNA identification information database and DNA quality control service. Samples are collected from those accused or convicted of 11 violent crimes, including robbery, arson, drug trafficking, rape, and sexual assault against minors. The DNA Identification Management System maintained by the National Forensic Service has three indexes (arrestee, crime scene, and elimination index), while Korean DNA Database was operated by the Supreme Prosecutors Office including the convicted offender and crime scene index, and both database contain 179,848 profiles.

The Forensic DNA Databank of Malaysia (FDDM) was set up lately in 2015 by the Royal Malaysia Police Forensic Laboratory (RMPFL) and police (RMP) under the enactment of the Malaysian DNA Identification and Regulation Acts in 2009 and 2012, respectively. Approximately, database contains nearly 75,000 DNA profiles acquired from suspects, convicted offenders, crime scenes, detainees, drug dependents, missing persons, and volunteers from the members of the missing person and/or FDDM staff (Kumar et al. 2016).

Indonesian National Police and the Eijkman Institute for Molecular biology has DNA forensic unit for parentage, disaster victim, and perpetrator identification database containing mtDNA and forensic STRs for many Indonesian ethnics and contains 17,830 profiles. Bangladesh, Sri Lanka, the Philippines, Vietnam, and Thailand plan to set DNA database.

As of now, India doesn't hold any repository or DNA database at national level (Kumar et al. 2016). However, a survey conducted by Interpol in 2008 states use of a national design software to operate state database called HID (human identification). In India, DNA Technology (Use and Application) Regulation Bill, 2019 states "that no legal proceedings can be initiated against the Union government or any member of the board of the DNA National Data Bank for any action which is done in good faith." However, what constitutes "good faith" is vague here that fails to ascertain accountability on the board or the government of any responsibility in cases of security breach.

DNA evidence is undoubtedly conclusive and hence Centre is trying for fast implementation of the DNA Technology (Use and Application) Regulation Bill. It is under consideration with the parliamentary standing committee on science and technology, and once it comes into picture, it will go long way in faster and fair trials, standardization of protocols, and also prevention of repeat offence with a help of a regulated DNA databank (Anand 2020).

Ethical and Legal Standards

DNA databases all over the world change comprehensively on issues related to access and consent to support of both DNA tests similarly as the modernized profiles produced using them. Nonetheless, none global principles as well as restricted shields are there for ensuring security along with basic liberties. The development of criminological DNA information bases overall is frequently described as improvement in policing on public request. Different nations need essential quality affirmation for research centers or a dependable framework to follow DNA proof from the wrongdoing scene to the court and forestall misunderstandings, tainting, and unnatural birth cycles of equity. US people are effectively advancing DNA information bases, regularly depicted as specialized answers for horror rates. FBI Laboratory has worked in accordance with more than 29 nations by using the CODIS software in order to advance peaceful accords and approving enactment (Williams and Johnson 2004; Issues 2020).

The advancement of legal DNA information bases is in no way, shape, or form restricted to governments. A private organization has been created to straightforwardly contract with unfamiliar governments for the construction and maintenance of DNA database, offering strategy suggestions generally displayed on US practice. In the recent time period of 10 years, Life Technologies has prompted more than 50 unfamiliar organizations related to legal DNA enactment, strategy, and special regulations providing a base to foreign countries. Japan Legislature in 2009 normalized DNA assortment and examination for nation's 47 prefecture research centers utilizing Life Technologies DNA testing frameworks. Life Technologies keeps on offering help to NPA (National Police Agency). In 2009 the Bermuda government marked a huge contract of about 1,000,000 dollar with Trinity DNA (Florida based firm) for setting up DNA database plan (Issues 2020; Williams and Johnson 2004).

Hereditary information bases for wrongdoing have become a public point after the capture of the Golden State killer, Joseph James De-Angelo. He was captured at the age of 72 years and committed in excess of 50 assaults and 12 homicides. While his capture was commended as a law requirement triumph, a large group of inquiries developed in view of the way law authorization authorities in the end discovered De-Angelo: through a mix of conventional criminologist work and use of information from a publicly supported hereditary information base. For this situation, police looked through GEDmatch. This case boomed various discussions on DNA-related issue and its utilization as a proof. The central issue is that specialists use hereditary information bases to recognize the groups of suspects. In spite of the fact that the open doors for wrongdoing settling by using DNA information base hunts might be huge, new advances and inventive employments of them don't happen in a vacuum. All things being equal, novel employments of innovation request thought of countless moral issues and order cautious cross-examination of the likely effect of DNA information bases on wrongdoing control (Carracedo 2008).

The maintenance of DNA profiles and tests taken from wrongdoing scenes can be promptly defended in light of the fact that they may be valuable if an examination should be re-opened later on (either to convict a culprit or to absolve a guiltless individual). The significant basic freedoms concerns identify with the extending people gathering for DNA collection and handling afterward. The DNA framework is on the grounds that:

- DNA can be utilized to follow people or their family members; thereby database of DNA can be abused by who can invade the framework, e.g., governments or any individual.
- It can prove to be fruitful in the case of finding a suspect. The stored DNA records are compared to the records stored in other PCs, for example, for the rejection of a visa the captured records can be used as a source.
- DNA tests and profiles contain private data about well-being and hereditary connections (counting paternity and non-paternity) (Issues 2020, Carracedo2008, Kobilinsky et al. 2007).

Major Ethical Concerns Include

Human Rights and Racial Contemplations

Basic liberties concern that includes DNA sample collection without full educated assent of subject's must be defended in restricted conditions. Director of a Human Rights Watch in China states "DNA assortment can have real policing utilizes in examining explicit criminal cases, yet just in a setting in which individuals have significant security insurances," as the system in which it functions there have been reported to infringe rights to privacy. Grand (ECHR) Chamber of the European Court of Human Rights in 2008 prohibited the assortment and inconclusive maintenance of fingerprints, cell tests, and DNA profiles. The European Court, in arriving at its decision, contemplated that broad, unpredictable DNA information bases

abused the privilege to individual protection. It added that DNA assortment might be suitable according to state security and crime deterrence, however just if the assortment framework is intensely managed by setup law and open to the cautious investigation of judiciary. The Supreme Court of the USA decided that the assortment and maintenance of DNA profiles on individuals indicted for rough violations were legitimate given the restricted kinds of assortment, investigation, and utilization of tests gave by rule, in Maryland v. Ruler. In China, the police collection of DNA samples seems to be limiting in nature, Article 130 of CrPC states that over the span of criminal examinations, to "learn certain highlights, states of wounds, or states of being of a casualty or a criminal suspect, an actual assessment might be directed, and fingerprints, blood, pee and other natural examples might be gathered. On the off chance that a criminal presume will not be inspected, the specialists, when they consider it fundamental, may lead a mandatory assessment." It clearly lacks a legal guidelines and principles on the time limit for the preservation of DNA, how the data can be used and shared by any organization. It is a clear lack of privacy protection measures and gives way to abuses (Carracedo 2008; Kobilinsky et al. 2007).

For instance, New York State division in 2007 regarding equity in criminal administrations relaxed family looking strategies. Currently the major focus is on the look, yet just in specific situations, and just under survey. The change had the help of lead prosecutors, in a similar manner of homicide victim dad who was public in his assistance of the augmentation of DNA framework. The ill effect of security issues can be seen on the people who are generally involved in criminal case. Because of the utilization of online media, wrongly charged people may wind up with profound reputational harm or notwithstanding expanded pressure from accepting they might be erroneously blamed for a wrongdoing.

Lastly, the most amazing investigations of the usage of data bases for familial DNA looking is the disproportionate impact it would have on dull and Latino individuals who are starting at now caught by the criminal value structure. Greely, Roberts, and Erin Murphy in their study found that racial incoherencies in imprisonment transform into an uneven grouping of African American men from whom DNA was obtained and kept in databases all across the world (Carracedo 2008).

In the year 2010, individuals of color represented roughly 27 percent of grown-up captures when the grown-up dark populace was 12 percent. In certain areas, where individuals of color comprises the populace more than 50%, for example, Shelby County, individuals of color establish 85 percent of lawful offense litigants. In the Maricopa County, 30% were ariz and 39 percent of the lawful offense litigant populace was Hispanics. Furthermore, in that very year, the Hispanic populace of Maricopa County, Ariz., was 30%, while Hispanics made up 39 percent of the lawful offense respondent populace. Significantly, while a considerable lot of those captured are dark and Latino, accordingly populating DNA information bases, many are frequently not sentenced (Kobilinsky et al. 2007).

Privacy (Lack of Safeguard)

Privacy is one of the significant setbacks for keeping up a DNA database as the data set comprising a large number of profiles of DNA which relate to biological samples databank. DNA Fingerprinting and Civil Liberties procedures were described by Journal of Law, Medicine, and Ethics in its issue of mid-2006. The issue of protection can be drawn nearer in various manners. For example, to begin with 13 STR loci, CODIS center which lies in non-coding genes of the human genome which has no relationship with any kind of hereditary sickness and henceforth the data in the information base is just valuable for human character testing. Secondly individuals or other portraying data are taken care of at public level with the DNA profiles. Explicit case information is made sure about and constrained by the law authorization organizations that present the information. In this way, just the crime lab that presented the DNA is such a kind of foolproof evidence that can give a positive result with a known individual. Thirdly the information is scrambled and is just open to some limited authorized organizations and CODIS directors. Lastly, government and state punishments regarding ill-advised utilization of DNA tests incorporate heavy fine along with conceivable detainment. If a person utilizes the DNA data for any for say reason except the authorized persons, then he or she has to face severe punishments. The punishments incorporate a fine of 250,000 dollars for unapproved divulgence of data (Herkenham 2006).

Familial Looking, Research Use, and Counter-Psychological Oppression

Familial looking, a cycle through which specialists search halfway coordinates between DNA profiles obtained from the crime scene and the DNA profiles belonging to people put away on the basis of their DNA information. It is used for recognizing suspected relative with a probability of whom he or she might have met, possibly prompting the speculator's ID and maybe a fruitful arraignment. Familial looking through prompts a not insignificant rundown of incomplete comparison analysis should be abbreviated by the help of more extensive DNA testing and the police official's investigations. In UK, the technique is spearheaded as it is assisted with illuminating various genuine wrongdoings. Nonetheless, it raises extra worries for the genuine people security who are not the culprit or suspect, but still their profile matches with the suspected one when specifically talking about non-paternity cases which may coincidentally be uncovered by familial looking cycle. Whenever utilized regularly, familial looking could prompt critical maltreatments by permitting examiners or any individual who invades the information base to find the family members of political protesters or to seek after foes (Maguire et al. 2014).

DNA data sets comprise assortments of organic examples (whenever put away), mechanized DNA profiles, and other data (e.g., criminal history and nationality) that might be significant to hereditary analysts. Nonetheless, much examination around there is disagreeable because of the historical backdrop of selective breeding. Specifically, endeavors to interface hereditary attributes to undermined ideas of

race or to distinguish "qualities for guiltiness" are disputable. Dissimilar to setup of information bases for research purposes, measurable DNA information bases contain information gathered without assent and additionally once in a while with assent for policing purposes as it were. Any attempt to use such databases to make inferences regarding inherited characteristics is thus in violation of established moral principles. Such breaks have just happened with some current information bases (Kobilinsky et al. 2007).

The utilization of DNA information bases in criminal examinations requires a person's personality to be uncovered in particular case where the DNA profiles stored in the databank match with the profile found at the crime scene. Up to this point, employments of DNA information bases were limited to a great extent to searching for matches with wrongdoing scene DNA profiles. Currently it is evolving. In UK, the DNA gathered and held under the act of Counter-Terrorism 2008 would be able to be utilized for "recognizable proof" of the individual from whom the material came. This is an ongoing difference in use which permits organic observation of specific people (e.g., the capacity to utilize a person's DNA to follow and distinguish them, regardless of whether they are associated with perpetrating a wrongdoing). Obviously, this might be valuable to security benefits; however it is additionally conceivably prone to mishandle. Government of proposition to gather fingerprints and DNA information regularly on capture of an offense and using it regularly for the purpose of recognizable proof (e.g., coordinating the person to their subtleties on unique mark information bases and DNA, utilizing offices set up in malls for such intentions) were dropped in 2008 after open objection. Notwithstanding, this remaining parts a likely use for DNA information bases later on, especially as new innovation creates that can permit DNA constant testing while on the spot of crime scene and coordinating with records of information base (Maguire et al. 2014; Ahmed 2020).

DNA Is Not Secure and Foolproof

DNA isn't foolproof and secure. By some coincidence a false match can occur among the DNA profile of a person with the crime scene DNA profile, or because of helpless lab strategies, and the ramifications of somebody's profile of DNA at the scene of crime likewise be misconstrued. The chance of a counterfeit match between an individual's DNA profile and a bad behavior scene DNA profile depends upon the course of action of DNA profiling that is used. With the passage of time, the rules used to make a DNA has changed and vary from country to country: the USA uses 13 STRs at better places in the genetic gathering; anyway most various countries use less STRs. The UK structure (10 STRs) is surveyed to have around a 1 out of a billion "facilitate probability": this is the likelihood that an individual's DNA profile arranges a bad behavior scene DNA profile by chance whether or not the DNA at the bad behavior scene didn't begin from. In spite of the fact that this probability is extremely low, the quantity of bogus matches that happen relies upon the quantity of examinations that are made between various

DNA profiles. On the off chance that each wrongdoing scene DNA profile is thought about against each put away DNA profile on an enormous information base by theoretical looking, few bogus matches are required to happen essentially by some coincidence. Bogus matches are bound to happen with family members; sibling of any individual who has committed some sort of crime will have in common a portion of the family general's DNA succession. The issue is exacerbated when only a few crime scene DNA profiles are completed, the risk of a false match increases, which could lead to some wrongdoing at the site. The risk of spurious similarity may increase dramatically if DNA profiles were completed (Maguire et al. 2014; Machado and Silva 2014).

The samples of DNA and their nature might change as indicated by the wellspring of the DNA, regardless of whether it has gotten debased after some time, whether the DNA belongs to a single person or more than one. Very small size sample for DNA analysis from the crime scene can make the case investigation move in a wrong direction regardless of the fact whether the person was actually present at the crime scene or not. Conversely, a huge amount of blood found at the location of a homicide or murder will offer dependable outcomes. A combination can be deciphered from numerous points of view, but there exists no such reliable method to tell from which piece of DNA profile originates from which implies that blended DNA database profiles are not entirely clear, especially if a criminological lab is one-sided by attempting to catch hold of the real culprit by having an exact match. This improves the probability of a bogus match with some unacceptable individual. DNA tests can likewise be wrongly examined or stirred up during lab techniques, bringing about a match with some unacceptable individual if there is no proper adaptation of quality affirmation methodology. Routine cross-outskirt theoreticallooking of DNA profiles at crime scenes with the DNA samples captured from the people in different nations is a way prone to hurl a lot more bogus matches than if such ventures are confined toone nation or restricted to justfew suspected people for DNA matching (Machado and Silva 2014).

Regardless of whether a DNA coordinate is real, an individual's essence at any scene of crime does not prove that the person is involved in any kind of wrongdoing. The accurateness of the profile matching ought to rely upon whether there is extra supporting proof. Any person whose DNA data information is already existing may be defenseless in the case proofs are being planted against him or her by the degenerate cops, ground-breaking government offices, or hoodlums. Regardless of whether an unsuccessful labor of equity doesn't happen, a person who is dishonestly blamed for a wrongdoing because of a DNA match might be exposed to a distressing police request, pre-preliminary confinement, or removal to an unfamiliar nation (Anderson et al. 2011; Maguire et al. 2014).

Errors and Improper Utilization of Police Resource

DNA is without a doubt a significant instrument in criminal examinations and has assisted with getting the culprits of some intense violations, including assaults and

murders. In any case, the possibility of having no more sexual assaults or murders is only when every single individual gets his/her DNA information recorded on an information base. Notwithstanding worries about protection and rights, the primary constraints to this thought are:

(i) The troubles in gathering applicable and valuable wrongdoing scene DNA proof
(ii) The extremely low probability of the vast majority perpetrating genuine violations for which DNA proof may be significant
(iii) The expenses and handy challenges related with gathering and keeping records of DNA profiles and related information from gigantic amounts of people
(iv) Open trust consequences in policing

England and Wales has a significant venture started by their government in 2000 for extending DNA utilization. The positive outcomes can be achieved by improvising the DNA assortment from scene of crime and accelerating the examination. Be that as it may, there were still genuine down as far as possible in how much useable DNA could be gathered along these lines: regardless of enhancements in systems, DNA profiles are as yet stacked to the DNA information base from under 1% of recorded violations. Numerous wrongdoing scenes don't uncover any DNA in the purest form (Machado and Silva 2014).

On an extreme, a monstrous expansion in the quantity of people's DNA profiles gathered and DNA information base is set aside in Europe region didn't improve probability of having the option to indict somebody for a wrongdoing. This gives off an impression of being on the grounds that DNA profiles were gathered and keep distant from large-scale individuals (everybody captured for an offense, whether or not they were at last sentenced). The incorporation of a huge number of guiltless individuals' records on the DNA information base likewise brought about lost public trust in policing. In spite of the way that it is difficult to assess the impacts, this may have made a couple of infringement all the more difficult to handle by making a couple of individuals less supportive with police assessments. In Scotland, the standards on the support of DNA profiles kept up open assistance, and their DNA information base stayed a successful device in criminal examinations regardless of most blameless individuals' records being erased (Ahmed 2020).

The DNA sample should be collected from the entire population instead from the ones who have been suspected would clearly cost generously more and furthermore raise down to earth and moral troubles about how to gather DNA from everybody without assent. Gathering DNA from unfamiliar guests would add further to the expenses and challenges which can impact sly affect individuals ready to head out to the nation. Gathering DNA from children upon entering the world would raise genuine moral issues about assent and the part of the clinical calling. Likewise, greater information bases – and more correlations between DNA profiles put away in various nations – improve the probability of bogus matches, as portrayed previously. These can burn through police time following bogus leads, regardless of whether they don't prompt premature deliveries of equity (Kobilinsky et al. 2007).

Conclusion

DNA databases since their inception have been continuously providing help to law enforcement agencies and are very much likely to grow in the coming years. With this, the search will widen the horizon and successfully enlarge the hunt net by including crimes from an extended rundown of wrongdoings which will acquire more DNA samples as new laws become effective. As DNA bases grow in size, various operational reformations will be needed; like additional core loci to avoid extraneous matches, algorithms enabling faster searches will be needed, expanded infrastructure, manpower, training and research resources, etc. There were innocent people who stayed behind bars for years; post-conviction exonerations have become possible for them only with the help of DNA databases. DNA database hits enabled to link serial crimes and helped in locating the perpetrators. This revolutionary tool is open to criticism; however, it holds immense power. Owing to this, it is imperative that it should be handled responsibly and with utmost care to ensure safety of public and their civil liberties.

References

Agreement Between the United States of America and New Zealand [Internet]. Mar 20, 2013. Available from: https://www.state.gov/documents/organization/277051.pdf

Ahmed A (2020) Ethical Concerns of DNA Databases used for Crime Control | Bill of Health [Internet]. Bill of Health. [cited 31 October 2020]. Available from: https://blog.petrieflom.law.harvard.edu/2019/01/14/ethical-concerns-of-dna-databases-used-for-crime-control/

Amankwaa AO (2018) Forensic DNA retention: public perspective studies in the United Kingdom and around the world. Sci Justice 58(6):455–464

Anand P (2020) Forensic DNA is a quick way to bring rapists to justice. But awareness and testing levels are low in India [Internet]. Times of India Blog. [cited 31 October 2020]. Available from: https://timesofindia.indiatimes.com/blogs/voices/forensic-dna-is-a-quick-way-to-bring-rapists-to-justice-but-awareness-and-testing-levels-are-low-in-india/

Anderson C, Stackhouse R, Shaw A, Iredale R (2011) The national DNA database on trial: engaging young people in South Wales with genetics. Public Underst Sci 20(2):146–162

Anker AST, Doleac JL, Landersø R (2017) The effects of DNA databases on the deterrence and detection of offenders. Working paper. Available at http://jenniferdoleac.com/wp-content/uploads/2015/03/DNADenmark.pdf

Belgium identified over 1,100 people through DNA databases in 2019. (2020, January 13). Retrieved November 04, 2020, from https://www.brusselstimes.com/news/belgium-all-news/88970/belgium-identified-over-1100-people-through-dna-databases-in-2019-adn-forensics-missing-persons-suspects-convicts-profile/

Biondo R, De Stefano F (2011) Establishment of Italian national DNA database and the central laboratory: some aspects. Forensic Sci Int 3(1):e236–e237

Borja-Santos R (2015) Base de dados de perfis de ADN só conseguiu 5393 amostras em cinco anos. Retrieved November 04, 2020, from https://www.publico.pt/2015/03/28/sociedade/noticia/base-de-dados-de-perfis-de-adn-so-conseguiu-5393-amostras-em-cinco-anos-1690613

Carracedo A (2008) Forensic DNA typing protocols, 5th edn. Human Press, New Jersey

CODIS – NDIS Statistics (2016) Retrieved November 04, 2020, from https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics

Council of the European Union. Draft Council Conclusions on the implementation of the "PR €UMDECISIONS" ten years after their adoption. http://data.consilium.europa.eu/doc/document/ST-10550-2018-INIT/en/pdf
Criminal Investigations (Bodily Samples) Act 1995 (n.d.) Retrieved November 04, 2020, from http://www.legislation.govt.nz/act/public/1995/0055/latest/DLM368904.html
Cyranoski D (2004) Japan's ethnic crime database sparks fears over human rights
da Silva LAF, Vilaça W, Azevedo D, Majella G, Silva IF, Silva BF (2009) Missing and unidentified person's database. Forensic Sci Int 2(1):255–257
De Moor S (2018) Forensic DNA databases as data sources for criminological research (Doctoral dissertation, Ghent University)
de Wet S, Oosthuizen H, Visser J (2011) DNA profiling and the law in South Africa. Potchefstroomse Elektroniese Regsblad 14(4):171–207
DNA (2020) Retrieved November 04, 2020, from https://www.interpol.int/How-we-work/Forensics/DNA
Ferreira ST, Paula KA, Maia FA, Svidzinski AE, Amaral MR, Diniz SA, Moraes AV (2013) A study of the first DNA database of biological evidence from sexual assaults and rapes in Brazil. Forensic Sci Int 4(1):e368–e369
Flaus A (2013) Familial searches and the New Zealand DNA profile databank: the thin edge of the genetic wedge
Florida International University – Digital Communications (2020) National Missing And Unidentified Persons System. Retrieved November 04, 2020, from https://www.nfstc.org/programs/namus/index.html
Frank PH, Găină RM (2014) Political corruption in Romania: the DNA database, 2014. Document disponibil în limba engleză, la: http://www.peterhfrank.com/political-corruption-in-romania-the-dnadatabase/. Ultima accesare, 30
Ge JY, Yan JW, Xie Q, Sun HY, Zhou HG, Li B (2013) Development of Chinese forensic Y-STR DNA database. Fa Yi Xue Za Zhi 29(3):212–215
Ge J, Sun H, Li H, Liu C, Yan J, Budowle B (2014) Future directions of forensic DNA databases. Croat Med J 55(2):163
Global summary (2020) Retrieved November 04, 2020, from http://dnapolicyinitiative.org/wiki/index.php?title=Global_summary
Harbison SA, Hamilton JF, Walsh SJ (2001) The New Zealand DNA databank: its development and significance as a crime solving tool. Sci Justice 41(1):33
Heathfield LJ (2014) Policy required for entry of DNA profiles onto the National Forensic DNA database of South Africa. S Afr J Sci 110(7–8):1–3
Herkenham M (2006) Retention of offender DNA samples necessary to ensure and monitor quality of Forensic DNA efforts: appropriate safeguards exist to protect the DNA samples from misuse. J Law Med Ethics 34(2):380–384
http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2019/0024/COM_COM(2019)0024_EN.pdf
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/778065/National_DNA_Database_anuual_report_2017-18_print.pdf
https://www.interpol.int/en/content/download/4875/file/GlobalDNASurvey.pdf
Interpol (2008) Annual report 2008. https://www.interpol.int/en/content/download/5253/file/AnnualReport2008-EN.pdf. Retrieved 6 June 2021
Interpol (2011) Annual report 2011. https://www.interpol.int/content/download/10958/file/Annual%20Report%202011-EN.pdf?inLanguage=eng-GB. Retrieved 6 June 2021
Interpol (2019) Global DNA profiling survey results 2019. https://www.interpol.int/fr/content/download/15469/file/INTERPOL%20Global%20DNA%20Profiling%20Survey%20Results%202019.pdf. Retrieved 6 June 2021
Issues T (2020) The Issues. [online] Forensic genetics policy initiative. Available at: http://dnapolicyinitiative.org/about-us/. Accessed 2 Nov 2020

Johnson P, Williams R (2004) DNA and crime investigation: Scotland and the 'UK National DNA database'. Scott J Crim Justice Stud 10:71
Kobilinsky L, Levine L, Margolis-Nunno H (2007) Forensic DNA analysis. Chelsea House, New York
Kumar S, Verma AK, Singh P, Singh R (2016) Current scenario of forensic DNA databases in or outside India and their relative risk. Egypt J Forensic Sci 6(1):1–5
Locarno L, Corradi D, Marino M (2019) Start-up of the criminal genetic database in Mendoza, Argentina. Forensic Sci Int 7(1):100–102
Machado H, Silva S (2014) "Would you accept having your DNA profile inserted in the National Forensic DNA database? Why?" results of a questionnaire applied in Portugal. Forensic Sci Int Genet 8(1):132–136
Maguire CN, McCallum LA, Storey C, Whitaker JP (2014) Familial searching: a specialist forensic DNA profiling service utilising the national DNA database® to identify unknown offenders via their relatives – the UK experience. Forensic Sci Int Genet 8(1):1–9
Marchese V, Cerri N, Caenazzo L (2013) Italian National Forensic DNA database in an European perspective. Forensic Sci Int 4(1):e246–e247
Milot E, Lecomte M, Germain H, Crispino F (2013) The national DNA data bank of Canada: a Quebecer perspective. Front Genet 4:249
Ministerie van Justitie en Veiligheid (2018) Jaarverslag 2017. Retrieved November 04, 2020, from https://dnadatabank.forensischinstituut.nl/dna-dossier/documenten/publicaties/2018/11/01/jaarverslag-2017
Ministerie van Justitie en Veiligheid. (2019) Databank vermiste personen. Retrieved November 04, 2020, from https://dnadatabank.forensischinstituut.nl/dna-databanken/dna-databank-vermiste-personen
Ministerie van Justitie en Veiligheid (2020) NFI research method included in FBI DNA databank system. Retrieved November 04, 2020, from https://www.forensicinstitute.nl/news/news/2020/02/05/nfi-research-method-included-in-fbi-dna-databank-system
Montague S (2011) Genetic suspects: global governance of Forensic DNA profiling and Databasing. Yale J Biol Med 84(3):326
Penacino GA (2008) The new genetic database of Argentina. Forensic Sci Int 1(1):658–659
Perepechina IO (2019) Forensic DNA registration in the Russian Federation: background and the current state. Forensic Sci Int 7(1):688–689
Santos F, Machado H, Silva S (2013) Forensic DNA databases in European countries: is size linked to performance? Life Sci Soc Policy 9(1):12
Smith M, Mann M (2015) Recent developments in DNA evidence. Trends Issues Crime Crim Justice 506:1
The National DNA Database Is Watching You (2020) Retrieved November 04, 2020, from https://www.sydneycriminallawyers.com.au/blog/the-national-dna-database-is-watching-you/
Toom V (2018) Cross-border exchange and comparison of forensic DNA data in the context of the Prüm decision. Civil liberties, justice and home affairs
Toom V, Granja R, & Ludwig A (2019). The Prüm decisions as an aspirational regime: Reviewing a decade of cross-border exchange and comparison of forensic DNA data. *Forensic Science International: Genetics, 41,* 50–57
Voultsos P, Njau S, Tairis N, Psaroulis D, Kovatsi L (2011) Launching the Greek forensic DNA database. The legal framework and arising ethical issues. Forensic Sci Int Genet 5(5):407–410
Wee S (2020) China is collecting DNA from tens of millions of men and boys, Using U.S. Equipment. Retrieved November 04, 2020, from https://www.nytimes.com/2020/06/17/world/asia/China-DNA-surveillance.html
Williams R, Johnson P (2004) 'Wonderment and dread': representations of DNA in ethical disputes about forensic DNA databases. New Genet Soc 23(2):205–223
Williams R, Johnson P (2008) Genetic policing: the use of DNA in criminal investigations

InDel Loci in Forensic DNA Analysis

47

Tugba Unsal Sapan

Contents

Abstract

As a part of forensic science, in the investigations of forensic cases, such as murder, sexual assault, and theft, DNA profile, which is obtained from biological materials found at the crime scene, is used to determine whether there is a connection between the suspect people and the offense. Depending on the type of case, autosomal or gonosomal polymorphism, which is obtained from X and Y chromosome, can be used in solving forensic cases. Although STR (Short Tandem Repeats) loci is still widely used in routine forensic genetic analysis, in many cases, typing problems can occur in highly degraded biological samples collected from the crime scene, and results cannot be obtained. In order to solve these problems, forensic scientists have been looking for alternative genetic markers.

InDels, which have been used in forensic science in recent years, can achieve more successful results in forensic identification when used together with STR and SNP loci. InDels are the most common type of polymorphism in the genome

T. Unsal Sapan (✉)
Institute of Addiction and Forensic Sciences, Uskudar University, Istanbul, Turkey
e-mail: tugba.unsalsapan@uskudar.edu.tr

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_49

after SNPs. InDel occurs as a result of insertion and/or deletion of one or more nucleotides, and it is a kind of polymorphism which can be used in population studies since insertion/deletion is seen with a frequency of more than 1% of the population. These loci can be used for identification, biogeographic genealogical research analysis, evolutionary research, and revealing kinship relationships since they allow to do multiplexing PCR study and have high heterozygosity rate and also small amplicon lengths (60–200 bp). Research on the use of InDels in forensic science has gained momentum in recent years. The number of ready-made commercial kits for these loci is quite limited. For this reason, researchers create their own InDel multiplex panels to use in their studies.

Keywords

Forensic biology and genetics · Human identification · Insertion/deletion (InDel) polymorphism · InDel loci

Introduction

Advances in molecular genetics in the 1980s allowed the study of polymorphic traits directly at the DNA level. The use of DNA in forensic science developed rapidly after Alec Jeffrey's discovery of polymorphic repeat sequences in the DNA molecule in 1985. Since then, the identification of biological samples collected from both individuals and from the crime scenes in the determination of paternity-kinship relations and other criminal investigations has carried out using DNA analysis techniques. In forensic cases, such as terrorism, murder, sexual assault, and theft, it is possible to link up between the suspect and the crime scene by using DNA profiles, which are isolated from biological materials detected at the crime scene (Jeffreys et al. 1985; Robertson et al. 1990; Chan 1992). Although VNTR (variable number of tandem repeats) loci, which were used in the first period of DNA analysis in forensic sciences, have a high discrimination power, it has been replaced by new technologies on account of the need of good qualified (non-fragmented) and excessive amounts (300–500 ng) of DNA, long and laborious analysis times, exposing radioactive materials, and so on. For the last 20 years, STR loci have been widely used in forensic identification (Lee et al. 1994; Robertson et al. 2002).

Short tandem repeats (STR) loci have been used as ideal genetic markers in forensic science in recent years due to their small amplicon size, successful results in degraded biological samples, capability to do multiplex analysis, and no requirement of expensive equipment (Weber and May 1989; Edward et al. 1992). Although STR loci have been still used today, typing problems are experienced in extremely degraded biological samples, which are collected from the crime scene, and successful results for comparison cannot be obtained many times. In order to solve this problem, mini short tandem repeats (miniSTRs) that allow typing even in degraded samples were put on the market in the early 2000s (Coble and Butler 2005). In recent years, for the same purpose, researchers have been focused on

different DNA polymorphisms, such as single-nucleotide polymorphism (SNP), which takes up a very little space in DNA, and InDel polymorphism (Pereira et al. 2009a, b).

Insertion

It is a type of mutation that disrupts the natural sequence as a result of addition of one or more bases to the DNA base sequence. This addition can be a base or as much as a whole chromosome (Fig. 1). As a result of DNA polymerase shift, it is usually thought to be formed by addition of base sequences, which are not adjacent to microsatellite sites, to the main sequence (Gelbart et al. 2002; Kondrashov and Rogozin 2004; Rodriguez-Murillo and Salem 2013).

Deletion

It is a type of mutation that occurs as a result of the deletion of one or more bases in the DNA sequence. Deletion, as well as insertion, can occur on one or more bases, as well as in chromosome size (Fig. 2). If a deletion occurs when part of the

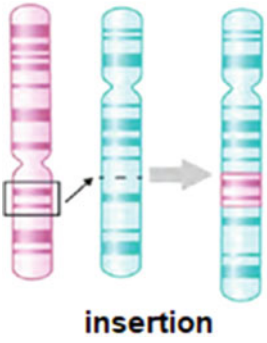


Fig. 1 Insertion on a chromosome level (Gordon and Egner 2013)

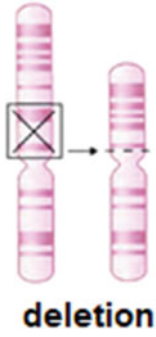


Fig. 2 Deletion on a chromosome level. (Gordon and Egner 2013)

chromosome breaks off and disappears, it can lead to serious genetic diseases (Gelbart et al. 2002; Kondrashov and Rogozin 2004; Rodriguez-Murillo and Salem 2013).

InDel (Insertion/Deletion)

Point or gene mutations are defined as changes that occur in the DNA sequence and can be passed on to subsequent generations. These mutations usually occur in one or more nucleotides and cause changes in the structure of the genome. Point mutations can also occur in the form of insertion or deletion (Campbell et al. 2006) (Fig. 3). However, insertion or deletion mutations are much more important changes than point mutations. Because insertion or deletion of one or more bases in the DNA chain usually leads to the shift of the genetic code that starts from the point where the insertion or deletion occurs, this causes polymorphism by creating significant changes in the structure of the gene (Pereira et al. 2009a, b).

InDels and somatic/gonosomal chromosome mutations are terms used to express mutation combinations, which include deletion or insertion separately or together, in the studies of forensic molecular genetics, evolution, and population genetics (Gelbart et al. 2002; Kondrashov and Rogozin 2004; Gregory 2004).

InDels, which account for 16–25% of all genetic variations in the genome, are the most commonly seen DNA polymorphism after SNPs, and 1.6–2.5 million InDel polymorphisms have been identified in human population studies. This, in turn, suggests that hereditary changes that occur as a result of insertion or deletion (InDel) mutations can be used as a genetic marker since they are seen in the human genome frequently. Even though they are so common, studies with InDels are limited (Hongbao 2005). However, since InDels have power of discrimination and heterozygosity, it is possible to use InDels in studies of human identification, ancestry, and evolutionary and molecular anthropology. For this reason, today, it

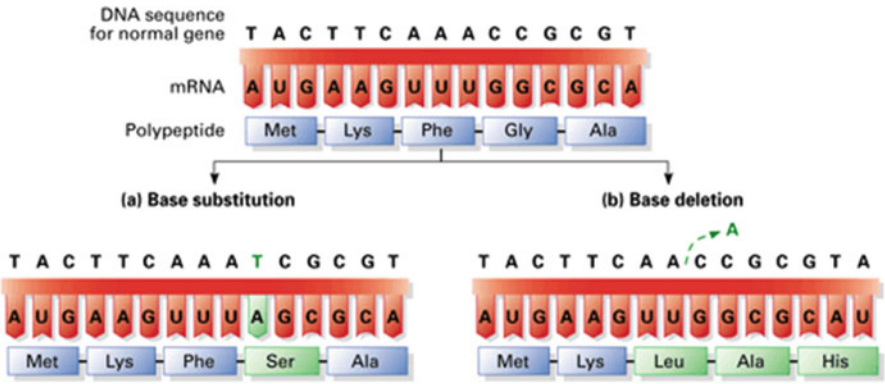


Fig. 3 Schematic representation of insertion and deletion on the DNA sequencing. (Gordon and Egner 2013)

is an alternative polymorphism to SNPs and STRs, which are used in DNA profiling in forensic identification (Reiner et al. 2005; Pereira et al. 2009a, b, 2012; Martínez-Cortés et al. 2015).

Use of Insertion/Deletion Polymorphism in Forensic Genetics

Insertion/deletion polymorphism (InDel) is known to be as a length polymorphism, which formed as a result of insertion and/or deletion of one or more nucleotides on the genome. Differences are used to discrimination of two people from each other in forensic identification, which makes polymorphisms the basis of identification (Rodriguez-Murillo and Salem 2013).

Polymorphism occurs as a result of successive mutations and is passed down from generation to generation according to Mendelian laws. In 2002, James Weber and colleagues identified over 2000 biallelic insertion/deletion polymorphisms on the human genome. Thus, studies of InDel polymorphism began for the first time in forensic sciences (Weber and May 1989). InDel loci have been used in forensic identification since they are able to study with multiplex PCR and have high heterozygosity and small amplicon lengths (60–200 bp) (Fig. 4). They can also be used to identify disaster victims in mass fatality cases, such as aircraft accidents, terrorist attacks or natural disasters, ancestry determination, evolutionary research, and molecular anthropology (Manta et al. 2012; Martínez-Cortés et al. 2015). More successful results can be achieved in identification by using InDel polymorphism, which are called next-generation genetic variations together with STR and SNP loci (Sanchez et al. 2006; Pereira et al. 2012).

InDel loci have been used in forensic sciences, especially in recent years. Commercial kit production for InDel loci is highly restricted. So forensic scientists often create their own InDel panels and try to popularize them (Guangyao et al. 2015).

Commercial Kits Based on InDel Analysis

InDel loci have been used in forensic science in recent years. However, InDel kit production is also quite limited. Currently, there are three commercial InDel kits made for the use in forensic science.

Investigator® DIPplex Kit

It is a kit containing 30 InDel loci found in somatic chromosomes that have been marketed by QIAGEN firm in the last 2–3 years. The kit also includes the locus of amelogenin. Selected InDel loci are smaller than 160 bp and are specifically designed for the use in the identification or anthropological research. In addition, they are analyzed by ABI 310, 3130, 3130XL, 3500, and 3500XL capillary electrophoresis (Investigator® DIPplex Handbook 2014).

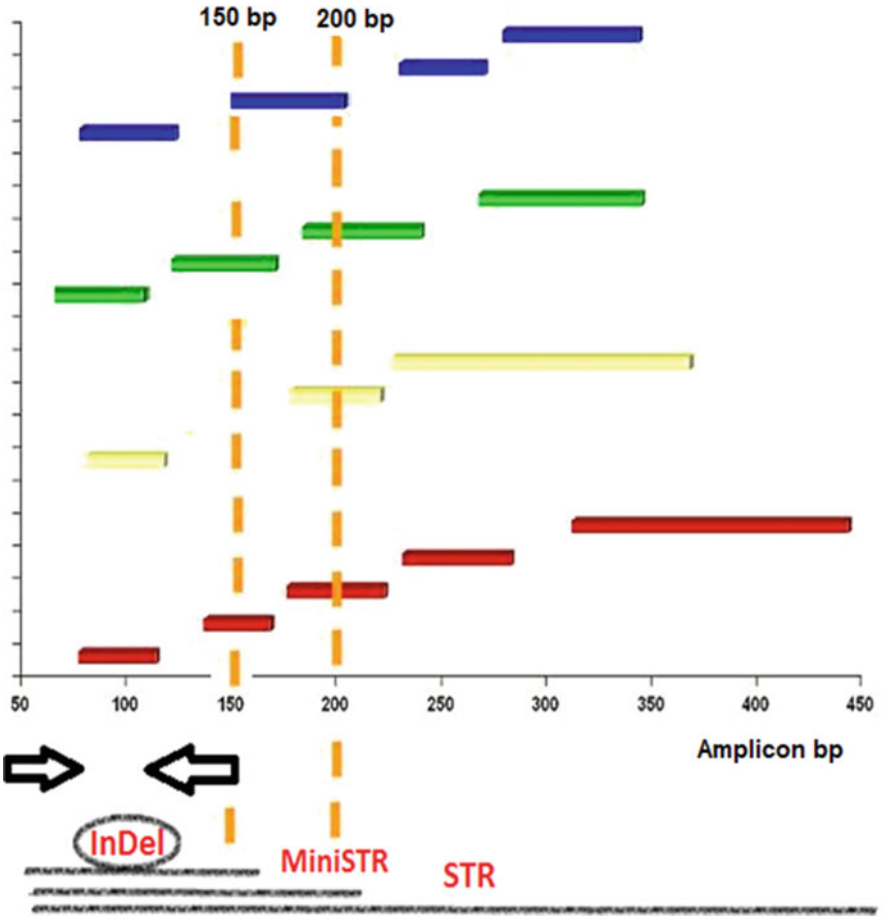


Fig. 4 Comparison of PCR product sizes of InDels with STRs

InDelPlex INDEL Polymorphism Detection Kit

It is a kit developed by Pereira R. et al. in partnership with the Institute of Molecular Immunology and Pathology of Oporto University and the University of Santiago de Compostela and commercialized with Genomica firm over the past few years. The kit allows multiplex PCR amplification of 38 InDel region and produced for the use of forensic identification and clinical diagnostic purposes. These analyses can be done by using ABI 310, 3130, 3130XL, 3500, and 3500XL capillary electrophoresis (Pereira et al. 2009a, b, 2012).

Mentype® DIPplex PCR Amplification Kit

It is a kit containing 30 InDel loci found on somatic chromosomes that have been marketed in recent years by the Biotype firm and also includes the locus of amelogenin. This kit restricted amplicon length to ~150 bp, which makes the kit

perfectly suitable for analyzing critical stains. The kit can be analyzed by using ABI 310, 3130, 3130XL, 3500, and 3500XL capillary electrophoresis (Mentype® DIPplex PCR Amplification Kit Handbook 2009).

Use of Gonosomal InDel Loci in Forensic Genetics

InDel loci exist on both somatic and sex chromosomes. Compared to somatic chromosomes, there are InDel variations in similar ratios on the X and Y chromosomes too. It can be used in addition to autosomal genetic markers in father/daughter relationship in X-linked polymorphism analysis. In the cases of paternity of two relative men (such as father/son), paternity can be determined by the use of polymorphism on the X chromosome if the child is a girl. It is because suspicious fathers will have different X chromosomes since they have different mothers (Szibor et al. 2005; Prinz and Sansone 2001).

Since the Y chromosome is transferred unchanged from father to son, it is observed in male members of the same family in the same form, except for mutations and genetic abnormalities. For this reason, population genetics, biogeographic lineage determination, male kinship evaluation or link analysis, and forensic genetic identification studies can be done by utilizing Y chromosome-related polymorphism analysis. In paternity cases, if the child is a boy, it is possible to get results by typing Y chromosome loci of any man (grandfather, uncle, cousin, etc.), who is in the family tree of the father candidate, especially when DNA cannot be obtained for various reasons (such as the father candidate cannot be found, DNA cannot be obtained from his biological material or he is dead, etc.) (Roewer et al. 2000; Corach et al. 2001; Prinz and Sansone 2001). Similarly, in paternity cases, where the father candidate is dead or could not be found, it is possible to get results by analyzing the X-linked polymorphism between the grandmother candidate and child (if the child is a girl) because the child gets one of her X chromosomes that her father gets from her grandmother. As a result, the paternal X chromosome in a girl will necessarily be coming from one of her grandmother's X chromosomes (Szibor et al. 2005, Prinz and Sansone 2001).

Pregnancies that occur after sexual assault crimes can result in abortion. In 6–8 weeks of abortions, the tissue and maternal blood materials are found together with miscarriage, and it is difficult to separate them microscopically. In this case, if the fetus is female, the presence of the X chromosome, which was inherited from the suspected father, can be determined by X-linked polymorphism analysis (Szibor et al. 2005; Prinz and Sansone 2001).

The Y chromosome is also used in illuminating sexual assault crimes. In cases where the victim is a woman, the victim's vaginal swab sample contains a mixture of DNA belonging to the victim and the perpetrator. Since there is no Y chromosome in women, only the profile of the perpetrator is obtained when identification is performed using Y chromosome loci in the mixture DNA (Roewer et al. 2000; Corach et al. 2001; Prinz and Sansone 2001).

By analyzing the polymorphism associated with X chromosome, it can also be determined whether two girls' father is the same, regardless of whether their mothers are different or the same. Since the father has one X chromosome, he will pass on the same X chromosome to all girls. But if their father is not the same, the X chromosome loci transferred from their father will also be different (Szibor et al. 2005; Prinz and Sansone 2001).

By using gonosomal and autosomal InDels together with STR and SNP loci, more successful results can be achieved in such cases or in identification of samples, which were collected from the crime scene (Szibor et al. 2005; Prinz and Sansone 2001).

Analysis of InDel Loci

Although InDel loci are biallelic, it is analyzed by the method of fragment analysis, such as STRs. These steps are followed in the identification of InDels as in STRs: first, extraction and quantification of DNA sample, then amplification of InDel loci of DNA by using fluorescence-marked InDel primers in multiplex PCR, and, finally, obtaining the profile of person by separating these loci in capillary electrophoresis with ABI 310, 3130, 3130XL, 3500, and 3500XL devices.

As a result of the analysis of InDel loci in capillary electrophoresis, three types of alleles can be seen in electropherogram because either deletion, insertion, or a combination of both can be transferred from the mother and father. These are expressed as deletion (Del – minor allele), insertion (Ins – major allele), and insertion-deletion (InDel), where both deletion and insertion are transferred together. The base pair size of the identified InDel loci on the DNA is defined by researches. Since it is defined how many base deletions or insertions they undergo, determining only the alleles as insertion or deletion will be sufficient to say the number of base pairs.

Before the analysis in capillary electrophoresis, this base pair data is defined in panel manager. It is created with a panel for each dye color (primer) with bins. In this way, the data of all alleles that exist in the population, just like in analysis of STR loci, is defined into the device. In capillary electrophoresis of DNA, if a person's peak was observed in the first bin, as shown in the figure, this shows the person's allele is "Del" because it has an incomplete base sequence. If the insertion has been transferred from the parent, the peak will be observed in the second bin, as shown in Fig. 5, which means that the person has the insertion as a result of addition of base, and the allele is "Ins." If a peak is observed in both bins, it means that the person has received both an insertion and deletion polymorphism from the parent. This, in turn, refers to the third allele, which we call "InDel" (Fig. 5).

In this way, the profile of people analyzed with InDel multiplex panel kits in all regions can be determined just like in STR analysis.

The number of loci that will reach sufficient (99.9%) discrimination power for human identification of InDel loci differs from STR loci number. In order to use InDels for identification purposes, as in STR analysis, a sufficient number of loci to

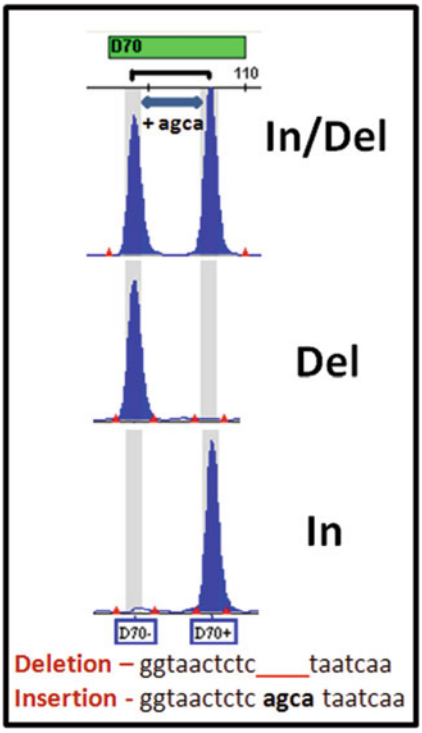


Fig. 5 Electropherogram image of insertion/deletion (InDel) polymorphism. (Fondevila Álvarez et al. 2011)

use are determined, depending on the heterozygosity rates and discrimination power of the locus contained in the multiplex kit or panel. The number of loci contained in InDel kits or panels developed by researchers, which exist on the market today, has a 99.9% discrimination power, and this locus number ranges from 20 to 40 (Ünsal et al. 2017).

While expressing D70+ as the insertion allele in the figure, D70- refers to the deletion allele.

Conclusion

Currently, STR systems are widely used for human identification in criminal laboratories. Forensic scientists are developing new polymorphic systems that can obtain results from all kinds of biological samples as an alternative to polymorphic systems (Coble and Butler 2005).

Insertion/deletion polymorphism occurs in the form of the addition or loss of one or more bases in the human genome, which is causing polymorphism and can be used in human identification and genealogical determination in illuminating forensic cases. Especially in cases where DNA obtained from biological evidence from the

crime scene is degraded or trace amount; As an alternative to STRs, small-sized systems on DNA are preferred for analysis in forensic sciences. In such cases, successful identification results can be obtained by using InDel polymorphism (Pereira et al. 2009a, b, 2012).

The loci that these systems occupy on DNA are quite small (60–200 bp), and the potential for a successful DNA profile is high, even if the DNA is degraded. The amount of DNA that can be obtained from a degraded or trace biological sample is either non-existent or ranges from 100 pg to 1 ng. Since it is known that the amount of DNA required for multiplex PCR used for identification studies is 0.5–1 ng, it has been determined by studies that identification with InDel loci gives successful results.Therefore InDels have been used in forensic science in recent years, and studies continue to develop InDel panels as an alternative or complement to STR and SNP loci in identification (Ünsal et al. 2017).

References

Campbell N.A, Williamson, B., Heyden, R.J. (2006) Biology: exploring life. Boston: Pearson Prentice Hall, 11(6): 242–243

Chan L (1992) Advances in molecular biology with applications in clinical medicine. Klin Lab 38: 2–4

Coble MD, Butler JM (2005) Characterization of new MiniSTR loci to aid analysis of degraded DNA. J Forensic Sci 50:43–53

Corach D, Risso FL, Marino M, Penacino G, Sala A (2001) Routine YSTR typing in forensic casework. Forensic Sci Int 118:131–135

Edward SA, Hammond HA, Jin L, Caskey T, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeats. Genomics 12:241–253

Fondevila Álvarez M, Pereira R, Gusma L, Phillips C, Lareu MV, Carracedo A, Butler JM, Vallone PM (2011) Forensic performance of short amplicon insertion-deletion (InDel) markers. Forensic Sci Int Genet Suppl Series 3:e443–e444

Gelbart WM, Lewontin RC, Griffith JF, Miller JH (2002) Modern genetic analysis: integrating genes and genomes. W.H. Freeman and CO., New York

Gordon A, Egner S (2013) Dynamic biology for high school. Sapling Learning, Austin

Gregory TR (2004) Insertion-deletion biases and the evolution of genome size. Gene 324:15–34

Guangyao F, Yi Y, Haibo L, Yiping H (2015) Screening of Multi-InDel markers on X-chromosome for forensic purpose. Forensic Sci Int Genet 5:42–44

Hongbao M (2005) Development application of polymerase chain reaction (PCR). J Am Sci 1(3):1–47

Investigator® DIPplex Handbook (2014) For multiplex amplification of 30 deletion/insertion polymorphisms, plus Amelogenin. Qiagen. https://www.qiagen.com/us/resources/download.aspx?id=2e1b66cf-e497-4054-82fc-3e1d2e14faba&lang=en. Accessed 12 Sep 2020

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

Kondrashov AS, Rogozin IB (2004) Context of deletions and insertions in human coding sequences. Hum Mutat 23(2):177–185

Lee HC, Ladd C, Bourke MT, Pagliaro E, Tirnady F (1994) DNA typing in forensic science I. Theory and background. Am J Forensic Med Pathol 15:269–282

Manta F, Caiafa A, Pereira R, Silva D, Amorim A, Carvalho EF, Gusmão L (2012) INDEL markers: genetic diversity of 38 polymorphisms in Brazilian populations and application in a paternity investigation with post mortem material. Forensic Sci Int Genet 6(5):658–661

Martínez-Cortés G, Gusmão L, Pereira R, Salcido VH, Favela-Mendoza AF, Muñoz-Valle JF, Inclán-Sánchez A, López-Hernández LB, Rangel-Villalobos H (2015) Genetic structure and forensic parameters of 38 Indels for human identification purposes in eight Mexican populations. Forensic Sci Int Genet 17:149–152

Mentype® DIPplex PCR Amplification Kit Handbook (2009) Biotype Diagnostic GmbH

Pereira R, Phillips C, Cíntia AA, Amorim Carracedo A, Gusma L (2009a) A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis 30:3682–3690

Pereira R, Phillips C, Cíntia AA, Amorim Carracedo A, Gusma L (2009b) Insertion/deletion polymorphisms: a multiplex assay and forensic applications. Forensic Sci Int Genet Suppl Series 2:513–515

Pereira R, Phillips C, Pinto N, Santos C, Batista dos Santos SE, Amorim A, Carracedo A, Gusma L (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS One 7:e29684

Prinz M, Sansone MY (2001) Chromosome-spesific short tandem repeats in forensic casework. Croat Med J 42(3):288–291

Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, Phong A, Burchard EG, Harris TB, Psaty BM, Kwok PY (2005) Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. Am J Hum Genet 76: 463–477

Robertson J, Ross AM, Burgayne LA (1990) DNA in forensic science: theory. Techniques and applications. Ellis Horwood, London

Robertson J, Ross AM, Burgayne LA (2002) Properties of hypervariable single locus polymorphism and their application to identity testing. In: Balazs I (ed) DNA in forensic science: theory, techniques and applications. Taylor& Francis e-library, London

Rodriguez-Murillo L, Salem RM (2013) Insertion/deletion polymorphism. In Gellman, Marc D., Turner, J. Rick (eds) Encyclopedia of Behavioral Medicine. Springer-Verlag New York 1:1076.

Roewer L, Kayser M, Knijff P, Anslinger K, Corach D, Füredi S, Geserick G, Henke L, Hidding M, Kärgel HJ, Lessig R, Nagy M, Pascali VL, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M (2000) A new method for the evaluation of matches in recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. Forensic Sci Int 114:31–43

Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 9:1713–1724

Szibor R, Hering S, Edelmann J (2005) A new web site compiling forensic chromosome X research is now online. Int J Legal Med 25:1–3

Ünsal T, Filoğlu G, Aşıcıoğlu F, Bülbül Ö (2017) Population data of new 21 mini-InDels from Turkey. Forensic Sci Int Genet Suppl Series 6:e189–e191

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

The Role of DNA Profiling in Landscape of Human Migration

48

J. A. Lorente, Christian Haarkötter, María Saiz, M. I. Medina-Lozano, X. Gálvez, M. J. Alvarez-Cubero, L. J. Martínez-González, B. Lorente-Remon, and Juan Carlos Alvarez

Contents

J. A. Lorente
Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

GENYO, Centre for Genomics and Oncological Research (Pfizer / University of Granada / Andalusian Regional Government), Granada, Spain
e-mail: jlorente@ugr.es

C. Haarkötter (✉) · M. Saiz · M. I. Medina-Lozano · X. Gálvez · J. C. Alvarez
Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain
e-mail: chaarkotter@ugr.es; msaiz@ugr.es

M. J. Alvarez-Cubero · L. J. Martínez-González
Department of Biochemistry and Molecular Biology III, Faculty of Medicine, University of Granada, Granada, Spain

Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), Granada, Spain
e-mail: mjesusac@ugr.es; luisjavier.martinez@genyo.es

B. Lorente-Remon
Program International Liaison Officer (ILO), Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, DNA-PROKIDS, Granada, Spain

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_50

Abstract

Human migration in general and forced human migration in particular is an increasing problem worldwide. The significant number of people involved in migrations and the extent of their needs have attracted the attention of criminal networks. However, the most tragic cases are those individuals who disappear, either physically or "virtually." In both scenarios, forensic science and particularly genetic identification play key roles in solving and preventing such crimes. Human migration and the need for human identification is an immense and growing problem that needs to be addressed using forensic technology such as DNA typing, which has been proven to be effective in solve a large percentage of criminal cases. The aim of this chapter is giving an overview about the internationally accepted terms of human migration as well as the attempts to measure the problem, followed by the discussion of the importance of forensic DNA typing as a powerful tool in order to try to give an answer to this issue. DNA migrant identification databases come as an interesting instrument due to its robustness, power of identification, comparison possibilities, versatility, and managing characteristics. Finally, two implemented programs, DNA-PROKIDS and DNA-ProORGAN, both created by our team in the University of Granada (Spain), will be exposed, as well as a few future perspectives in order to try to give an answer to such a complex problem as human migration.

Keywords

Human migration · Human trafficking · DNA typing · Forensic science · Illegal adoptions · Organ trafficking

Introduction

Human migration in general and forced human migration in particular is an increasing problem worldwide. As a result of natural catastrophes and human-related causes (such as wars and civil strife), the number of immigrants has grown year after year. However, not all migration is related to violence or natural disasters but rather to the pressing need to search for better futures for oneself or one's family.

The significant number of people involved in migrations and the extent of their needs have attracted the attention of criminal networks. Human smuggling and trafficking organizations can be found in regions and along routes with a constant flow of migrants.

A large number of migrants, especially women, are exploited and assaulted during their migration. Sexual abuse and labor exploitation are common during migration, especially for those considered illegal immigrants, who will likely be exploited even after arriving at the destination.

However, the most tragic cases are those individuals who disappear, either physically or "virtually." The physical disappearances occur when individuals are killed or die from whatever cause during migration and are buried without identification in mass or single graves, with no record.

"Virtual" disappearances are those victims whose identity is changed, such as with children who are kidnapped and offered up for illegal adoption or enslaved for sexual or labor exploitation.

In both scenarios, forensic science and particularly genetic identification play key roles in solving and preventing such crimes. Human migration and the need for human identification is an immense and growing problem that needs to be addressed using forensic technology such as DNA typing, which has been proven to be effective in solve a large percentage of criminal cases.

Definitions and Numbers: The Magnitude of the Problem

Given that this chapter concerns the technical and scientific aspects of DNA typing, the legal aspects related to human migration will not be reviewed in depth. Due to the complexity of this topic, however, we will cover the internationally accepted definitions and terms that need to be considered to understand and approach this issue.

Human migration can occur within a country or across borders and, depending on the reason for the migration, can be voluntary or forced. International migration can be legal if all administrative and border control documents and procedures are duly completed. Migration can be termed illegal if any of the former items are missing, incomplete, or falsified. People can illegally enter another country, usually under the guidance of experienced teams of smugglers belonging to criminal networks, as illegal immigrants or as victims of human trafficking (see Fig. 1).

We therefore need appropriate definitions for the terms "migrant," "human smuggling," and "human trafficking." We will therefore follow the internationally accepted consensus definitions.

The United Nations International Organization for Migration (IOM) defines the word "migrant" as an "*umbrella term, not defined under international law, reflecting the common lay understanding of a person who moves away from his or her place of usual residence, whether within a country or across an international border, temporarily or permanently, and for a variety of reasons. The term includes a number of well-defined legal categories of people, such as migrant workers; persons whose particular types of movements are legally defined, such as smuggled migrants; as well as those whose status or means of movement are not specifically defined under international law, such as international students*" (International Organization for Migration 2019).

In 2019, the number of international migrants worldwide was estimated at 272 million, approximately 3.5% of the world's population, 48% of whom were female. In 2018, the global refugee population was estimated at 25.9 million, with more than half of these under 18 years of age. The number of individuals displaced in their own country reached 41.3 million, while the number of stateless persons was estimated at 3.9 million.

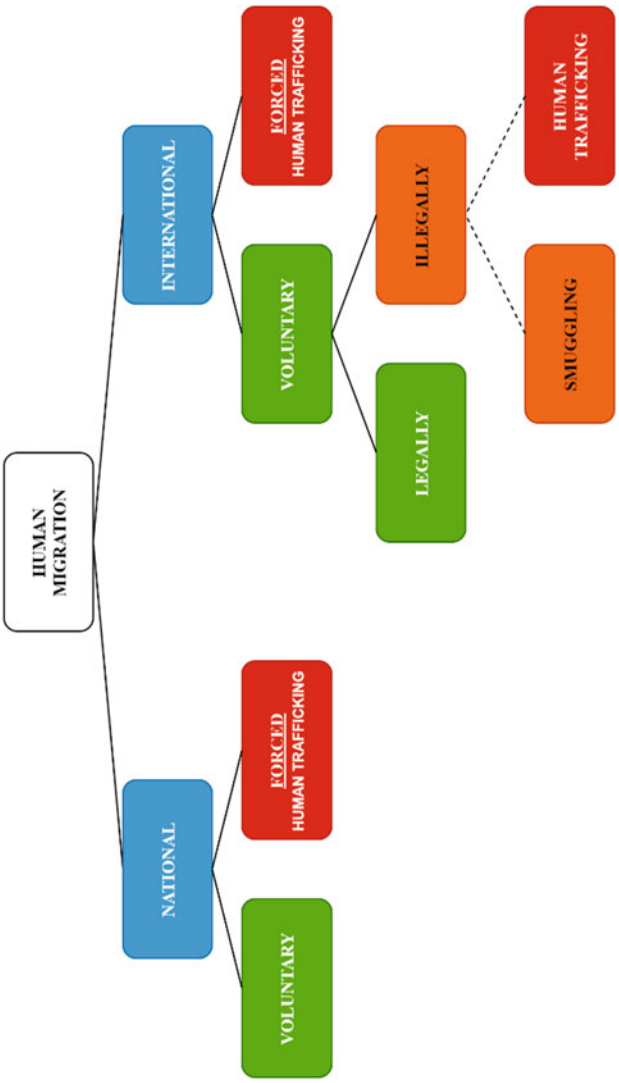


Fig. 1 Types of human migration

Migration patterns differ by region. Most migrants born in Africa, Asia, and Europe remain within their country, while most migrants from Latin America leave their country of birth. The number of migrants is increasing daily, and the patterns of migration are changing (International Organization for Migration 2021).

Since 1990, there has been an increase in the number of migrants. In 2020 alone, the IOM identified 22,221,538 individuals (11,651,019 males and 10,570,519 females) as migrants (United Nations. Department of Economic and Social Affairs. Population Division 2020).

Human trafficking involves the recruitment, movement, or harboring of people for purposes of exploitation (sexual exploitation, forced labor, slavery, or organ removal). Victims are trafficked through threats or use of violence, fraudulent schemes, deception, and abuse of power. Human trafficking can occur within a country or across borders (United Nations Office on Drugs and Crime 2018).

In contrast, *migrant smuggling* occurs only across borders and consists of assisting migrants in entering or staying in a country illegally. Given that migrants consent to being smuggled by irregular methods into countries, they are not considered victims. However, smuggled migrants are frequently placed in dangerous situations (e.g., crossing rivers, seas, deserts, and wild forests with their respective hazards). Migrants are also often exploited, sexually abused, harmed, and even killed, due to the smugglers' sense of impunity resulting from the migrants' inability and/or unwillingness to report them (United Nations Office on Drugs and Crime 2018).

Among the many potential scenarios, the most common characteristics of immigrants are their long distance from their homes and their escape from violence, hunger, and extreme poverty. Migrants travel alone or with their families, often times without their personal belongings and without protection, a perfect scenario for unscrupulous individuals to exploit the vulnerable with impunity.

Missing Persons and Migrations: A Difficult Crime to Solve

One of the most tragic situations in migration is the case of individuals who disappear during transit. This relevant but underestimated and complex problem presents a real challenge for the international community.

In 2014, the IOM described this problem as "an epidemic of crime and abuse" and sought to coordinate efforts and data, creating the IOM Missing Migrants Project (https://missingmigrants.iom.int/), which responds to the needs of the 2030 Agenda for Sustainable Development, specifically 10.7.: "*Facilitate orderly, safe, regular and responsible migration and mobility of people, including through the implementation of planned and well-managed migration policies*." (United Nations General Assembly 2015).

According to available worldwide data and estimates regarding this crime, 1754 migrant fatalities were recorded in 2021 alone, with 1689 in 2014, 3522 in 2015, 4510 in 2016, 3653 in 2017, 2844 in 2018, 3556 in 2019, and 1457 in 2020. The lower number in 2020 is most likely the effect of the COVID-19 pandemic. Based on

these data, 22,055 migrants have died from 2014 to the end of June 2021. The main reported causes of death were drowning, hypothermia, dehydration, starvation, vehicle accidents, and shootings. More relevant data divided by region, country of origin, country of destination, age, sex, etc., are available at the referenced webpage (https://missingmigrants.iom.int/).

There are no reliable data on the percentage of deceased migrants who have been and who are regularly identified, but the number is low. Forensic science plays a unique and essential role in identifying such victims and establishing the cause and date of death. Coordination between forensic anthropology, forensic odontology, and forensic genetics is needed to solve the problems that arise when a dead body or remains are found.

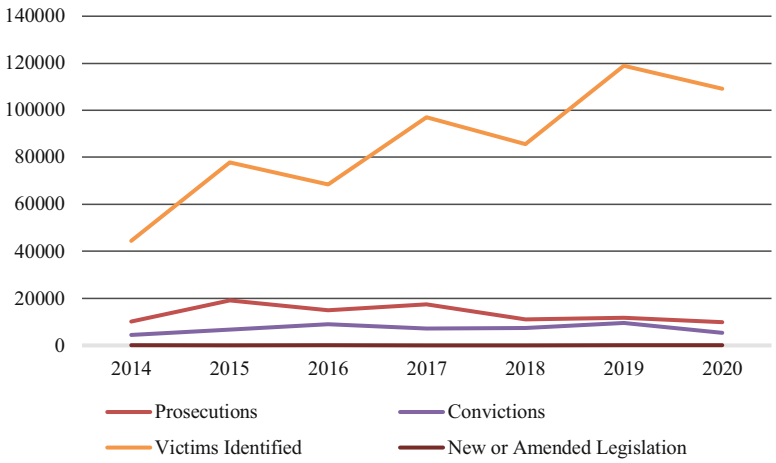
Human Trafficking: The Most Dangerous Scenario

The internationally accepted definition for "human trafficking" is that adopted by the United Nations Convention against Transnational Organized Crime, in resolution A/RES/55/25 of the 15th of November 2000 at the 55th session of the United Nations General Assembly, originally signed in Palermo, Italy, in December 2000, as defined in Article 3 (United Nations 2000):

> "For the purposes of this Protocol: (a) "Trafficking in persons" shall mean the recruitment, transportation, transfer, harbouring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation. Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation, forced labour or services, slavery or practices similar to slavery, servitude or the removal of organs"; (b) The consent of a victim of trafficking in persons to the intended exploitation set forth in subparagraph (a) of this article shall be irrelevant where any of the means set forth in subparagraph (a) have been used; (c) The recruitment, transportation, transfer, harbouring or receipt of a child for the purpose of exploitation shall be considered "trafficking in persons" even if this does not involve any of the means set forth in subparagraph (a) of this article; (d) "Child" shall mean any person under eighteen years of age.""

According to the 2020 United Nations Office on Drugs and Crime Global Report on Trafficking in Persons, the number of victims detected and reported in 2018 was 49,032 worldwide. Approximately 50% of reported trafficked victims are sexually exploited, 38% are placed into forced labor, 6% are pressed into criminal activities, 1.5% are forced into marriages, and the rest are exposed to various forms of exploitation, including baby selling, organ removal, and other forms of exploitation (United Nations Office on Drugs and Crime 2020).

The US State Department Office to Monitor and Combat Trafficking in Persons publishes the yearly Trafficking in Persons (TIP) Report. According to their data,



Graphic 1 Number of prosecutions, convictions, identified victims, and legislation (Department of State. United States of America 2021)

there were 109,216 victims identified globally in 2020. That same year, only 9876 prosecutions were conducted, resulting in 5271 convictions. According to the report, the number of cases is growing year after year (from 44,462 victims in 2014 to 118,932 in 2019); however, the number of prosecutions is not growing at the same rate, from 10,051 in 2015 to 11,841 in 2019 (Department of State. United States of America 2021), as shown in Graphic 1.

These findings do not reflect the scope of the crime or the efforts of the international community. These cases are difficult to prosecute and subject to corruption and lack of necessary data from the countries of origin. The US State Department's TIP Report therefore classifies countries into three separate tiers, depending on the seriousness with which they operate internally and internationally to prevent and prosecute this crime.

Another concern raised in the 2021 report is related to the COVID-19 pandemic, which has led to significant economic difficulties worldwide, problems that exacerbate the risk for the poor and marginalized, with children and women being especially vulnerable. This pandemic has facilitated human trafficking activities. Alerts and special surveillance need to be enforced in the countries of origin and, in cases of international forced migration, in those countries of transit and origin.

From a genetic, DNA typing perspective, there are three situations where DNA is of major interest: identification of the deceased, identification of trafficked children, and identification of trafficked organs. The procedures and existing programs, as well as the challenges, will be covered in sections "Missing Children: The DNA-PROKIDS Program" and "Organ Trafficking: The DNA-ProORGAN Program" of this chapter.

The Role of Forensic Science: DNA Typing

Forensic science faces two main challenges: identifying the victim and determining the cause of death, the solutions to which will depend on the circumstances and laws of the country where the victim is found.

In terms of the cause of death, forensic pathologists can offer valuable information if the body is well preserved and if a full autopsy, including complimentary analysis and techniques, can be performed. If the body is not well preserved or if only partial remains are found, forensic pathologists will need to work with anthropologists to determine the cause of death. The cause of death could be easy to determine in some cases (e.g., those with severe skull fractures), while being next to impossible in others.

Identifying victims in cases of human migration is a complex problem, mainly due to the lack of data and records with which to compare and create reference databases. Although the international forensic community has the necessary tools for solving most cases, there is a lack of appropriate national and international legislation, insufficient collaboration between agencies, fear among the relatives of victims, and government corruption, which all contribute towards perpetuating this unacceptable situation.

The four main disciplines that can help identify cadavers are forensic pathology, forensic anthropology, forensic odontology, and forensic genetics. Depending on their amount and quality, the data obtained from victims can vary in terms of support, usefulness, and implementation. Provided they coordinate the data, these disciplines can be additive and compatible. The only partially destructive discipline is that of genetics, in those cases in which bone or teeth fragments are pulverized for DNA extraction and analysis. Appropriate coordination among the various specialists in terms of workflow is therefore essential.

Forensic genetics has a number of advantages when implemented worldwide. Although not intrinsically better than other methodologies, genetics offers unique tools for solving cases of persons gone missing during migration and in human-trafficking crimes.

The Advantages and Need for DNA Migrant Identification Databases

We are not proposing building and using only DNA databases for identifying missing persons but rather the generation and storage of all types of records that can be used in the context of forensic pathology, odontology, and anthropology. Forensic DNA does not always yield useful results, and biological samples from relatives are not always available.

The main advantages of DNA typing for identifying missing persons during migration are detailed below. For human trafficking that includes missing children, illegal adoptions, and organ removal, DNA offers unique features compared with the other mentioned disciplines. Genetic identification is a robust, reliable, universal

technology that has a high power of identification, has the ability to compare different relatives, can be used with all types of biological samples, and facilitates the creation of databases.

Robust, Reliable, Universal Technology

Forensic DNA typing has become the gold standard in human identification since its introduction in the mid-1980s following the publications by Alec Jeffreys using restriction fragment length polymorphism techniques and especially after the adoption of polymerase chain reaction-based techniques and the use of internationally validated and accepted commercial kits. The International Society of Forensic Genetics (ISFG) has from the start played a key role in harmonizing the technical and scientific advances in forensic DNA typing. Following the ISFG recommendations, commercial companies have developed kits with a common set of loci, and countries now have compatible databases. There are continuous advances being applied to human genetic identification (Parson et al. 2016). The development of the Combined DNA Index System (CODIS) by the US Federal Bureau of Investigation in the mid-1990s was another milestone. This system has been donated to numerous countries and is continually expanding (Ge et al. 2012). Regional forensic networks, such as the Asian Forensic Sciences Network and the Ibero-Latin American Academy of Criminalistics and Forensic Studies and its DNA Working Group (GITAD) have also played a key role in harmonizing scientific protocols (Lorente 2001), with a significant focus on quality and the tracking of database development (Carneiro da Silva Junior et al. 2020). Other countries, such as China, have adapted their scientific developments and loci to be compatible with existing international datasets, such as CODIS and the European Standard Set (Guo et al. 2014). The use of the highest standards and the most demanding procedures, required by ISO certification and accreditation, positioned DNA analysis as one of the most reliable and trusted technologies for human identification. One of the goals of The International Forensic Strategic Alliance, a global network of regional forensic science organizations, is to "encourage the exchange of information related to experience, knowledge and skills between the member networks and other operational forensic experts as appropriate" and could play a role in facilitating data exchange.

In short, DNA typing is a reliable, compatible, and universal technology whose data can be easily and anonymously compared among countries.

Power of Identification

Combining the results of the analysis of autosomal DNA, Y and X chromosomes and mitochondrial DNA should theoretically provide likelihood ratios showing the biological and genetic relationships.

There are, unfortunately, certain limitations in the context of missing persons during migration and in human trafficking-related problems. Generating sufficient genetic information from samples recovered at sea, from single or mass graves can be problematic. Samples are at times highly compromised, and the quality and quantity of the DNA is very low. There have been numerous published scientific

papers on overcoming this problem using novel strategies and approaches (Zeng et al. 2019; Emery et al. 2020). Certain authors, such as David Caramelli et al. at the University of Florence, have focused their efforts on identifying the best source for DNA, suggesting petrous bone as an ideal source (Pilli et al. 2018). Other authors have employed powerful next-generation sequencing-based technologies to generate as much genetic information as possible (Zeng et al. 2019; Elwick et al. 2019).

Once forensic data have been generated from unidentified remains or cadavers, the data should be compared with that from potential relatives, which can be a challenge and requires caution. Even in mass disaster cases where most if not all victims are known, a statistical analysis might be difficult, especially if there are biologically related individuals or if the relatives to be compared with the victims are genetically distant (Ge et al. 2010; Ge and Budowle 2021; Marsico et al. 2021). The most difficult situations, however, are those where the human remains to identify are tentatively from many different countries or geographical regions, given that in most of these cases there are few reference samples from relatives (Vigeland et al. 2020; Marsico et al. 2021).

There is a need to coordinate efforts in collecting samples from the migrants' relatives, especially those who have not received news of them for some time. Coordinated national and international efforts need to be reinforced and should focus on the close genetic relatives (parents, siblings, progeny) of the missing persons (Alvarez-Cubero et al. 2012).

Nonprofit, humanitarian proposals to generate and maintain DNA databases to compare data between unidentified remains and all those relatives who wish to donate biological material would be a highly welcome measure to help solve this problem. Bruce Budowle of the University of North Texas Center for Human Identification put forth just such a proposal, providing a procedure for receiving genetic profiles and biological samples from relatives of missing persons who might have died while crossing the Mexico-United States border (Budowle et al. 2020).

Comparing Victim's Data with Relatives' Data

From an operational standpoint, this option basically entails identifying the victim by comparing their data with that from numerous potential relatives. Certain relatives (parents and children) are ideal candidates from a genetic standpoint; however, other distant relatives (grandparents, grandchildren, nephews, nieces, cousins, etc.) could provide relevant information or at least clues for continuing the investigation.

Other powerful identification methodologies do not allow for anything other than direct comparisons; for example, a reference sample or record from the missing person, such as fingerprints and dental records. This does not imply that DNA is the best technology, just that it might be the only applicable one in these cases.

Possibility to Analyze All Kind of Biological Samples

Genetic databases for migrations and for identifying missing persons have two main indexes: (1) profiles in question (from unidentified persons or remains) and (2) reference profiles (from known persons).

Samples can be obtained from unidentified human remains, usually bones and teeth but also hair, muscles, skin, and even body fluids, depending on the preservation status of the remains. If the unidentified victim is alive, as is usually the case with children and mentally impaired adults, reference samples can be taken from epithelial buccal cells, saliva, and blood.

Samples can be taken from known sources such as potential relatives of the missing persons and usually include epithelial buccal cell, saliva, and blood. Samples can also be taken from the victim's personal items such as a hairbrush, comb, toothbrush, razor blade, wristwatch, glasses, rings, and tight-fitting clothes. In a small number of cases, reference samples can originate from hospitals (e.g., tissues previously taken for biopsies and through voluntary donation such as blood, semen, and bone marrow).

Despite the broad variety of sources for extracting DNA, the results will be identical, a unique advantage for DNA analysis and genetic databases.

Building, Updating, and Managing Genetic Databases

Once the genetic analysis has been completed, DNA databases can be generated from the sets of numbers and letters (depending on the type of analysis), short tandem repeats, sequencing, and single nucleotide polymorphisms. The information that is statistically unique to each individual can be stored in a single row composed of 60–120 characters on average. These data are exact and do not deviate, provided they are generated under the highest standard of quality. The data are easy to store, due to the small space they take up compared with data generated from images (such as anthropological, fingerprint, and weapon images). A million profiles would require between 120 and 180 megabytes of storage.

Another advantage is that the alphanumeric structure of this information facilitates rapid comparisons in the search for exact matches and other types of coincidences.

There are a number of applications that not only store data but also automatically perform searches and establish statistical calculations in the event of tentative matches, such as CODIS and other commercially available systems, most of which are compatible with CODIS, such as BONAPARTE (https://www.bonaparte-dvi.com/index.php), which was recently chosen by INTERPOL to support its efforts in the search for missing persons (INTERPOL 2021a) in coordination with its I-Familia program (INTERPOL 2021b).

Another commercially available application worth mentioning is the Mass Fatality Identification System (M-FISys, https://www.genecodesforensics.com/software/) developed by the Gene Codes Corporation (Ann Arbor, Michigan, USA), which helps identify missing children (DNA-PROKIDS) and works with other missing person databases.

When creating these databases, the highest standards of quality (certification, accreditation) need to be followed, and the databases must meet all legal and ethical requirements (data dissociation, restricted access) of the country in which it operates.

Missing Children: The DNA-PROKIDS Program

Of the almost 110,000 victims of human trafficking identified by the US State Department in 2020, a large percentage were children, defined as individuals under 18 years of age.

Nevertheless, our research group at the University of Granada (Spain) efforts are focused on the most vulnerable children: newborns (0–2 months), infants (2 months–1 year), and toddlers (1–4 years). At these ages, children are easy targets for traffickers, who exploit them for various purposes, including sex, labor, begging, and illegal adoptions. The lack of identification is compounded by the children's difficulty or inability to explain who they are or to provide clues that could help identify them or their families. Children are kidnapped, issued fake documents and new identities, and adopted by another family to start another life, while the true biological family remains powerless. As an identification tool, DNA typing could help not only to solve such crimes but also prevent them.

Human migrations facilitate these illegal activities, especially in cases of forced migration caused by conflicts. Europe has recently been affected by large numbers of migrants from Iraq and Syria, who generally attempt to enter the European Union/Schengen space through the Greek-Turkish and Italian borders (Olivieri et al. 2018; Pavlidis and Karakasi 2019).

DNA-PROKIDS is a humanitarian, nonprofit program created between 2002 and 2004 and run by the University of Granada (Spain) and is supported by the Spanish Government and the Andalusian Regional Government, with the collaboration of Zogbi (a provider of forensic equipment in Mexico), the US Returned.org Foundation, and the QSDglobal Foundation for missing persons (Spain). The first pilot program for DNA-PROKIDS was conducted in Guatemala between 2004 and 2006 (Alvarez-Cubero et al. 2012).

The DNA-PROKIDS program is open to collaboration with countries willing to combat the trafficking of children at the origin and supports interested countries in creating a database composed of two separate indexes. The first index (or Questioned Database) is constructed from DNA data from unidentified missing children (usually newborns, infants, and toddlers) who reside in orphanages under the protection of the authorities. The second index consists of DNA data from the relatives of missing children who voluntarily donate biological samples after signing an informed consent.

DNA-PROKIDS is currently under development in various countries and is in permanent collaboration with state governments, including Guatemala (Guatemala National Institute of Forensic Sciences [INACIF]), Mexico (Mexico City Attorney General's Office), Honduras (National Office for Infants, Adolescents and Family [DINAF] and Honduras Institute of Forensic Medicine), and Thailand (Royal Thai Police).

So far, more than 1800 children have been identified and returned to their families. Through DNA typing, DNA-PROKIDS can help stem the tide of illegal adoptions.

Organ Trafficking: The DNA-ProORGAN Program

Organ trafficking, considered part of human trafficking, which in turn is linked to human migration, is a difficult crime to track and prosecute but one that could be investigated with the use of forensic DNA analysis. The United Nations Office on Drugs and Crime defines organ trafficking as the "illicit removal of organs." To help combat organ trafficking, the Declaration of Istanbul on Organ Trafficking and Transplant Tourism was signed in 2008 (The Transplantation Society 2008), with an updated edition signed in 2018 (The Transplantation Society 2018).

According to the Global Observatory on Donation and Transplantation (the official WHO observatory supported by The Spanish Transplant Organization), there were 163,141 organ transplantations worldwide in 2019. Due to the restrictions imposed due to the COVID-19 pandemic, there were 79,242 transplantations in 2020. Kidneys represent approximately 65% of all transplanted organs, 5–10% of which could be illegal and thus considered organ trafficking (López-Fraga et al. 2014). Although the data are hard to track due to their criminal and secretive nature, between 5000 and 10,000 kidneys might have been transplanted illegally in 2019.

Despite the growing number of transplantations worldwide, organ trafficking persists due to the fact that the number of patients needing organs is much larger than the number of donors and potential transplants. If unavailable locally, patients and their families often attempt to obtain organs from other sources, even traveling abroad (Ambagtsheer et al. 2016).

The dramatic reduction in transplantations during the COVID-19 pandemic has two main reasons. First, hospitals rooms and intensive care units were needed for the thousands of infected patients. Second, transplantation requires varying levels of immunosuppression, and SARS-Cov-2 virus infection presents a high mortality risk for immunosuppressed patients. Despite the efforts of the global medical community, the waiting lists for transplantation have increased throughout 2020 and are still increasing. The number of cases of organ trafficking will therefore likely increase in parallel (official data still pending).

The lack of prosecution of organ trafficking is mainly due to the lack of documentation and the easy falsification of documents. Organs can be extracted from individuals who might sign informed consent forms under false pretenses, under duress, or with forged documents are, falsifying the names and signatures. Tracking the origins of organs is therefore almost impossible. Genetic analysis can help identify victims and the individuals involved, which might help prevent future occurrences (Lorente et al. 2020).

In 2016, the University of Granada started the pilot DNA-ProORGAN program in an effort to develop genetic databases that could help track the identities of organ donors and recipients. The program has been supported since 2017 by the Spanish Medical Association and, since 2019, by the Spanish National Transplant Organization. The program's efforts are now focused on kidney transplantations, given that they account for an estimated 65% of all transplants and approximately 80% of all illegal transplants. DNA from transplanted kidneys (or graft) can be obtained from

the kidney recipient's urine. There is therefore no need to perform biopsies or expensive and dangerous surgical procedures (Lorente et al. 2020).

Future Perspectives: Enlarging and Coordinating the Databases

The main challenge as of today is not a scientific one but rather a legal, social, humanitarian one: to use the currently available genetic technology as much as possible. All countries should have a national civil database for identifying missing persons, whose main purpose is to identify individuals as a way of restoring their dignity.

Enacting legislation for implementing databases to monitor such criminal activity, focused on identifying those who traffic humans and their organs, could pose difficulties for certain countries. However, databases that focus on human dignity and are based on the voluntary donation of samples by relatives of missing persons should be easier to implement. There are already numerous countries that have passed such legislations, which can act as examples for other countries considering the creation of such databases.

These databases need to contain an index with DNA profiles from all unidentified remains and another index with profiles from reference samples (relatives, personal belongings, etc.). National authorities need to pay special attention to the relatives of missing persons. Without reference samples for comparison, identification will not be possible. International organizations and nongovernmental organizations can also help, especially when working on-site.

Forensic scientists can also play an additional relevant nonscientific role by informing and convincing people and authorities as to the advantages and limitations of genetic identification.

Conclusion

The world continues to face the problems of identifying missing persons and unidentified remains, as well as the kidnapping and exploitation of children for illegal adoptions and other purposes. These problems are intimately connected to the flow of human migration, the perfect environment for human traffickers.

Human identification is a multidisciplinary task in the forensic sciences, requiring the coordination of these specialists. These specialists include forensic pathologists, anthropologists, odontologists, and fingerprint specialists, employing tools such as forensic genetics and DNA typing, a powerful, reliable, universal, and compatible scientific technology. There are commercially available kits and software that can help generate databases to aid in these investigations.

There is a pressing need to focus on identifying missing persons, creating national and regional databases, and connecting them through the established channels of international cooperation following national and international laws.

DNA typing specialists should include among their goals the raising of awareness among the public as to the benefits of this technology, advocating its creation and permanent implementation.

Although we are faced with this problem, we have the tools that could help solve a large part of this crime.

References

Alvarez-Cubero MJ, Martínez-González LJ, Álvarez JC et al (2012) Genetic identification of missing persons: DNA analysis of human remains and compromised samples. Pathobiology 79:228–238

Ambagtsheer F, de Jong J, Bramer WM, Weimar W (2016) On patients who purchase organ transplants abroad. Am J Transplant 16:2800–2815. https://doi.org/10.1111/AJT.13766

Budowle B, Bus MM, Josserand MA, Peters DL (2020) A standalone humanitarian DNA identification database system to increase identification of human remains of foreign nationals. Int J Legal Med 134:2039–2044. https://doi.org/10.1007/S00414-020-02396-9

Carneiro da Silva Junior R, Wirz LN, Solares Reyes E, Del Moral Stevenel MA (2020) Development of DNA databases in Latin America. Forensic Sci Int 316:110540

Department of State. United States of America (2021) Trafficking in Persons Report. Retrieved from https://www.state.gov/wpcontent/uploads/2021/07/TIP_Report_Final_20210701.pdf

Elwick K, Bus MM, King JL et al (2019) Utility of the Ion S5™ and MiSeq FGx™ sequencing platforms to characterize challenging human remains. Legal Med 41:101623. https://doi.org/10.1016/j.legalmed.2019.08.001

Emery MV, Bolhofner K, Winingear S et al (2020) Reconstructing full and partial STR profiles from severely burned human remains using comparative ancient and forensic DNA extraction techniques. Forensic Sci Int Genet 46. https://doi.org/10.1016/J.FSIGEN.2020.102272

Ge J, Budowle B (2021) Forensic investigation approaches of searching relatives in DNA databases. J Forensic Sci 66:430–443. https://doi.org/10.1111/1556-4029.14615

Ge J, Budowle B, Chakraborty R (2010) DNA identification by pedigree likelihood ratio accommodating population substructure and mutations. Investig Genet 1(1):1–9. https://doi.org/10.1186/2041-2223-1-8

Ge J, Eisenberg A, Budowle B (2012) Developing criteria and data to determine best options for expanding the core CODIS loci. Investig Genet 3(1):1–14. https://doi.org/10.1186/2041-2223-3-1

Guo F, Shen H, Tian H et al (2014) Development of a 24-locus multiplex system to incorporate the core loci in the Combined DNA Index System (CODIS) and the European Standard Set (ESS). Forensic Sci Int Genet 8:44–54. https://doi.org/10.1016/J.FSIGEN.2013.07.007

International Organization for Migration (2019) Glossary on migration. International Organization for Migration, Geneva

International Organization for Migration (2021) IOM SNAPSHOT. https://www.iom.int/sites/default/files/about-iom/iom_snapshot_a4_en.pdf

INTERPOL (2021a) INTERPOL unveils new global database to identify missing persons through family DNA. News. https://www.interpol.int/News-and-Events/News/2021/INTERPOL-unveils-new-global-database-to-identify-missing-persons-through-family-DNA

INTERPOL (2021b) I-Familia. In: How we work. https://www.interpol.int/en/How-we-work/Forensics/I-Familia

López-Fraga M, Domínguez-Gil B, Capron AM et al (2014) A needed convention against trafficking in human organs. Lancet 383:2187–2189. https://doi.org/10.1016/S0140-6736(14)60835-7

Lorente JA (2001) Correspondence: the AICEF/GITAD: Latin American Academy of Criminalistics and Forensic Studies. J Forensic Sci 46:196–197

Lorente JA, Saiz M, Haarkötter C et al (2020) Genetic identification against traffic in human beings. WIREs Forensic Sci 3(1):e1392. https://doi.org/10.1002/wfs2.1392

Marsico FL, Vigeland MD, Egeland T, Piñero MH (2021) Making decisions in missing person identification cases with low statistical power. Forensic Sci Int Genet 54:102519. https://doi.org/10.1016/J.FSIGEN.2021.102519

Olivieri L, Mazzarelli D, Bertoglio B et al (2018) Challenges in the identification of dead migrants in the Mediterranean: the case study of the Lampedusa shipwreck of October 3rd 2013. Forensic Sci Int 285:121–128. https://doi.org/10.1016/J.FORSCIINT.2018.01.029

Parson W, Ballard D, Budowle B et al (2016) Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Sci Int Genet 22:54–63. https://doi.org/10.1016/J.FSIGEN.2016.01.009

Pavlidis P, Karakasi MV (2019) Greek land borders and migration fatalities – humanitarian disaster described from the standpoint of Evros. Forensic Sci Int 302:109875. https://doi.org/10.1016/J.FORSCIINT.2019.109875

Pilli E, Vai S, Caruso MG et al (2018) Neither femur nor tooth: petrous bone for identifying archeological bone samples via forensic approach. Forensic Sci Int 283:144–149

The Transplantation Society (2008) The Declaration of Istambul on Organ Trafficking and Transplant Tourism. Retrieved from https://www.declarationofistanbul.org/images/documents/doi_2008_English.pdf

The Transplantation Society (2018) The Declaration of Istambul on Organ Trafficking and Transplant Tourism. Retrieved from https://www.declarationofistanbul.org/images/documents/doi_2018_English.pdf

United Nations (2000) United Nations Convention Against Transnational Organized Crime A/RES/55/25. Palermo. Retrieved from https://www.unodc.org/pdf/crime/a_res_55/res5525e.pdf

United Nations General Assembly (2015) Transforming our world: the 2030 Agenda for Sustainable Development A/RES/70/1. Retrieved from https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

United Nations Office on Drugs and Crime (2018) Human trafficking and migrant smuggling. In: Hum. Traffick. Migr. Smuggling. https://www.unodc.org/e4j/en/secondary/human-trafficking-and-migrant-smuggling.html

United Nations Office on Drugs and Crime (2020) Global report on trafficking in persons 2020. New York

United Nations. Department of Economic and Social Affairs. Population Division (2020) International Migrant Stock 2020. Retrieved from https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa_pd_2020_international_migrant_stock_documentation.pdf

Vigeland MD, Marsico FL, Herrera Piñero M, Egeland T (2020) Prioritising family members for genotyping in missing person cases: a general approach combining the statistical power of exclusion and inclusion. Forensic Sci Int Genet 49:102376. https://doi.org/10.1016/J.FSIGEN.2020.102376

Zeng X, Elwick K, Mayes C et al (2019) Assessment of impact of DNA extraction methods on analysis of human remain samples on massively parallel sequencing success. Int J Legal Med 133:51–58

Challenges in the DNA Analysis of Compromised Samples

49

Christian Haarkötter, María Saiz, M. J. Alvarez-Cubero, Juan Carlos Alvarez, and J. A. Lorente

Contents

C. Haarkötter · M. Saiz (✉) · J. C. Alvarez
Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain
e-mail: chaarkotter@ugr.es; msaiz@ugr.es; juanca@ugr.es

M. J. Alvarez-Cubero
Department of Biochemistry and Molecular Biology III, Faculty of Medicine, University of Granada, Granada, Spain

Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), Granada, Spain
e-mail: mjesusac@ugr.es

J. A. Lorente
Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, University of Granada, Granada, Spain

GENYO, Centre for Genomics and Oncological Research (Pfizer / University of Granada / Andalusian Regional Government), Granada, Spain
e-mail: jlorente@ugr.es

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_51

Abstract

DNA analysis is a vital tool in forensic sciences – particularly criminal science – since the introduction of PCR at the beginning of the 1990s. However, many challenges remain that must be solved, such as inhibition of the reaction, DNA degradation, low copy number (LCN) DNA, and DNA contamination. The aim of this chapter is to examine these issues and give the reader not a solution but options for dealing with these challenges. First, inhibitory factors and mechanisms are discussed. Next, DNA degradation and LCN DNA are introduced, followed by issues and possible sources of DNA contamination. Finally, recommendations are provided to identify, treat, or at least avoid these problems.

Keywords

Degraded DNA · Forensic DNA challenges · Low copy number DNA · DNA contamination · PCR inhibition

Introduction

Frequently, forensic genetic laboratories must work with DNA samples that are in suboptimal condition. These samples might be scarce, fragmented, or poorly preserved or might have been exposed to environmental agents. Consequently, the techniques that are used in forensic laboratories are crucial to optimize the recovery of DNA from compromised samples. Thus, the type of sample that is used in the DNA analysis is critical.

The best tissues to analyze ancient DNA are the bones and teeth (Miloš et al. 2007), because they protect DNA from degradation and biological processes due to their physical and chemical robustness.

Teeth are the hardest tissue in the human body, due to tooth enamel, and DNA is preserved by calcified tissues (Girish et al. 2010). The tooth is made up of dentin, a connective tissue that forms a large part of the structural axis of the tooth. Dentine in the dental crown is covered by enamel, an extremely mineralized, acellular, and avascular tissue of ectodermal origin. The dentin at the root of the tooth is covered by cementum, another type of calcified connective tissue. The soft tissue in the pulp consists of odontoblasts, fibroblasts, endothelial cells, peripheral nerves, undifferentiated mesenchymal cells, and other nucleated components of blood, making it rich in DNA (Muruganandhan and Sivakumar 2011). The pulp of the tooth is used to recover DNA, where it abounds and is unlikely to be contaminated by nonhuman DNA (Girish et al. 2010) (Fig. 1).

Bone tissue is made up primarily of proteins and minerals. The two most abundant proteins in this tissue are collagen and osteocalcin. Approximately 70% of the mineral component of bone comprises hydroxyapatite, which includes calcium phosphate, calcium carbonate, calcium fluoride, calcium hydroxide, and citrate.

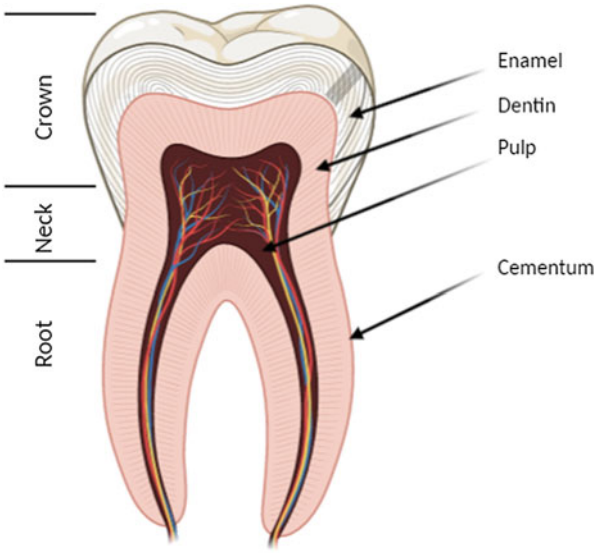


Fig. 1 Anatomy of a tooth

With regard to the structural organization of bone, the mineral component provides support for the protein element and thus physically excludes potentially harmful extracellular agents and enzymes (Collins et al. 2002). DNA has a high affinity for hydroxyapatite, and DNA degradation is linked to a loss in the crystallinity of hydroxyapatite, although it has also been associated with the loss of collagen. Bony characteristics are directly related to the survival of DNA and its protection from degradation. Bone density is one of the most important intrinsic factors in bone survival, differing significantly between men and women. In addition, there are disparities in bone density between parts of the skeleton, with the middle region of bones having higher levels.

Mechanisms that preserve DNA well in bones are not well characterized (Collins et al. 2002; Pruvost et al. 2007). However, there are several factors that affect DNA preservation, such as temperature, pH, humidity, and time.

Many studies have measured success in obtaining an STR (short tandem repeat) profile, based on the type of sample, tooth, and bone (Fig. 2), reporting that the molars, petrous bone, and femur are the best tissues (Miloš et al. 2007; Gamba et al. 2014).

Keratinized tissue, such as hair and nails, as well as horn, feathers, and scales in vertebrates, can also be used. Because these tissues originate from living cells that die during biogenesis and undergo natural desiccation, DNA might survive for longer periods. However, these cells have lost their nuclei, reducing the analysis to that of mitochondrial DNA.

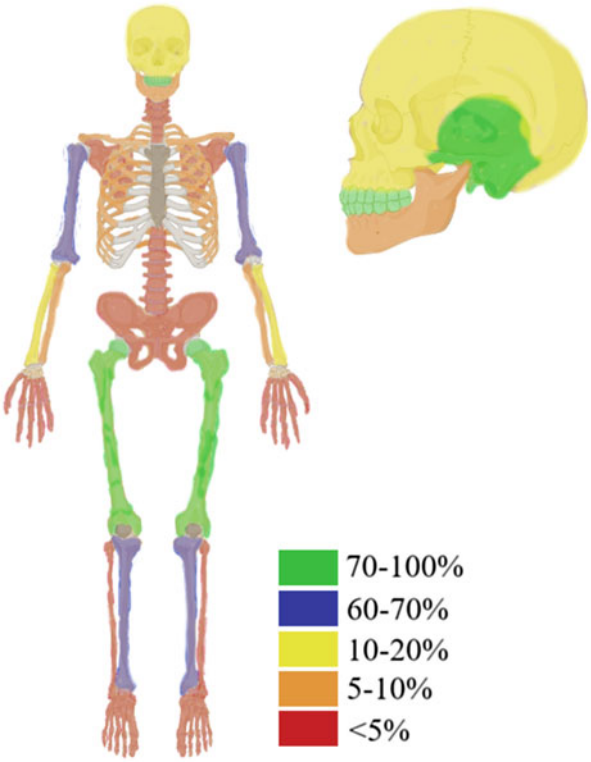


Fig. 2 Percentage of success in DNA recovery and STR typing by bone type (Miloš et al. 2007; Behar et al. 2012)

It is difficult to estimate the conservation of DNA through simple macroscopic inspection. However, there are several features of a good sample. The bones must be hard and heavy and have a compact structure. Long bones are preferable to porous bones, such as the pelvis, scapula, and parts of the skull. The bones must not have cracks or evidence of microbial attack. Especially for human identification, teeth with well-preserved roots are useful. Burnt material or material that has been affect by heat is unsuitable. With a large number of individuals, sampling the same anatomical element is recommended to avoid double-typing.

Inhibition of PCR

PCR inhibitors usually bind directly to single- or double-stranded DNA, preventing or reducing amplification. Other inhibitors block the enzyme, primarily by altering the main cofactor, Mg^{2+}. One of the chief limitations in forensic cases is the presence of inhibitors in the samples; thus, it is preferable to know how to eliminate them to ensure amplification of the sample, usually by improving DNA purification methods to remove inhibitors.

Inhibitory Mechanisms

Common samples in forensic cases, such as skeletal remains, hair, and blood, are usually exposed to harsh environmental conditions, resulting in damaged or degraded DNA and samples that contain PCR inhibitors. For example, human remains could contain PCR-inhibiting agents, such as collagen, calcium, humic acid, melanin, and hematin that reduce downstream DNA typing success. Blood samples also have harbor inhibitors, including anticoagulants, such as EDTA and heparin, and calcium, which is commonly co-extracted with DNA from skeletal remains. Most DNA extraction methods are effective in removing high amounts of inhibitors from the blood, hair, and bone tissue in forensic samples (Zeng et al. 2019).

However, some inhibitors remain in the PCR process and could interfere primarily, not only in DNA amplification processes binding directly to single- or double-stranded DNA or blocking the enzyme but also with the library preparation in MPS (massive parallel sequencing) analysis or qPCR by quenching the fluorescence of the fluorophore. Inhibitors could have a negative effect on 5′-3′ exonuclease activity and thus lead to inhibition, because the DNA polymerase will be unable to hydrolyze the probe, resulting in underestimation of DNA levels. Humic acid and several molecules in blood, such as hemoglobin and hematin, quench the fluorescence of DNA-binding dyes. Consequently, these inhibitors could decrease the amplification and corrupt the analytical data, specifically for samples with low amounts of template and high amounts of background material, which are common in forensic scenes. There are several strategies for avoiding or reducing the effects of PCR inhibitors, such as purification, dilution of DNA extracts, and the use of inhibitor-tolerant DNA polymerases (Sidstedt et al. 2020).

Other inhibitors, such as calcium, suppress STR typing as their concentration increases, whereas humic acid has no inhibitory effect on STR profiling but intercalates in the DNA template, limiting its availability for amplification. Collagen strongly inhibits STR typing, effecting allele dropout with rising concentration. Hematin does not seem to inhibit STR typing, but, as it is a metal-chelating agent that binds to magnesium, it can cause complete allele dropout in qPCR (Kieser et al. 2020).

Several reagents for extraction, such as phenols, that are introduced during DNA purification can act as inhibitors by denaturing the polymerase (Sidstedt et al. 2020). Thus, several kits for extraction or amplification in forensic cases are designed to reduce inhibitors, such as the Investigator 24plex, with a higher tolerance to common PCR inhibitors, including humic acid and tannic acid (Lin et al. 2017). Many efforts have been made to improve PCR kits, such as a modified version of AmpFlSTR SGM Plus, replacing AmpliTaq Gold DNA polymerase, with a customized blend of two alternative polymerases: ExTaq Hot Start and PicoMaxx High Fidelity. These changes have improved the results for allele calls and stutter sizes, primarily by increasing the resistance to PCR inhibitors that plague forensic identification (Hedman et al. 2011). Certain substances can also help to avoid the effect of PCR inhibition; bovine serum albumin (BSA) may improve between five and ten times

inhibitor tolerance (Sidstedt et al. 2019), which has led to the addition of BSA in several commercial kits available nowadays.

As discussed, there are several types of inhibitor, depending on the mechanism, for which there are various strategies, based on whether DNA polymerization or fluorescence detection is primarily affected (Sidstedt et al. 2020).

Inhibition Factors

(i) *Soil*: Many chemical reactions occur naturally in soil, which contains inorganic and organic compounds that can affect DNA. Many of these substances cannot be removed during DNA purification and can influence PCR (Alaeddini 2012). Depending on the type of soil, there will be varying proportions of fulvic acid or humic acid, which will have disparate effects on the quality of the DNA. Humic acid predominantly affects the initial PCR cycles, during which DNA is mainly genomic, but not the final cycles, when short amplicons abound (Sidstedt et al. 2017). It is difficult to eliminate with extraction protocols (Opel et al. 2010). Besides, soil pH and humidity levels have a negative effect on ancient DNA conservation, and they may ease the introduction of PCR inhibitors in the bone tissue (Dabney et al. 2013).

(ii) *Blood*: Blood is one of the main samples that are collected for human identification, but many of its components inhibit PCR (Al-Soud et al. 2000). Hemoglobin and hematin in blood samples inhibit amplification and thus impede correct genetic identifications. The inhibitory power of each component has been studied, and it has been concluded that hemoglobin is a more powerful inhibitor, requiring lower amounts to completely inhibit amplification (Ambers et al. 2018); however, unlike hemoglobin, hematin directly alters the activity of DNA polymerase (Opel et al. 2010).

(iii) *Bone*: Inhibitory substances in bone become apparent when amplification is performed with these samples (Loreille et al. 2007). The extraction of DNA from bone is complex and expensive; thus, it is essential to know the compounds with which PCR competes to avoid them (Pajnič 2016). Over 99% of calcium resides in bones and teeth (Beto 2015). But, calcium reduces the efficiency of DNA amplification (Opel et al. 2010), by interfering with the interaction between polymerase and magnesium ions in a PCR reaction (Bickley et al.1996).

(iv) *Saliva*: There are minor inhibitors in saliva samples. Amylase is the most abundant protein in saliva, and its function is to convert complex non-soluble polysaccharides into smaller soluble units (Carpenter 2013). Thus, it could affect amplification, although studies have not confirmed it.

During sample processing or DNA extraction, inhibitors can be added, including powder from gloves; salts, such as potassium and sodium chloride; detergents; and other organic molecules, including ethylenediaminetetraacetic acid (EDTA), sarkosyl, ethanol, phenol, and isopropylic alcohol. These substances, in addition to

being necessary for cell lysis and nucleic acid isolation, can inhibit PCR inhibition at certain concentrations. Ionic detergents, such as sarkosyl, SDS, and sodium deoxycholate, have high inhibitory activity, whereas nonionic detergents, including Triton X-100 and P-40, can suppress PCR reactions at high concentrations. EDTA is a common component in elution buffers and purification kits, but its chelating activity can exhaust magnesium ions and thus inhibit PCR. Even contact between PCR reagents and UV-irradiated plastic tubes and the components of cotton swabs can affect the sensitivity of PCR (Shrader et al. 2012).

For these issues, several *purification kits* are available. The DNeasy PowerClean Pro Cleanup Kit (QIAGEN, Hilden, Deutschland) combines DNA purification reagents and a silica membrane spin column: DNA is captured and purified by being washed and eluted from the spin columns, depending on the percentage recovery of inhibitors. Other kits, such as the RNA/DNA/Protein Purification Plus Kit, which uses silicon carbide columns (Norgen Biotek, Ontario, Canada), can perform both DNA extraction and purification.

Organic solvents, such as dimethyl sulfoxide (DMSO) and formamide, have been used as additives to increase the efficiency of DNA amplification, whereas bovine serum albumin has been added during PCR amplification of samples that might contain inhibitors. Consequently, BSA enhances PCR amplification yields when combined with organic solvents, improving their effects with its own effects during the initial PCR cycles (Farell and Alexandre 2012). These *PCR additives* have beneficial effects on amplification, but it is impossible to establish which agents are useful in each particular context (Gene Link 2014):

(a) *Betaine* (0.1–3.5 M), *DMSO* (2–8%), and *formamide betaine* (1–5%). DMSO and formamide reduce the secondary structure of GC-rich templates and aid in their amplification. The combination of DMSO and betaine is superior to formamide use alone.
(b) *TMAC* (tetramethylammonium chloride, 15–100 mM). TMAC increases the hybridization specificity and melting temperature and should be used in combination with DMSO and betaine.
(c) *BSA and Gelatin BSA*. These compounds stabilize Taq polymerase and have been especially effective in amplifying ancient DNA, which contains PCR inhibitors, such as melanin, and reducing enzyme loss by nonselective adsorption to tube walls.

Degraded DNA

DNA is constantly degraded in living beings by hydrolysis, oxidation, and methylation, which fracture and alter its primary structure, changing the sequence and thus its inherent information (Lindahl 1993). The cell has various DNA repair pathways (Iyama and Wilson 2013), which cease when death occurs, allowing DNA to continue to be attacked by physical, chemical, and biological elements; however, the ensuing damage is permanent. DNA degradation begins with two mechanisms:

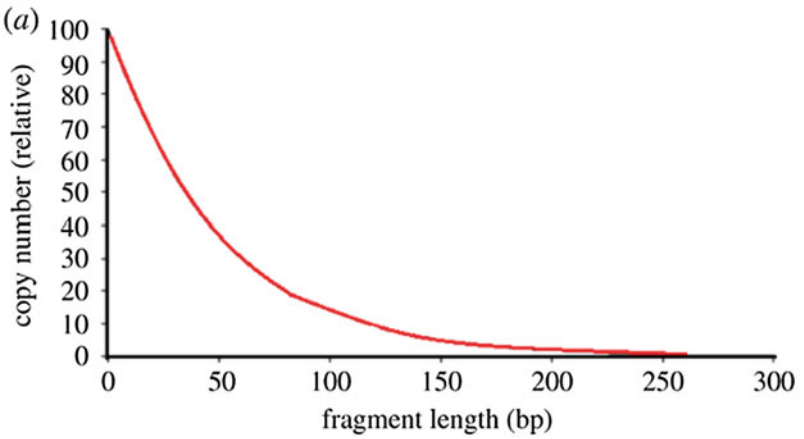


Fig. 3 Exponential decay of DNA: copy number versus fragment length (Allentoft et al. 2012)

nucleases cleave DNA into fragments, after which DNA is digested by microorganisms, effecting random DNA fragmentation, reflected by a negative exponential correlation (Fig. 3) between the number of molecules and DNA length at a rate of $5.50 \cdot 10^{-6}$ nucleotides per year (Allentoft et al. 2012).

DNA Degradation Mechanisms

DNA damage can be classified into two groups: fragmentation and modification (Pääbo et al. 2004), which occur through the following mechanism simultaneously and randomly:

(A) *DNA bond cleavage*. Due to the degradative activities of microorganisms and *postmortem* nucleases or other chemical components. Nucleases are phosphodiesterases that cleave one of the two (3′ or 5′) phosphorous-oxygen bridges in a nucleic acid polymer by acid-base catalysis (Yang 2011). DNase I is the most common enzyme that is used in artificial DNA degradation assessments in the forensic literature (Timken et al. 2005).

(B) *Oxidative damage*. Free radicals, such as $-O_2^{2-}$ and $-OH^-$, are produced by metabolic reactions and external sources, including ionizing radiation and decomposing aerobic bacteria. Free radicals react with DNA in a process that is called oxidative DNA damage and can induce DNA modifications, such as base and sugar lesions, strand breaks, and cross-links by reduction-oxidation reactions (Dizdaroglu 2002). Guanine is more susceptible to this type of damage due to its double bonds (Cadet et al. 2017).

(C) *Cross-links*. DNA cross-linking is the formation of a covalent bond between two DNA nucleotides. Cross-links appear with high frequency and can block *Taq*

polymerase during DNA amplification, which can be seen by electronic microscopy (Pääbo 1989).

(D) *Hydrolysis-induced rupture*. Hydrolysis affects DNA through the 2′-deoxyribose backbone and deamination. Deamination can occur in cytosine, adenine, and guanine, causing the loss of the exocyclic amino group, which transforms them into uracil, hypoxanthine, and xanthine, respectively (Marrone and Ballantyne 2010). These transformations lead to changes in sequence.

(E) *Maillard reactions*. Described by Maillard in 1912, these reactions are well-known in food chemistry, because they elicit the brown color in beers, meats, and baked goods. The exact mechanism remains unknown, but it has been reported that DNA is broken in glucose and amino acid mixtures (Hiramoto et al. 1993), conditions that are common during decomposition.

DNA Degradation Factors

In the environment, there are several factors that can affect DNA preservation (Dean and Ballard 2001):

1. *UV Radiation*. Solar radiation is a well-known mutagenic agent, and UV light is even used in forensic laboratories for exogenous DNA decontamination. Ultraviolet radiation comprises electromagnetic waves from 10 to 400 nm: long-wave (UVA, 400–315 nm), medium-wave (UVB, 315–280 nm), and short-wave (UVC, 280–100 nm). UV light affects DNA in various manners: UV induces damage (oxidative and hydrolysis damage due to free radicals that are created by UV radiation), pyrimidine photoproducts (such as cyclobutane pyrimidine dimers and 6–4 pyrimidine photoproducts, which affect DNA strand structure); purine photoproducts, and double-strand breaks (Rastogi et al. 2010).
2. *Temperature*. Likely it is the most determining factor in DNA preservation. Low temperatures preserve DNA, because they slow degradative chemical reactions. High temperatures accelerate such reactions, but very high temperatures lead to tissue desiccation and can protect DNA from degradative processes. Ideal DNA preservation occurs at constant temperature, as reported for bones that have been found in caves (Collins et al. 2002).
3. *Humidity.* The solvating action of water allows for the presence of organic substances, which have DNA polymerase-inhibiting activity, allowing oxidative and hydrolytic degradation. Water can also help organic components from the environment penetrate into the sample if its surface is porous, as with bones or teeth. DNA levels fall slightly in a more humid environment, but real effects are detected on DNA quality (50% less reproducible alleles (Burger et al. 1999)).
4. *pH*. An acidic pH affects the oxidation state of the phosphorus content in DNA phosphate groups, and low pH levels fragment DNA into its elemental units (Young et al. 2014). Alkaline pH levels affect the hydrogen bridges between strands, and very high pH levels separate double-stranded DNA by deprotonation

(Ageno et al. 1969). Neutral or slightly high pH levels – near physiological pH – help preserve DNA.

5. *Soil components*. The minerals in soil have various effects on DNA preservation. For example, sand – composed primarily of quartz – has adsorbing capacity and sequestering DNA and protects it from the activity of degrading agents (Romanowski et al. 1991). The same phenomenon occurs in clays; thus, analyzing the soil itself to obtain DNA is promising (Alvarez et al. 1998). Humic substances, a heterogeneous group of substances that are derived from plants and animals that remain after decomposition, have also been described as PCR inhibitors. However, humic substances bind to DNA (Zipper et al. 2003), protecting it from other substances (Crecchio and Stotzky 1998), so its possible degradation mechanism is not clear yet.
6. *Microorganisms*. Nucleated human cells are a material of interest in forensic genetics, but they contain many nutrients that attract the decomposing activity of microorganisms. Hemoglobin is a source of iron and protein for *Plasmodium*, the calcium phosphate in bones attracts *Bacillus megaterium*, and collagen attracts collagenolytic bacteria; nutrients in semen or vaginal fluids induce rapid bacterial degradation by *Escherichia coli* and *Staphylococcus epidermidis* (Dash and Das 2018). Degrading microorganisms liberate various components, such as nucleases, that mutate DNA.

To mitigate the analytical problems that are associated to degraded DNA, several amplification strategies have been developed:

(a) *miniSTR* analysis is an amplification strategy that uses a low number of short tandem repeats (STRs) and redesigned primers so that they bind closer to the STR regions in DNA, and it has been used successfully with degraded DNA (Wiegand and Kleiber 2001; Coble and Butler 2005). Several commercial miniSTRs kits have been developed, such as MiniFiler™ (Thermo Fisher), a 5-dye chemistry kit with FGA, CSF1PO, D18, D16, D21, D2, amelogenin, D7, and D13 loci, generating 33–200-bp primers. However, miniSTRs have a major disadvantage: few loci can be multiplexed in the same assay (Butler 2011).
(b) *Mitochondrial DNA* can also be used to analyze degraded DNA. Compared with a single molecule of nuclear DNA in the cell nucleus, there are up to 100,000 copies of mtDNA in a single cell (Butler 2011). In cases in which STR analysis has failed, mtDNA sequencing has been successful (Holland et al. 1993). However, there are several disadvantages: heteroplasmy (different mtDNA sequences in the same individual), maternal inheritance, and difficult statistical interpretation.
(c) *Next-generation sequencing (NGS)* and massive parallel sequencing (MPS) are promising tools for analyzing degraded DNA, sequencing large regions of DNA (or even the whole genome) to search for single-nucleotide polymorphisms (SNPs) for human identification and ancestral and phenotypic assessments (Butler 2011). Groups of SNP markers, known as microhaplotypes [40], have been proposed for degraded DNA, obtaining profiles even with highly degraded

samples (Turchi et al. 2019). There are two NGS platforms that have been validated for forensic analysis: the Ion S5 System (Thermo Fisher), with a whole-mtDNA genome and control region panel; an STR panel with GlobalFiler markers, 48 ancestry SNPs, and 54 identification SNPs; and ForenSeq™ (Verogen, San Diego, CA, USA), with 27 STR markers, 24 Y-STRs, 7 X-STRs, 94 identity SNPs, 22 phenotypic SNPs, and 56 ancestry SNPs.

Low Copy Number DNA

DNA input that is less than 200 pg (Budowle et al. 2009) or even 100 pg (Gill et al. 2000), translating into 15 or 30 diploid cells, respectively, is known as low copy number DNA, low DNA testing, low-template DNA, or LCN DNA.

There are at least two scenarios in which LCN DNA typing is essential: touch DNA and challenging samples. *Touch DNA* is based on the analysis of swabs of various objects or surfaces, looking for skin cells to demonstrate that a person touched an object or a crime scene surface (Oorschot and Jones 1997). In addition to individuals having different tendencies to deposit DNA when they touch an object, several people could have been in contact with crime scene items and surfaces but not involved with the crime, which is why mixtures can be expected in touch DNA analysis. Further, passive transfer of DNA of a person who has touched a suspect's hands is possible (Lowe et al. 2002). Recently, several touch DNA studies have been performed, based on the increased sensitivity of commercial kits, which have generated similar conclusions: touch DNA mixtures are difficult to interpret, and passive transfer is possible (Cale et al. 2016).

Conversely, LCN DNA typing can be used for *victim identification* when conventional DNA typing has failed. LCN DNA STR typing was successfully applied to four sets of skeletal remains from World War I, World War II, and the Vietnam War (Irwin et al. 2007). Despite its limitations, an optimal DNA extraction protocol and miniSTR technology are promising tools when analyzing challenging samples (Mameli et al. 2014).

Problems with LCN DNA Analysis

DNA analysis and interpretation have several well-known issues, which are amplified when the DNA input is in the recommended range. There are four main problems in LCN DNA analysis: stochastic effects, contamination, replicate analyses, and controls (Budowle et al. 2009).

(a) *Stochastic effects*. Stochastic variation occurs when two assays randomly produce different alleles in the same locus (Butler and Carolyn 2010) (Fig. 4).
   1. *Stochastic detection threshold*. This threshold is an arbitrary value that is generated during laboratory validation assays; above this value, single

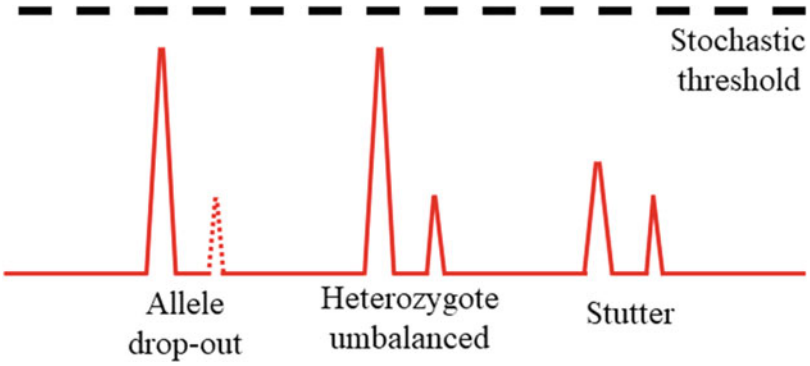


Fig. 4 Stochastic effects in LCN DNA typing

alleles are assumed to be homozygous (Butler 2014). Because LCN DNA analysis usually proceeds under this threshold, it cannot be applied.

2. *Allele dropout*. The preferential amplification of only 1 allele due to stochastic variations can result in fake homozygotes (Butler 2014).
3. *Heterozygote peak imbalance*. Heterozygote balance is obtained by the arithmetic division of high- and low-peak areas, theoretically resulting in a value of 1; however, generally, the value ranges from 0.6 to 1.67, and LCN DNA has a greater tendency toward heterozygote imbalance (Kloosterman and Kersbergen 2003).
4. *Stutter*. Stutters are allelic products with 1 less unit than the associated allele, and in LCN DNA analysis, they can have even larger areas than allele peaks (Gill et al. 2000).

(b) *Contamination*. Contamination during LCN DNA analysis comes from reagents and laboratory consumables, laboratory personnel, cross-contamination between samples, and evidence collection (Budowle et al. 2009). It has to be noted that at equal levels of contamination, this would be more noticeable in low copy number samples than in reference samples due to the great concentration disproportion between the contamination and the reference sample and it would not be detectable in most occasions.

(c) *Replicate analyses*. Because random allele variations might be detected in the same locus, various aliquots of LCN DNA extracts should be amplified and analyzed to determine a consensus profile (Taberlet et al. 1996).

(d) *Appropriate controls*. Given that LCN DNA analysis is (or should be) sufficiently sensitive to detect a single DNA molecule and because the purpose of a negative control is to detect gross laboratory contamination, controls must be redefined to detect laboratory contamination in the detected alleles (Gill et al. 2000).

To mitigate the analytical problems that are associated with LCN DNA, several procedural strategies have been developed:

(a) *DNA extraction protocol* is crucial. Its efficiency will determine the amount of DNA that is obtained and can even improve situations with low copy number DNA. Various extraction protocols have been compared in LCN DNA extraction, of which multi-silica-based DNA extraction protocols succeed, whereas spin columns fail (Mameli et al. 2014).
(b) *DNA amplification* can determine success in LCN DNA analysis. Given the small amount of DNA input, miniSTR primers have been successfully applied (Mameli et al. 2014). Notably, touch DNA profiles are more likely to be obtained with Identifiler® Plus due the use of Hot Start DNA polymerase, pre-PCR hold at 95 °C, or buffer components (Martin et al. 2018). But, the classical solution to LCN DNA analysis is *increasing the cycle number to 34* cycles (Gill et al. 2000): the peak area of stutters and the imbalance in heterozygotes increase compared with amplification using 28 cycles with less than 100 pg of DNA. However, it has been suggested that a combination of cleanup of PCR products, concentration, increased sample loading, and increased injection variables can achieve the same or better results with products from 28 cycles of amplification (Foster et al. 2008). In addition, *mtDNA* analysis and *NGS* platforms are suitable alternatives if STR typing is unsuccessful; both strategies analyze SNPs and take the advantage of having smaller amplicon size.
(c) *Post-PCR purification* in LCN DNA increases the sensitivity of the PCR products, for which silica membrane-based columns are the best option versus filtration and enzyme-mediated hydrolysis. Further, electropherogram artifacts were similar, and no contamination was observed (Smith and Ballantyne 2007).

DNA Contamination

DNA contamination implies the accidental transference of DNA to a sample (Butler 2005). DNA contamination with exogenous DNA is complicated and can be detected when working with challenging samples. The analysis of forensic DNA contamination is a tremendous challenge, for which there remain two main issues (Yang et al. 2019). The first is a minor contributor that cannot be assessed due to stochastic effects and cannot be interpreted. The second is contamination by more than three profiles, which is difficult to distinguish between all contributors. Sources of DNA contamination will vary, depending on the sample and analysis: both endogenous and exogenous DNAs can be co-extracted or coamplified, or only the contaminating DNA can even be amplified and detected. The high sensitivity of PCR and its ability to amplify LCN DNA can generate problems in the management of challenging samples, necessitating extraordinary measures and validation protocols in the laboratory to avoid sample contamination.

DNA Contamination Sources

It is important to determine the circumstances when contamination can occur to prevent it or, at least, be alert. Samples can be compromised by contamination through three periods by direct deposition or secondary transfer (Lee et al. 1998):

1. *Prior or during sample collection*. Compromised samples can be found anywhere, and there will be many types of biological evidence on surfaces and common items that can be a source of DNA contamination, known as background or environmental DNA (Fonneløpa et al. 2016). Further, inattention during the collection of the sample by personnel can lead to contamination due to passive transfer from the investigator.
2. *During sample preparation*. Because endogenous DNA of samples is degraded and at low levels, it is important to clean the sample as much as possible so that exogenous DNA contamination is avoided, to prevent mixture and to preferentially amplify exogenous DNA. The first DNA identification study of skeletal remains pointed out the need to distinguish between exogenous and endogenous DNA when analyzing an ancient DNA profile (Hagelberg et al. 1991). There are three main techniques to prepare samples prior to DNA extraction: surface abrasion, chemical wash, and ultraviolet light. *Surface abrasion* consists of applying a *Dremel*-type drill to the bone surface to eliminate most exogenous material (Carracedo 2005). *Chemical wash* of the bone surface and submersion in diluted bleach has been a controversial topic, because it can penetrate bone pores, affecting endogenous DNA; however, there are contradictory conclusions on this issue (Koehn et al. 2020). Finally, *ultraviolet light irradiation* (Kaestle and Horsburgh 2002) is also used due its ionizing effect, which fragments exogenous DNA, but it is not useful with irregular surface samples, and if the exposure time is exceeded, endogenous DNA can also be degraded.
3. *Post-PCR products*. Separation of pre-PCR and post-PCR areas is essential to avoid contamination with post-PCR products. Further, laboratory coats can be contaminated easily with PCR products, for which special and exclusive pre-PCR protective equipment for personnel must be used when working with compromised samples, including whole suits, double gloves, shoe covers, and surgical masks. In addition, the inclusion of negative controls in the analytical process is necessary. *Good laboratory practices* and adequate cleaning of laboratory equipment and tools are mandatory to avoid environmental DNA and cross-contamination in samples (Scientific Working Group on DNA Analysis Methods 2017).

It has been reported that carry-over contamination can be regulated by substituting deoxyuridines (dUTP) for thymidines (dTTP) or by introducing *uracil* during primers synthesis and by treating starting reactions with UDG (uracil DNA glycosylase) followed by its thermal inactivation. This enzyme cleaves uracil of uracil-containing DNA with no effects on thymine containing DNA. With these

measures PCR carry-over contamination can be treated if contaminants contain uracil in place of thymine (Longo et al. 1990).

Next-generation sequencing (NGS) technologies are powerful tools for validating results in ancient DNA studies. Ion semiconductor sequencing is a more precise and lower-cost tool versus molecular cloning for distinguishing between endogenous and exogenous DNA in ancient bones. The Ion Torrent™ PGM® has been used successfully to verify results obtained in Sanger sequencing of human chalcolithic teeth with a coverage of more than 1000 copies of mtDNA for the amplified fragments. Furthermore, NGS sequencing arises information about the damage of DNA (Palencia-Madrid and de Pancorbo 2015).

Laboratory Recommendations for the Analysis of Compromised Samples

There are several *laboratory criteria* (Poinar 2003; Pääbo et al. 2004) that have been established for ancient DNA laboratory work, including the analysis of compromised DNA samples:

1. *Working in a physically isolated area*. To avoid contamination, it is important to work in a separate laboratory that is dedicated exclusively to degraded samples and, if possible, another area where DNA work is not performed. DNA extraction and amplification should be performed in this area, and the following steps should be implemented in daily laboratory tasks. Personnel must wear laboratory coats, face masks, and gloves; equipment should be washed with bleach and irradiated with UV light; negative test should be included with every assay; duplicate PCR assays must be run; and all results should be compared against a staff and investigator database (Gill 2001).
2. *PCR amplification control*. It is convenient to make periodic checks of amplification with blank samples that contain only water and the necessary reagents for PCR, and every analysis should include a negative control so that contamination can be detected.
3. *Molecular behavior*. The obtained results should be proportional to DNA degradation index: an entire DNA sequence from ancient DNA samples should not be expected; such a case would likely be contamination rather than a successful analysis.
4. *Quantification*. DNA copy number in samples must be assessed by real-time PCR or competitive PCR; thus, if the molecule number is under 1000, contamination might be impossible to exclude.
5. *Reproducibility*. The obtained results should be repeated with the same DNA extracts or other DNA extracts from the same specimen.
6. *Clone*. PCR product must be verified by sequencing and cloning of amplification products (at least ten) to assess if contamination has occurred and to what extent.
7. *Replication*. Extraction, amplification, and sequencing of separate samples from the same individual by independent laboratories should achieve the same results.

8. *Biochemical preservation*. Ancillary information on DNA conservation can be obtained by measuring the quantity, composition, and degradation of amino acids or other residues. If a forensic analysis is performed, other markers, such as oxidation of fats, should be revised.
9. *Associated remains*. When working with human samples, if animal remains are found to be associated to them, the DNA should have the same degree of conservation in both types of samples; thus, PCR amplification of animal remains is advised.
10. *Phylogenetic sense*. The obtained sequences should be interpreted phylogenetically with other known haplotypes to confirm the results, especially when mitochondrial pseudogenes are detected (noncoding mitochondrial DNA contained in nuclear DNA).

Conclusions

Forensic DNA analysis developed rapidly in the past several decades; however, there are still unsolved issues when compromised samples are involved (see Table 1).

PCR inhibition is caused by two mechanisms: the inhibitor binds to DNA to prevent the reaction, or it inactivates the DNA polymerase. Inhibition must be considered when working with blood or bones or if samples have been exposed to such components as soil humic and fulvic acids. An adequate DNA extraction protocol, purification, or certain DNA polymerases are recommended for these issues, and the use of purification kits and PCR additives is recommended.

Table 1 Problems with DNA analysis and possible solutions and recommendations

Problem	Possible solutions
Challenging samples	Adequate sample selection if possible Laboratory requirements
PCR inhibition	Inhibitor-tolerant DNA polymerases Developed PCR kits Purification kits PCR additives
Degraded DNA	miniSTRs mtDNA analysis NGS
Low copy number DNA	DNA extraction protocol DNA amplification Post-PCR purification mtDNA analysis NGS
DNA contamination	Good laboratory practices Chemical wash Ultraviolet light irradiation NGS

DNA degradation is caused by several mechanisms: bond cleavage, oxidative damage, cross-links, hydrolytic rupture, and Maillard reactions – all of which are mediated by the activity of UV radiation, certain temperatures, humidity, pH ranges, soil components, and microorganism activity. Mini-STRs, mtDNA analysis, and NGS are possible strategies to be considered.

Low copy number DNA is a problem that is related to DNA degradation, for which several applications have been recently explored, such as touch DNA and victim identification. Nevertheless, stochastic effects, contamination, analysis replication, and the use of appropriate controls are problems that can arise when working with LCN DNA. Specific DNA extraction and amplification protocols and post-PCR purification are recommended.

Finally, DNA contamination can have different sources: prior to sample collection, during sample handling, or after PCR is performed. For avoiding sample contamination, and as a general guideline for all of these challenges, several laboratory criteria have been created for ancient DNA work: working in a physically isolated area, PCR amplification control, molecular behavior, quantification, reproducibility, cloning, replication, biochemical preservation, associated remains, and phylogenetic sense.

Forensic DNA analysis has advance quickly in quite short time, but there are many challenges that remain. The aim of this chapter was to introduce possible solutions that are being applied by forensic scientists. However, more research is needed for addressing the problems that are posed by an even more complex society.

References

Ageno M, Dore E, Frontali C (1969) The alkaline denaturation of DNA. Biophys J 9:1281–1311

Al-Soud WA, Jönsson LJ, Rådstrom P (2000) Identification and characterization of immunoglobulin G in blood as a major inhibitor of diagnostic PCR. J Clin Microbiol 38:345–350. https://doi.org/10.1128/JCM.38.1.345-350.2000

Alaeddini R (2012) Forensic implications of PCR inhibition – A review. Forensic Sci Int Genet 6:297–305. https://doi.org/10.1016/j.fsigen.2011.08.006

Allentoft ME, Collins M, Harker D et al (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dates fossils. Proc R Soc B Biol Sci 279:4724–4733

Alvarez AJ, Khanna M, Toranzos A, Stotzky G (1998) Amplification of DNA bound on clay minerals. Mol Ecol 7:775–778

Ambers A, Wiley R, Novroski N, Budowle B (2018) Direct PCR amplification of DNA from human bloodstains, saliva, and touch samples collected with microFLOQ® swabs. Forensic Sci Int Genet 32:80–87. https://doi.org/10.1016/J.FSIGEN.2017.10.010

Behar DM, van Oven M, Rosset S et al (2012) A "copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90:675–684

Beto JA (2015) The role of calcium in human aging. Clin Nutr Res 4:1. https://doi.org/10.7762/CNR.2015.4.1.1

Budowle B, Eisenberg AJ, van Daal A (2009) Validity of low copy number typing and applications to forensic science. Croat Med J 50:207–217

Burger J, Hummel S, Herrmann B, Henke W (1999) DNA preservation: a microsatellite-DNA study on ancient skeletal remains. Electrophoresis 20:1722–1728

Butler JM (2005) Forensic DNA typing, 2nd edn. Elsevier, Burlington

Butler JM (2011) Advanced topics in forensic DNA typing: methodology. Academic, San Diego
Butler JM (2014) Advanced topics in forensic DNA typing: interpretation. Academic, San Diego
Butler JM, Carolyn RH (2010) Scientific issues with analysis of low amounts of DNA. In: Promega Profiles DNA. https://www.promega.es/resources/profiles-in-dna/2010/scientific-issues-with-analysis-of-low-amounts-of-dna/. Accessed 15 May 2020
Cadet J, Davies KJA, Medeiros MH et al (2017) Formation and repair of oxidatively generated damage in cellular DNA. Free Radic Biol Med 107:13–34
Cale CM, Earll ME, Latham KE, Bush GL (2016) Could secondary DNA transfer falsely place someone at the scene of a crime? J Forensic Sci 61:196–203
Carpenter GH (2013) The secretion, components, and properties of Saliva 4:267–276. https://doi.org/10.1146/annurev-food-030212-182700
Carracedo Á (ed) (2005) Forensic DNA typing. Humana Press, Totawa
Coble MD, Butler JM (2005) Characterization of new MiniSTR loci to aid analysis of degraded DNA. J Forensic Sci 50:43
Collins MJ, Nielsen-Marsh CM, Hiller J et al (2002) The survival of organic matter in bone: a review. Archaeometry 44:383–394
Crecchio C, Stotzky G (1998) Binding of DNA on humic acids: effect on transformation of Bacillus subtilis and resistance to DNase. Soil Biol Biochem 30:1061–1067
Dabney J, Meyer M, Pääbo S (2013) Ancient DNA damage. Cold Spring Harb Perspect Biol 5: a012567
Dash HR, Das S (2018) Microbial degradation of forensic samples of biological origin: potential threat to human DNA typing. Mol Biotechnol 60:141–153
Dean MD, Ballard WO (2001) Factors affecting mitochondrial DNA quality from museum preserved Drosophila simulans. Entomol Exp Appl 98:279–283
Dizdaroglu M (2002) Introduction to serial reviews on oxidative DNA damage and repair. Free Radic Biol Med 32:667
Farell EM, Alexandre G (2012) Bovine serum albumin further enhances the effect of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. BMC Res Notes 5:257
Fonneløpa AE, Johannessen H, Egeland T, Gill P (2016) Contamination during criminal investigation: detecting police contamination and secondary DNA transfer from evidence bags. Forensic Sci Int Genet 23:121–129
Foster L, Thomson J, Kutranov S (2008) Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle "low copy number" (LCN) method for analysis of trace forensic DNA samples. Forensic Sci Int Genet 2:318–328
Gamba C, Jones ER, Teasdale MD et al (2014) Genome flux and stasis in a five millenium transect of European prehistory. Nat Commun 21:5257
Gene Link (2014) PCR additives & enhancers. In: Certif. Anal. Prod. Man. https://www.genelink.com/Literature/ps/M40-3021-PCR_Additives_Ver5.1.pdf. Accessed 25 Feb 2021
Gill P, Whitaker J, Flaxman C et al (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 112:17–40
Gill P (2001) Application of low copy number DNA profiling. Croat Med J 42:229–232
Girish K, Rahman FS, Tippu SR (2010) Dental DNA fingerprinting in identification of human remains. J Forensic Dent Sci 2:63–68
Hagelberg E, Gray IC, Jeffreys AJ (1991) Identification of the skeletal remains of a murder victim by DNA analysis. Nature 352:427
Hedman J, Duvfa C, Norén L et al (2011) Applying a PCR inhibitor tolerant DNA polymerase blend in forensic DNA profiling. Forensic Sci Int Genet Suppl Ser 3:E349–E350
Hiramoto K, Kato T, Kikugawa K (1993) Generation of DNA-breaking activity in the Maillard reaction of glucose-amino acid mixtures in a solid system. Mutat Res 285:191–198
Holland MM, Fisher DLDL, Mitchel LG et al (1993) Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war. J Forensic Sci JFSCA 38:542–553

Irwin JA, Leney MD, Loreille O et al (2007) Application of low copy number STR typing to the identification of aged, degraded skeletal remains. J Forensic Sci 52:1322–1327
Iyama T, Wilson DM (2013) DNA repair mechanisms in dividing and non-dividing cells. DNA Repair 12:620–636
Kaestle FA, Horsburgh KA (2002) Ancient DNA in anthropology: methods, applications, and ethics. Am J Phys Anthropol 35:92–130
Kieser RE, Buś MM, King JL et al (2020) Reverse Complement PCR: a novel one-step PCR system for typing highly degraded DNA for human identification. Forensic Sci Int Genet 44:102201
Kloosterman AD, Kersbergen P (2003) Efficacy and limits of genotyping low copy number (LCN) DNA samples by multiplex PCR of STR loci. J Soc Biol 197:351–359
Koehn K, Buettner A, Lindner I (2020) Effect of sodium hypochlorite decontamination on the DNA recovery from human teeth. Int J Legal Med 134:93–99
Lee H, Ladd C, Scherczinger CA, Bourke MT (1998) Forensic applications of DNA typing: part 2 collection and preservation of DNA evidence. Am J Forensic Med Pathol 19:10–18
Lin S, Li C, Ip SCY (2017) A selection guide for the new generation 6-dye DNA profiling systems. Forensic Sci Int Genet 30:34–42
Lindahl T (1993) Instability and decay of the primary structure of DNA. Nature 362:709–715
Longo MC, Berninger MS, Hartley JL (1990) Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. Gene 93:125–128
Loreille OM, Diegoli TM, Irwin JA, et al (2007) High efficiency DNA extraction from bone by total demineralization. Forensic Sci Int Genet 1:191–195. https://doi.org/10.1016/j.fsigen.2007.02.006
Lowe A, Murray C, Whitaker J et al (2002) The propensity of individuals to deposit DNA and secondary transfer of low level DNA from individuals to inert surfaces. Forensic Sci Int 129:25–34
Mameli A, Piras G, Delogu G (2014) The successful recovery of low copy number and degraded DNA from bones exposed to seawater suitable for generating a DNA STR profile. J Forensic Sci 59:470–473
Marrone A, Ballantyne J (2010) Hydrolysis of DNA and its molecular components in the dry state. Forensic Sci Int Genet 4:168–177
Martin B, Blackie R, Taylor D, Linacre A (2018) DNA profiles generated from a range of touched sample types. Forensic Sci Int Genet 36:13–19
Miloš A, Selmanović A, Smajlović L et al (2007) Success rates of nuclear short tandem repeat typing from different skeletal elements. Croat Med J 48:486–493
Muruganandhan J, Sivakumar G (2011) Practical aspects of DNA-based forensic studies in dentistry. J Forensic Dent Sci 3:38–45
Opel KL, Chung D, McCord BR (2010) A study of PCR inhibition mechanisms using real time PCR. J Forensic Sci 55:25–33. https://doi.org/10.1111/J.1556-4029.2009.01245.X
Pääbo S (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci 86:1939–1943
Pääbo S, Poinar H, Serre D et al (2004) Genetic analyses from ancient DNA. Annu Rev Genet 38:645–679
Pajnič IZ (2016) Extraction of DNA from human skeletal material. In: Methods in molecular biology. Humana Press, New York, pp 89–108
Palencia-Madrid L, de Pancorbo MM (2015) Next generation sequencing as a suitable analysis for authentication of degraded and ancient DNA. Forensic Sci Int Genet Suppl Ser 5:e338–e340
Poinar HN (2003) The top 10 list: criteria of authenticity for DNA from ancient and forensic samples. Int Congr Ser 1239(1239):575–579
Pruvost M, Schwarz R, Correia VB et al (2007) Freshly excavated fossil bones are best for amplification of ancient DNA. Proc Natl Acad Sci USA 104:739–744
Rastogi RP, Richa KA et al (2010) Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. J Nucleic Acids 2010:592980
Romanowski G, Lorenz MG, Wackernagel W (1991) Adsorption of plasmid DNA to mineral surfaces and protection against DNase I. Appl Environ Microbiol 57:1057–1061

Scientific Working Group on DNA Analysis Methods (2017) SWGDAM contamination prevention and detection guidelines for forensic DNA laboratories. In: https://www.swgdam.org/publications

Shrader C, Schielke A, Ellerbroek L, Johne R (2012) PCR inhibitors – occurrence, properties and removal. J Appl Microbiol 113:1014–1026

Sidstedt M, Romsos EL, Hedell R, et al (2017) Accurate digital polymerase chain reaction quantification of challenging samples applying inhibitor-tolerant dna polymerases. Anal Chem 89:1642–1649. https://doi.org/10.1021/ACS.ANALCHEM.6B03746

Sidstedt M, Steffen CR, Kiesler KM et al (2019) The impact of common PCR inhibitors on forensic MPS analysis. Forensic Sci Int Genet 40:182–191

Sidstedt M, Rådström P, Hedman J (2020) PCR inhibition in qPCR, dPCR and MPS – mechanisms and solutions. Anal Bioanal Chem 412:2009–2023

Smith PJ, Ballantyne JA (2007) Simplified low-copy number DNA analysis by post-PCR purification. J Forensic Sci 52:820–829

Taberlet P, Griffin S, Goossens B et al (1996) Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res 24:3189–3194

Timken MD, Swango KL, Orrego C, Buoncristiani MR (2005) A duplex real-time qPCR assay for the quantification of human nuclear and mitochondrial DNA in forensic samples: implications for quantifying DNA in degraded samples. J Forensic Sci 50:1–17

Turchi C, Melchionda F, Pesaresi M et al (2019) Performance of a massive parallel sequencing microhaplotypes assay on degraded DNA. Forensic Sci Int Genet Suppl Ser 7:782–783

van Oorschot RAH, Jones MK (1997) DNA fingerprints from fingerprints. Nature 387:767

Wiegand P, Kleiber M (2001) Less is more – length reduction of STR amplicons using redesigned primers. Int J Legal Med 114:285–287

Yang W (2011) Nucleases: diversity of structure, function and mechanism. Q Rev Biophys 44:1–93

Yang J, Lin D, Deng C et al (2019) The advances in DNA mixture interpretation. Forensic Sci Int 301:101–106

Young JM, Rawlence NJ, Weyrich LS, Cooper A (2014) Limitations and recommendations for successful DNA extraction from forensic soil samples: a review. Sci Justice 54:238–244

Zeng X, Elwick K, Mayes C et al (2019) Assessment of impact of DNA extraction methods on analysis of human remain samples on massively parallel sequencing success. Int J Legal Med 133:51–58

Zipper H, Buta C, Lämmle K et al (2003) Mechanisms underlying the impact of humic acids on DNA quantification by SYBR Green I and consequences for the analysis of soils and aquatic sediments. Nucleic Acids Res 31:e39

Validating Forensic DNA Workflows

50

Iman Muharam and Carla Paintner

Contents

Abstract

Method validation is a small but critical step in a forensic DNA laboratory's quality assurance system. Performing method validation is not only considered to be good science, to substantiate and to test the integrity of scientific processes in the laboratory, but also adheres to establishing quality standards and guidelines as dictated by standardization bodies. By performing rigorous validation, the forensic laboratory achieves confidence in the reliability of a testing method with respect to defining expert witness testimony and admissibility of forensic DNA evidence in the courts.

Despite the publication of international standards and guidelines for validating forensic DNA methods, laboratories often struggle to define their validation scopes, resulting in either insufficient validation or, often, over-validation. The

I. Muharam (✉)
HID Professional Services, Genetic Sciences Division, Thermo Fisher Scientific, Scoresby, VIC, Australia
e-mail: iman.muharam@thermofisher.com

C. Paintner
HID Professional Services, Genetic Sciences Division, Thermo Fisher Scientific, Singapore, Singapore
e-mail: Carla.Paintner@thermofisher.com

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_52

authors here will share several important considerations when performing validations, to support the laboratory's implementation of new technologies and provide a pathway to achieving international accreditation.

Keywords

Forensics · DNA · Validation · Quality assurance · Accreditation · PCR · Rapid DNA · Next-generation sequencing

Introduction

Forensic DNA evidence has the potential to change an individual's life to the extent of imprisonment or, in the case of some jurisdictions, even the sentencing of a death penalty. Forensic DNA testing therefore must be performed to the highest standards and survive any legal scrutiny in court. Each technical process in the forensic DNA workflow must be robust, reliable, and reproducible, as demonstrated through a validation process, and performed, analyzed and interpreted by trained and qualified scientists (Butler 2010). To provide confidence in the DNA evidence, each laboratory must determine which validation studies are relevant to the methodology (SWGDAM 2016). It is the laboratory's responsibility to scrutinize and review their methods through a rigorous and standardized validation process.

Knowledge and expertise on method validation and quality assurance systems are accumulated over a period of time, often as a necessity when the laboratory is undertaking preparations for accreditation under an international standard such as ISO/IEC 17025:2017 that specifies general requirements for the competence of testing and calibration laboratories. Accreditation is the formal recognition, often by an accreditation or standardization body, that a laboratory meets or exceeds a list of standards to perform specific tests (QAS 2020). For many forensic DNA laboratories in Australia, Western Europe, the USA, and other countries, operating under an international standard such as ISO/IEC 17025:2017 has been a fact of life for almost two decades, as this is a requirement without which they would not be able to operate. In many countries, accreditation or compliance against an international standard is mandated by a national agency. Examples of accreditation bodies include the ANSI National Accreditation Board (ANAB) based in North America, National Association of Testing Authorities (NATA) in Australia, and National Accreditation Board for Testing and Calibration Laboratories (NABL) in India. In other parts of the world, the concept of standardization and accreditation is new, with many forensic DNA laboratories in Southeast Asia and South Asia achieving accreditation in recent years, and many others are currently in progress. Obtaining accreditation can be a lengthy and expensive process and therefore demonstrates the laboratory's commitment to quality. Regardless of whether a testing laboratory is officially accredited or not by a standardization body, all forensic DNA laboratories should at minimum be performing good, responsible science, either through following accepted relevant forensic guidelines or standards or operating under Good Laboratory Practice

(OECD 1998), with the vision to eventually be capable of obtaining accreditation under an internationally accepted standard. A small but significant step in the process of becoming accredited is for the laboratory to test and evaluate their procedures, to determine their efficacy and reliability for forensic analysis through the process of validation.

What Is Validation?

According to ISO/IEC 17025:2017, validation is defined as the provision of objective evidence that a given item fulfils specified requirements, where the specified requirements are adequate for an intended use. This standard is broad in scope, as it applies to general testing or calibration laboratories, and not specifically forensic DNA testing facilities. For definitions that are more relevant to forensic DNA testing, we can refer to either the Federal Bureau of Investigation Director's Quality Assurance Standards for Forensic DNA Testing, or the Scientific Working Group on DNA Analysis Methods Validation Guidelines for DNA Analysis Methods:

> "Validation is a process by which a method is evaluated to determine its efficacy and reliability for forensic casework analysis. . ." (QAS 2020)

> "Validation is a process by which a procedure is evaluated to determine its efficacy and reliability for forensic casework and/or database analysis." (SWGDAM 2016)

The QAS and SWGDAM documents are two of the most comprehensive documents that standardize and guide the forensic DNA laboratory's validation and implementation efforts and often are referenced as critical resources as part of the laboratory's quality assurance system. The QAS document specifically describes the requirements that laboratories performing forensic DNA testing or using the Combined DNA Index System DNA database in the USA must follow, to ensure the quality and integrity of DNA evidence generated by the laboratory. The standards were recently updated to include rapid DNA testing, next-generation sequencing, and use of interpretation software. The SWGDAM document provides guidelines on the validation process itself, that is consistent in its messaging with the QAS document, that the laboratory should or may follow. Due to their focus on forensic DNA and their acceptance throughout the international forensic DNA community, these two documents are often used by the forensic DNA laboratory to provide a more granular and standardized approach to meet ISO/IEC 17025:2017 requirements around implementing new methods or technologies. Though not discussed in detail in this chapter, globally there are various standards and guidelines utilized by the forensic community and accreditation bodies. In this chapter, we will highlight the requirements and recommendations from the QAS and SWGDAM documents. It is important for your laboratory to identify the appropriate standards and guidelines based on your governmental or regional requirements.

For the laboratory's purposes, validation is performed to test a method for its reliability, limitations, and adequacy to suit its intended purpose. Validation minimizes reinvention of methods in different laboratories. Methods that have been validated are also more readily accepted by the scientific community, more likely to have already undergone scrutiny in a court, and allow results from different laboratories to be compared, either between branch laboratories in a domestic setting or across countries, as crimes are not often limited by international borders.

There are two types of validation required to implement or modify forensic DNA testing methods: developmental validation and internal validation. *Developmental validation* is the acquisition of test data and determination of conditions and limitations of a new or novel DNA method for use on forensic samples, whereas *internal validation* is an accumulation of test data within the laboratory to demonstrate that established methods and procedures perform as expected in the laboratory (QAS 2020).

Developmental validation pertains to the rigorous series of tests performed on a *new* method or technology preceding implementation. As an example, a new commercial STR (short tandem repeat) PCR kit must be developmentally validated prior to commercial release. Depending on the method or technology, developmental validation studies include characterization of the genetic marker, species specificity, sensitivity studies, stability studies, case-type samples, population studies, mixture studies, precision and accuracy studies, and PCR-based studies (QAS 2020), with the underlying scientific principles of the technology to be published in a peer-reviewed publication that is accessible to the community (SWGDAM 2016). A DNA laboratory can refer to another laboratory's developmental validation studies to reduce the need to reinvent the wheel. This assists with the uptake of new methods and technologies, reducing the burden on each laboratory's implementation process.

To implement a new technology in the laboratory, however, internal validation must be performed prior to using a procedure for forensic DNA applications. To put it simply, a developmental validation is performed by the vendor or creator of a *new* method, whereas an internal validation is performed by someone *using* that method in their laboratory. Although not as extensive as developmental validation, internal validation studies include similar experiments such as sensitivity and stochastic studies, precision and accuracy studies, mixture studies, known and non-probative sample studies, and contamination assessment studies (QAS 2020), often building on the developmental validation data. Additionally, an internal validation may accumulate data points that are *specific* for that laboratory's testing environment, population samples, or other local conditions, ensuring that the method remains reliable and performs as expected. A validated internal protocol will have an associated validation report detailing the validation process for the specific protocol that can be reviewed by internal stakeholders and external quality assessors. Although there is no requirement to publish an internal validation in a peer-reviewed publication, laboratories are encouraged to disseminate their internal validation by sharing a presentation or poster in a scientific meeting. Again, this approach allows forensic DNA analysts to learn the findings from other laboratories and speed up validation and implementation of that method in their own lab. Importantly, the

validation findings will form the foundation of the laboratory's standard operating procedure (SOP) for that specific method, as attributes such as *limitations* of the method would be determined.

Validation should be distinguished from other method-assessment processes, such as performance check or evaluation. A *performance check* is typically required when a laboratory obtains an additional critical instrument, of the same model already validated in the laboratory, prior to use in the routine workflow. The intent of a performance check is generally to assess a new piece of critical instrument for similarities and differences compared to the originally validated instrument. The performance check studies are usually limited to studies that highlight instrument performance (contamination, sensitivity, precision, and accuracy studies) versus the performance of samples or assays (mixture, non-probative studies). For example, if the laboratory performs their capillary electrophoresis on an Applied Biosystems 3500xL Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) that they have validated internally and, due to an increase in their sample throughput, decides to bring online an additional 3500xL Genetic Analyzer, then the new unit does not undergo the full validation process but instead will be subjected to a smaller-scale performance check study. A performance check may also be performed when relocating a critical instrument to a new location or another part of the laboratory (Iyavoo et al. 2017). Other critical instruments that will require performance checks include robotic systems, thermal cyclers, and other equipment that perform DNA typing results (QAS 2020). An *evaluation* is typically an assessment of a modified procedure in comparison to the original procedure, or other procedures, using similar DNA samples. For example, results from a number of DNA extraction methods are compared, using similar DNA samples and conditions, to determine which is more suitable for an intended task, such as success rate with low-level DNA samples (Sturk-Andreaggi et al. 2011), or performance across DNA yield and inhibitor removal (Zimmermann et al. 2009). Although the detailed criteria encountered in a validation are not a prerequisite in these processes, both performance checks and evaluations should be sufficiently documented, reported, and available for review. Depending on the application or intended use, the laboratory may decide to expand the performance check or evaluation testing to an internal validation (Iyavoo et al. 2017). Both the QAS and SWGDAM documents are specific in their language and provide clear guidance on what scenarios require which scope of testing. In this chapter, we will be reviewing the concept of validation as it pertains to *internal validation* that is relevant to the forensic DNA laboratory in implementing a new method or technology.

Why Validate?

Validations are everywhere. Unconsciously, we perform validations outside of the laboratory daily, without having a second thought, before making certain decisions. These tests or studies are vital in helping us reach well-educated, reliable results in

Table 1 Examples of approaches possible when purchasing a new vehicle

Your approach	Your accumulated evidence
Chose a new vehicle at random, not bothering to take it for a test drive	The vehicle was tested at the factory. Vehicle features were not personally researched or tested.
Chose a new vehicle at random and quickly test drive down the road	Quick test drive to test time taken to reach 100 km/hr. Assess comfort of the seats. Not all vehicle features were not researched or tested.
Research various makes/models of vehicles and thoroughly test drive to compare	Test the car at low, average, and high speeds. Ensure that the features of the car meet your daily driving needs that are important to you, for example, safety features, gas mileage, air conditioning, etc.

our personal lives. Sometimes, these studies can impact the lives and safety of not only ourselves but also our loved ones.

Consider the scenario of a long day at work. After spending the entire day in the laboratory, performing validation studies to implement a new STR kit, you are relieved to call it a day and make your way home to your family. Unfortunately, you are so focused on thinking about the validation results you generated today, you lose control of your car and collided with a tree. Thankfully you are safe and no one is injured, but your car is badly damaged and no longer drivable. After a short insurance claim process, you are now ready to purchase a new car. Reviewing the approaches in Table 1 below, which one would you most likely follow?

The sensible approach that most of us will select would be the last option. When making an important decision, we often compare advantages and disadvantages to perform an analysis on specific metrics that are important to us. In the example above, when choosing a new car to purchase, we consider many aspects and features, from safety and security to gas mileage and comfort features, such as air conditioning and heated seats. We need a vehicle that is not only comfortable for the long drives but also, perhaps more importantly, ensures that the occupants safely reach their destination. A car is also a big investment, in some cases costing more than a house, so we rightfully take our time to perform the necessary assessments to ensure we are satisfied with our final decision. If we were not to undertake these assessments, it would be of no surprise if we ended up at the beginning again, with a non-drivable car.

In the context of a validation, the car can be equated to a method or technology (e.g., a new STR kit), and our assessments, including testing at low, average, high speeds and assessing safety features, are our validation studies. When implementing a new STR kit, it is not advisable to purchase the kit and immediately start using it to amplify forensic DNA samples without any prior testing. Performing a small test of amplifying a subset of ten samples prior to implementation is a minor improvement but most likely not sufficient for the laboratory to understand the complexity and limitations of the kit. Only through performing a thorough validation study using

Table 2 Example approaches when implementing a new STR kit, with only the last approach satisfying international standards and guidelines

Your approach	Your accumulated evidence
Immediately amplify samples with the new kit	You use all the data generated and apply the protocol in your laboratory. Although instruments and environments were different, at least they were tested somewhere.
Amplify a small subset of samples and implement the new kit	Assess the new kit using a few samples within your laboratory. Shows that profiles were generated, so approved for use on casework samples.
Perform thorough validation studies and generate protocols based upon data generated	Test the dynamic range of the kit on each instrument. Appropriate samples/controls are concordant. Results are accurate, reliable, and repeatable.

standardized methods, followed by critical analysis of the data and presenting the data in a report for critical review by other stakeholders, will provide the laboratory with enough confidence to implement the new technology. The validation results would then be used to form the foundations of a new SOP and associated training module. Table 2 below compares various approaches a laboratory may take when implementing a new STR kit, but only the last approach would satisfy international standards and guidelines.

Validations enable the laboratory to gain confidence and quality in their laboratory processes and to confirm that a method is robust and reliable and the performance is reproducible under differing reasonable scenarios (e.g., when performed by different operators or using different sets of calibrated pipettes). Validation data can be compared against results published or presented elsewhere, against data from other laboratories, therefore showing consistency in results between labs. The accumulated validation evidence provides the forensic DNA analyst with confidence in the evidence that they are reporting to the court, for use in the justice system to convict a suspect or exonerate an innocent person. By validating methods and technologies, the laboratory is providing the best possible testing service for their clients, the people, and the public in their jurisdiction.

In many countries, validation is a key requirement of the ISO/IEC 17025:2017 international standard and must be completed for all methods and protocols in use in a forensic DNA laboratory, in order to obtain accreditation. There is no option to *not* perform a validation, because this can result in the laboratory losing their accreditation and therefore their right to perform the laboratory testing services. Laboratories that fail to complete appropriate validation testing may also be at risk for loss of funding, such as institutional grants or governmental subsidies. Occasionally, a laboratory may find itself having performed inadequate validation, resulting in an overflow impact on the appropriateness of their protocols, analysis guidelines, and even internal training programs. In the best-case scenario, the laboratory's quality assurance system will recognize this gap and trigger a reassessment of the validation and identify what additional work may be required. To reestablish the confidence of

not only the analysts but also the public and the court, a laboratory may need to repeat portions of the validation. The cost of this exercise can be large, depending on the revalidation scope, but the cost to the lab may be much larger if the revalidation is not performed, or if a validation is not conducted correctly in the first place. In the worst-case scenario, for example, inadequate validation can result in inaccurate results and miscarriage of justice, resulting in arrest and imprisonment of an innocent suspect (or conversely, the inadvertent release of a criminal back into the general population). These outcomes can result in media attention and international scrutiny, in many cases resulting in lab closures, cost to lab morale or reputation, and even imprisonment of the forensic DNA analyst. The recent report of the Texas Forensic Science Commission (2018) provides an unfortunate example of the impact on the laboratory when the quality assurance system breaks down and weaknesses were identified in the lab's contamination control procedures and training programs.

How to Validate?

Although many forensic DNA laboratories are accustomed to performing validations, the majority still struggle with the concept due to lack of time, resources, and sometimes knowledge (e.g., if their validation "expert" has left the lab). During periods of transformation, when laboratories adopt newly accepted technologies, such as rapid DNA, next-generation sequencing, or probabilistic genotyping, some laboratories can often find themselves in a continuous cycle of validating, or heavy investment in research because of their unfamiliarity with the new technology, and therefore confusion in determining which validation studies are appropriate. These situations may result in the laboratory experiencing validation fatigue, causing lengthy delays in implementing the new technology. For laboratories that have very little or almost no experience in validation, validating a technology is approached with caution and concern, due to lack of experience and in-house expertise. In either scenario, labs may wind up either under- or over-validating.

When creating an experimental design for a validation, performance check, or any other scope, it is important to first have a clear understanding of what standards the laboratory will need to meet. Throughout this chapter, we have focused on the FBI Director's Quality Assurance Standards and ISO/IEC 17025:2017 documents and therefore will continue to reference these standards. To have clarity on the requirements, firstly, laboratories will refer to the relevant audit checklist documents that assessors will actually use as they perform the accreditation audit and will either assess the lab as being compliant or not, or whether a particular clause does not apply. Secondly, the lab will need to determine how to perform the relevant studies required. Unfortunately, the standards only define *what* studies must be performed, and not the details of *how*. To answer the *how*, laboratories often refer to other supplementary sources of information or guidelines, such as the SWGDAM Validation Guidelines for DNA Analysis Methods (SWGDAM 2016) or Validation and Verification of Quantitative and Qualitative Test Methods (NATA 2018), formerly known as Technical Note 17 and most relevant in Australian testing facilities. These

guidelines provide the granularity on the validation process, by defining the terminologies used and outlining the experiments that should be performed, including specifics on what data need to be defined and sometimes what samples need to be analyzed. For an internal validation, SWGDAM recommend the following studies: known and non-probative evidence samples or mock evidence samples, sensitivity and stochastic studies, precision and accuracy, mixture studies, and contamination assessment (SWGDAM 2016). Further to that, SWGDAM specifies that the sensitivity and stochastic studies should demonstrate sensitivity levels of the test, by determining the dynamic range, ideal target range, limit of detection, limit of quantitation, heterozygote balance, and signal-to-noise ratio (SWGDAM 2016). SWGDAM does provide flexibility for the laboratory to determine the suitability of each study, based on the methodology under examination, and may determine that a particular study is not required and also provides the laboratory with the freedom to evaluate the appropriate sample number and types (SWGDAM 2016). This helps to ensure that the laboratory's validation is always fit for purpose and as per required to demonstrate the reliability and potential limitations of that specific method.

Often, a laboratory still finds it challenging to start their validation project and prefers to seek expert support or outsource their validation project to a consultant with deeper knowledge, who can guide the laboratory on their validation. Alternatively, the lab may work with a vendor to create a bespoke validation support solution that includes providing experimental designs and personnel to perform the actual laboratory work, data analysis, and report writing. The Human Identification (HID) Professional Services (HPS) team from Thermo Fisher Scientific was set up in 2007 to address validation support required by forensic DNA laboratories in the USA. Since then, HPS team has performed more than 700 projects across 36 countries, including Singapore, Thailand, Japan, India, Scotland, Italy, Brazil, Mexico, South Africa, United Arab Emirates, and Australia. The HPS validation scope follows the QAS requirements and SWGDAM guidelines, to help the laboratory achieve ISO/IEC 17025:2017 accreditation, by providing a standardized experimental design together with the kits, reagents, consumables, and expert personnel to perform the laboratory work and generate the requisite data. Data analysis and report writing are completed by trained and experienced personnel, with the results provided back to the laboratory in a comprehensive validation report, followed by teachback training to transfer the validation knowledge to the DNA analyst. This type of solution provides the laboratory with a standardized implementation program with defined studies, timelines, and costs and ensures the validation project does not extend unnecessarily for an unpredictable timeframe or result in a budget blowout.

An example of a standardized scope for validating an STR kit, as developed by HPS, is presented in Table 3 below. Also highlighted in this table is the QAS section addressed by each study and how the lab may use those results as part of their validation, foundation for SOPs, or data analysis and interpretation guidelines.

For most laboratories, the scope above provides an adequate number of data points to validate a commercial STR kit such as the GlobalFiler™ IQC PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA, USA). The scope has been optimized to allow data from the same sample set to be used to address more

Table 3 Example of a standardized scope for validating a commercial STR PCR kit (such as GlobalFiler™ IQC PCR Amplification Kit on an Applied Biosystems 3500xL Genetic Analyzer), including the QAS that is addressed and how the laboratory may use the data for their validation and SOP. This scope was developed by HID Professional Services (Thermo Fisher Scientific).

Validation study	Scope	QAS audit	How this study can be used
Minimum threshold and contamination study	Negative amplification controls are analyzed at 1 RFU to determine noise of instrument and chemistry. Maximum, average, standard deviation, average plus three standard deviations (limit of detection), and average plus ten standard deviations (limit of quantification) are calculated to determine minimum thresholds to be used for analysis. Contamination is assessed in negative controls and blanks using the minimum thresholds.	8.3.1(5)	Assist in the determination of an analytical threshold. Determine how to address instances of contamination.
Sensitivity and stochastic study	A genomic DNA dilution series (4, 2, 1, 0.5, 0.25, 0.125, 0.063, 0.031, and 0.016 ng) is amplified in triplicate. Average peak heights, peak height ratios, dropout of alleles/loci, peak height of surviving sister allele, and interlocus balance are assessed.	8.3.1(3)	Establish analytical and stochastic thresholds Develop interpretation guidelines. Evaluate input amounts that minimize pull-up peaks, off-scale data, and other artifacts.
Precision study: repeatability and reproducibility	Repeatability of size variation is assessed using allelic ladders injected four times; standard deviation of allele size is calculated by injection, by capillary, and for all data together and evaluated for values greater than 0.15 bp. Repeatability of peak heights is assessed using replicate samples from the sensitivity and stochastic study dilution series. Reproducibility of peak heights is assessed using Control DNA 007.	8.3.1(2)	Establish protocols for the number and placement of allelic ladders on plates for injection. Understand variation in average peak heights for samples with the same DNA input amounts.

(continued)

Table 3 (continued)

Validation study	Scope	QAS audit	How this study can be used
	Reproducibility of genotypes is assessed using Control DNA 007 and NIST Standard Reference Material™ 2391d Component.		
Accuracy study	NIST Standard Reference Material™ 2391d Component is amplified with 1.0 ng DNA target input. Genotype results are compared to the certificate of analysis	8.3.1(2) and 8.4	Demonstrate compliance with Standard 8.4 of the *Quality Assurance Standards for Forensic DNA Testing Laboratories*
Known and non-probative sample study	20 known and non-probative samples are supplied by the laboratory and amplified with 1.0 ng DNA target input. Samples are analyzed for concordance and expected performance when compared to previous results supplied by the laboratory.	8.3.1(1)	Identify parameters for the STR kit, including the amount of input DNA, analytical and stochastic thresholds, and interpretation guidelines. Establish standard operating procedures.
Mixture study	Male-female genomic DNA mixture series (1:0, 1:1, 1:2, 1:4, 1:10, 1:20, 1:40, 1:80, 0:1) is amplified at 1.0 ng target input. Peak heights are analyzed for mixture ratios and presence, absence, or masking of minor component alleles.	8.3.1(4)	Establish protocols for data interpretation and statistical analyses, including the mixture ratios of components in mixed DNA samples and the amount of DNA to amplify.
Assessment of non-allelic peaks	Samples are assessed for the following PCR artifacts: minus stutter, plus stutter, incomplete adenylation, and dye artifacts. Samples are assessed for capillary electrophoresis artifacts such as pull-up.	N/A	Establish protocols for data analysis and mixture interpretation. Determine whether the default marker-specific stutter filters are appropriate.

than one validation study. For example, the sensitivity study sample set, consisting of genomic DNA dilution series from 4 ng input to 0.016 ng input amplified in triplicate, is used to assess peak height ratios for heterozygous loci (Fig. 1), interlocus balance (Fig. 2), and repeatability of peak heights (Table 4).

The peak height ratio data show the peak balance for heterozygous loci across different DNA input amounts (Fig. 1). These results show that peak imbalance increases

Dilution	Input DNA (ng)	Rep	D3S1358	vWA	D16S539	AMEL	D8S1179	D21S11	D19S433	TH01	FGA	D5S818	D13S317	D7S820	SE33	D10S1248	D1S1656	D12S391	D2S1338
A	4	1	91%	90%	95%	85%	96%	95%	89%	99%	86%	98%	89%	91%	90%	99%	94%	85%	95%
		2	81%	95%	94%	96%	98%	98%	98%	100%	85%	97%	94%	88%	84%	93%	87%	100%	99%
		3	88%	94%	86%	79%	96%	99%	91%	96%	97%	93%	98%	89%	86%	87%	85%	87%	98%
B	2	1	96%	92%	98%	88%	92%	98%	99%	95%	97%	97%	96%	97%	81%	96%	89%	96%	99%
		2	97%	94%	81%	83%	82%	83%	94%	98%	98%	95%	100%	89%	93%	97%	94%	95%	93%
		3	86%	88%	97%	82%	95%	94%	99%	90%	87%	90%	95%	88%	94%	93%	86%	87%	91%
C	1	1	87%	93%	94%	94%	94%	89%	71%	88%	92%	97%	94%	74%	92%	96%	91%	78%	86%
		2	95%	94%	99%	65%	86%	87%	88%	99%	99%	90%	95%	94%	87%	76%	96%	78%	98%
		3	99%	98%	91%	81%	77%	88%	98%	90%	91%	94%	86%	96%	91%	86%	87%	96%	98%
D	0.5	1	77%	90%	80%	95%	78%	81%	100%	98%	88%	94%	85%	97%	92%	86%	95%	75%	86%
		2	94%	91%	74%	93%	97%	82%	90%	81%	92%	99%	93%	86%	82%	89%	82%	98%	76%
		3	87%	96%	92%	89%	95%	92%	87%	91%	75%	100%	88%	90%	99%	97%	92%	92%	79%
E	0.25	1	76%	73%	92%	90%	66%	85%	84%	96%	80%	94%	77%	88%	92%	85%	86%	81%	57%
		2	88%	99%	68%	94%	88%	62%	79%	85%	99%	91%	90%	97%	68%	51%	88%	82%	51%
		3	84%	76%	100%	82%	99%	71%	94%	79%	67%	73%	98%	85%	96%	98%	77%	97%	62%
F	0.125	1	73%	62%	62%	80%	81%	83%	51%	96%	98%	88%	66%	93%	79%	71%	69%	76%	98%
		2	63%	80%	43%	67%	72%	84%	98%	56%	94%	70%	87%	95%	57%	71%	50%	65%	88%
		3	98%	62%	86%	83%	87%	100%	62%	86%	97%	72%	83%	59%	80%	80%	87%	64%	93%
G	0.063	1	79%	55%	53%	75%	99%	23%	55%	45%	56%	91%	81%	64%	78%	97%	80%	89%	61%
		2	92%	61%	97%	46%	99%	37%	84%	80%	97%	56%	85%	83%	47%	81%	64%	59%	60%
		3	65%	51%	71%	62%	73%	68%	92%	42%	83%	59%	85%	63%	78%	55%	89%	95%	72%
H	0.031	1	96%	33%	54%	53%	79%	58%	21%	58%	24%	83%	29%	58%	89%	81%	41%	86%	54%
		2	41%	57%	60%	36%	98%	92%		18%	65%	90%	43%	18%	14%	24%	38%	17%	60%
		3	57%	57%	91%	37%	99%	70%	13%	35%	83%	41%	91%		69%	48%	93%	56%	63%
I	0.016	1	55%	38%		51%	32%	43%	53%		30%	52%	29%	59%	69%	28%		73%	
		2	94%	76%		38%	40%	84%	55%	37%	32%	67%	13%	97%	24%	63%	67%	57%	97%
		3	73%		78%			22%	29%		80%	98%			39%	54%	86%		

Fig. 1 Example of peak height ratio data for heterozygous loci, observed after amplification of the sensitivity study DNA dilution series, in triplicate, when validating the GlobalFiler™ IQC PCR Amplification Kit

as the input DNA amount is reduced, especially once the input is within stochastic range or less than 0.1 ng. In this example data set, peak height ratios were generally above 70% with input amounts of at least 0.5 ng of DNA and above 50% with input amounts below 125 pg, demonstrating variability associated with low-template stochastic amplification effects, as evidenced by peak height ratios as low as 13% and/or allelic dropout. In addition, when graphing the heterozygous peak heights for each marker in the replicate amplifications of the sensitivity study DNA dilution series, we observe the interlocus balance is uniform across the DNA profile at higher input DNA amounts but

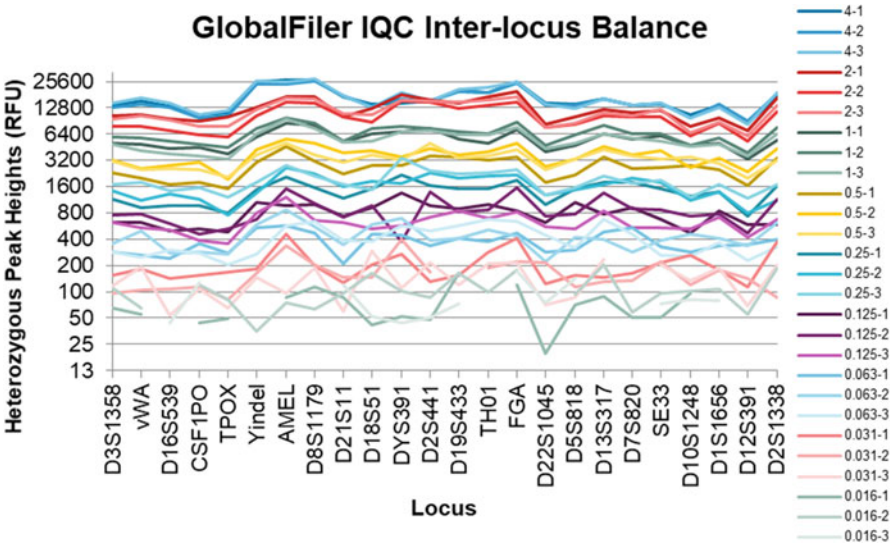


Fig. 2 Example of interlocus balance data obtained after amplification of the sensitivity study DNA dilution series, in triplicate, when validating the GlobalFiler™ IQC PCR Amplification Kit

Table 4 Example of data showing repeatability of peak heights from amplification of the sensitivity study DNA dilution series, in triplicate, when validating the GlobalFiler™ IQC PCR Amplification Kit

Input DNA (ng)	Replicate 1	Replicate 2	Replicate 3	Average (RFU)	Standard Dev (RFU)	RSD (%)
4	16,293	15,995	16,740	16,343	375	2
2	12,504	10,112	11,467	11,361	1200	11
1	5608	6510	5587	5902	527	9
0.5	2702	3668	3341	3237	491	15
0.25	1416	1605	1867	1629	226	14
0.125	791	869	634	764	120	16
0.063	375	442	410	409	33	8
0.031	206	166	148	173	30	17
0.016	72	106	96	91	18	19
					Average:	12

less so when the input is in stochastic range (Fig. 2). This data set also aids in setting a stochastic threshold at which it is reasonable to assume that sister allele dropout does not occur. The average heterozygous peak heights calculated for the three replicate amplifications of the DNA dilution series were generally similar at all levels of DNA not affected by stochastic effects, demonstrating repeatability of the GlobalFiler IQC kit, with an average 12% variation among replicate samples across all dilutions (Table 4). Based on these results, the laboratory has gained objective data that optimal sensitivity is seen when the target DNA input amplified under standard conditions ranges between 0.5 ng and 1 ng, resulting in optimal peak heights for interpretation of reliable and reproducible DNA profile data.

This example might come across as a massive leap from the question of "Why do you need to validate?" to "How do I validate?" If the laboratory is new to the concept of validation, or has not performed validations according to the standards reviewed here, the lab may attempt to devise an approach for smaller-scale validation or performance check projects and, within that, to focus on a subset of more critical studies first and phase in remaining studies over a number of weeks or months, depending on the time available. For example, a laboratory may elect to perform studies that provide them with an appreciation of the baseline performance of their new kit, method, or critical instrument, by focusing on experiments that utilize more standardized or controlled sample sets, such as contamination assessment, precision and accuracy, sensitivity, and stochastic studies that can be achieved using control DNA, allelic ladder, or purified genomic DNA samples. These fundamental studies can provide the laboratory with some immediate insight into the state of their laboratory or the system under investigation, such as:

- A precision study based on injecting allelic ladder samples multiple times and examining the variation in the resulting base pair sizes over a period of time can assist in showing the laboratory whether the environmental operating temperature allows collection of data with high consistency and robustness.
- A sensitivity study utilizing a range of input DNA can provide the laboratory with objective data on the performance of their system.
- A contamination assessment study can identify the presence or absence of either systemic or sporadic contamination in the laboratory and therefore trigger improvements in contamination management.

As the laboratory gains confidence, the studies can expand into additional experiments that utilize laboratory-type samples and therefore may aide the lab in creating SOPs, establishing thresholds, or improving data analysis and interpretation guidelines, such as:

- A known and non-probative study covering sample types that are routinely processed in the lab can help to demonstrate performance of routine samples using the new method.

- A mixture study can help to elucidate limitations in the system when dealing with samples containing multiple contributors at different mixture ratios and inputs, resulting in better analysis and interpretation guidance for the DNA analysts.

Below are additional suggestions that laboratories can consider as they prepare, plan, and execute their validation projects.

1. **Ensure your laboratory has essential infrastructure in place**. Forensic DNA facilities must have secure and controlled access, with pre-PCR techniques such as evidence examination, DNA extraction, and PCR setup to be performed at separate times or separate spaces, and separate again from post-PCR areas to reduce contamination risk (QAS 2020). Guidelines exist for the planning, designing, and construction of forensic DNA laboratories (NIST 2013) that can assist in ensuring the facility meets international standards. Before embarking on implementing a new method or technology, such as mitochondrial DNA sequencing using a next-generation sequencing platform such as the HID Ion GeneStudio™ S5 Prime System (Thermo Fisher Scientific, Waltham, MA, USA), the laboratory may need to consider modifying existing space or adding an additional section that will allow proper testing and analysis of mitochondrial DNA, as this workflow requires further contamination control measures to prevent double cross-contamination (NIJ 2018; Edson et al. 2004). In addition to the laboratory facility itself, the forensic DNA staff are also an integral part of the functioning of the laboratory. The QAS (2020) specifies that laboratory personnel shall have the education, training, and experience required to examine forensic evidence and provide testimony and further outlines minimum educational requirements, such as a minimum of a bachelor's degree, with at least nine semesters covering biochemistry, genetics, and molecular biology. To undertake an analyst role, the personnel requires 6 months of forensic human DNA laboratory experience and to have successfully completed the required internal training program with verification and approval by the technical leader (QAS 2020). Without a critical understanding of forensic DNA analysis and the effect of data quality on analysis and interpretation, the validation project will suffer from a lack of technical leadership. Moreover, in preparing for the validation, the laboratory should ensure they have appropriate PPE, calibrated pipettes, filtered pipette tips, and only use new, unopened, unexpired, and correctly stored kits, reagents, and consumables to safeguard the validity of the validation results obtained. An internal staff elimination database will also be beneficial to rule out laboratory staff as being a source of any contaminating peaks (Forensic Science Regulator 2020; OSAC 2019).
2. **Plan your project as a team**. In our experience, the success of any project is dependent on the planning that was performed in preparation for the project. Before starting the planning process, consider the project management tool that will best meet your needs. Basic software packages available on most computers, such as Microsoft® Excel, can be a good starting point and is powerful enough to

generate plans, schedules, and Gantt charts. If the laboratory is working on larger or more complex projects, various commercial project management software packages can be purchased for installation on a computer, or a subscription can be purchased for online project management platforms, such as Smartsheet (www.smartsheet.com), that can be accessed over the Internet using a browser on your computer or even through a mobile phone app. These software tools may include further functionalities, such as creation of project dashboards, sending alerts/notifications, and even tracking your budget. Projects are rarely an individual undertaking and usually involve a group of people with unique skill sets to address different project requirements, so it is important to define the project team and assign roles and responsibilities. For example, the project manager may be the laboratory manager, and the actual laboratory work is performed by a technician, with data analysis performed by a case reporting scientist and report writing by a scientist from the quality team. The laboratory should think about the project team holistically, starting from project conception to eventual implementation, and consider the stakeholders throughout. A responsibility assignment matrix may be beneficial as well, so the team can associate tasks within the project with at least one role based on who is responsible and accountable and needs to be consulted or informed (abbreviated as RACI). This ensures that the team members have clarity on their roles within the project, who to consult when support is required, and who needs to be informed once the project is completed and implementation can begin.

3. **Don't reinvent the wheel**. If the laboratory is performing internal validation, this means that the method has already had developmental validation performed by the vendor and most likely other laboratories have performed internal validation, and possibly published their work. When performing a validation, it is always useful to relate the validation studies back to the vendor's development validation. The vendor's validation is usually available as part of the method's user guide or manual and additionally published in a peer-reviewed journal. Commercial products such as the GlobalFiler™ PCR Amplification Kit undergo this process regularly (Ludeman et al. 2018). By reviewing these publications, the laboratory can become aware of the work already performed. This means that the lab can review previous data and identify any gaps that they may want to address in their own validation. For example, a species specificity study is performed in the developmental validation but often only utilizes common primates, such as chimpanzee, and common domesticates, such as dog, cat, and mouse, and a limited number of bacterial pools (Ludeman et al. 2018). Based on experiences with their own routine sample types, a laboratory may expand the species specificity study to include additional species relevant to their workflow within their own environment or geographical location. For example, additional human-associated microbial species may be examined due to their prevalence in common forensic samples analyzed in a specific laboratory (Martin et al. 2014). In the case of newer technology, such as rapid DNA testing using the Applied Biosystems RapidHIT™ ID System (Thermo Fisher Scientific, Waltham, MA, USA) or next-generation sequencing on HID Ion GeneStudio™ S5 Prime System (Thermo Fisher Scientific, Waltham, MA, USA), there is often a lack of published

validation guidelines; the vendor's developmental validation is the best source of information on what validation studies are relevant and what results can be expected (Salceda et al. 2017; Cihlar et al. 2020). Almost all journal publications will include the details of a corresponding author, who can be contacted should the laboratory wish to request further information, to share ideas, or simply to connect and collaborate on future projects. By reviewing the literature, a laboratory can consider approaches taken by other laboratories and create a validation plan that is fit for purpose.

4. **Review the validation data holistically and make sure to understand the limitations of the system**. When analyzing a complex DNA profile with multiple contributors, a forensic DNA analyst is taught to consider the data in the entire profile and not just a small number of loci (SWGDAM 2017). Similarly, a validation project consists of multiple studies with data sets that build on top of each other and should be considered in totality, with no particular piece of data scrutinized on its own. To calculate analytical and stochastic thresholds, a laboratory will not only need to consider baseline noise obtained from the contamination assessment study from amplifying negative controls but also consider the change in baseline noise when DNA is present in the PCR, average heterozygous peak heights across different DNA input amounts, peak imbalance, performance with case-type samples, and ability to characterize artifacts and differentiate between true peaks and artifact peaks. These additional considerations utilize data from the sensitivity and stochastic study, known and non-probative study, and assessment of non-allelic peaks. The laboratory should make a conscious effort to compare the results obtained against observations in the developmental validation or other published material, to ensure that the method is performing as expected and adequate for an intended use. Further to that, a method or technology will have inherent limitations, just like any other system, and this must be reflected when designing the validation experiments. Through appropriate experiment planning and analyzing the validation data objectively, the laboratory can not only determine the optimal conditions that will result in the best data quality but also understand what impact any adverse conditions may have on their data. For example, most commercial STR kits are optimized for a target DNA input of 0.5 ng to 1 ng amplifiable DNA; however, the lab should understand what to expect when either lower or higher amounts of DNA are amplified instead. Typically, lower DNA input (below 100 pg of total DNA) will result in stochastic effects and increased observations of peak imbalance, allelic dropout, and allelic drop-in, whereas higher input (above 2 ng total DNA) results in an increase in both PCR- and instrument-based artifacts, such as stutter and pull-up. Finally, ensure that the completed validation report is reviewed by key stakeholders in the project, who possibly come with different experiences or from different roles throughout the lab, so they can examine the results from their own lenses specific to their roles. This allows the validation to be reviewed from different perspectives and ensures the eventual implementation considers any concerns or requirements from all sections of the laboratory. Through this alignment across the lab, the implementation process will be smoother, and the benefits of the new method can be realized by all staff members.

Fig. 3 Internal validation and laboratory accreditation lead to laboratory excellence

Through performing an effective validation process, the laboratory can confirm the robustness, reliability, and reproducible performance of their method, providing confidence in reporting the results for use in the justice system. When combined with a quality assurance system that is independently audited against international standards and accredited by a standardization body, a laboratory shows commitment to quality and performance excellence, as illustrated in Fig. 3. All forensic DNA laboratories, regardless of the absence of any national requirement for accreditation, should strive to work toward the highest standards and continue to drive adoption of forensic DNA testing as an important and effective crime fighting tool.

Conclusion

Performing a validation may appear to be a complicated process; however, standards and guidelines exist and are well accepted in the forensic community. Forensic DNA laboratories are encouraged to standardize their approach to validating new methods and technologies and collaborate or connect with vendors and other laboratories to perform effective validation testing. The validation process ensures that methods are standardized, comparable, and fit for the purpose and is therefore a critical part of the laboratory's quality assurance system.

References

Butler JM (2010) Fundamentals of forensic DNA typing, validation. Elsevier, San Diego, p 300

Cihlar JC, Amory C, Lagacé R, Roth C, Parson W, Budowle B (2020) Developmental validation of a MPS workflow with a PCR-based Short amplicon whole mitochondrial genome panel. Genes 11:1345. https://doi.org/10.3390/genes11111345

Edson SM, Ross JP, Coble MD, Parson TJ, Barritt SM (2004) Naming the dead – confronting the realities of rapid identification of degraded skeletal remains. Forensic Sci Rev 16:63–90

Forensic Science Regulator (2020) Code of practice and conduct protocol: DNA contamination detection – the management and use of staff elimination DNA databases FSP-P-302 Issue 2. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/915377/302-Elimination_database_protocol_v2.pdf. Accessed 3 Dec 2020

ISO/IEC 17025:2017 (2017) General requirements for the competence of testing and calibration laboratories. 3rd edn

Iyavoo S, Draus-Barini J, Haizel T (2017) Validation of ABI 3500xL genetic analyzer after decommissioned and recommissioned at new premises. Forensic Sci Int 6:e542–e543. https://doi.org/10.1016/j.fsigss.2017.09.197

Ludeman M, Zhong C, Mulero JJ, Lagacé RE, Hennessy LK, Short ML, Wang DY (2018) Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. Int J Legal Med 132:1555–1573. https://doi.org/10.1007/s00414-018-1817-5

Martin P, Fernández de Simón L, Luque G, José Farfán M, Alonso A (2014) Improving DNA data exchange: validation studies on a single 6 dye STR kit with 24 loci. Forensic Sci Int Genet 13: 68–78. https://doi.org/10.1016/j.fsigen.2014.07.002

NATA (2018) National Association of testing authorities general accreditation guidance – validation and verification of quantitative and qualitative test methods. https://www.nata.com.au/phocadownload/gen-accreditation-guidance/Validation-and-Verification-of-Quantitative-and-Qualitative-Test-Methods.pdf. Accessed 5 Oct 2020

NIJ (2018) National Institute of Justice Office of Law Enforcement Standards Forensic Laboratories: handbook for facility planning, design, construction, and moving. https://www.ncjrs.gov/pdffiles/168106.pdf. Accessed 26 Nov 2009

NIST (2013) National Institute of Standards and Technology NISTIR 7941 forensic science laboratories: handbook for facility planning, design, construction, and relocation. https://doi.org/10.6028/NIST.IR.7941

OECD (1998) OECD principles of good laboratory practice. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/mc/chem(98)17&doclanguage=en. Accessed 3 Oct 2020

OSAC (2019) Organization of Scientific Area Committee for forensic science best practice recommendations for the management and use of quality assurance DNA elimination databases in forensic DNA analysis. https://www.nist.gov/system/files/documents/2020/08/04/Best%20Practice%20Recommendations%20for%20the%20Management%20and%20Use%20of%20Quality%20Assurance%20DNA%20Elimination%20Databases%20in%20Forensic%20DNA%20Analysis_OSAC%20Proposed.pdf. Accessed 3 Oct 2020

QAS (2020) Quality assurance standards for forensic DNA testing laboratories. https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view. Accessed 3 Oct 2020

Salceda S, Barican A, Buscaino J, Goldman B, Klevenberg J, Kuhn M, Lehto D, Lin F, Nguyen P, Park C, Pearson F, Pittaro R, Salodkar S, Schueren R, Smith C, Troup C, Tsou D, Vangbo M, Wunderle J, King D (2017) Validation of a rapid DNA process with the RapidHIT® ID system using GlobalFiler® express chemistry, a platform optimised for decentralised testing environments. Forensic Sci Int Genet 28:21–34. https://doi.org/10.1016/j.fsigen.2017.01.005

Sturk-Andreaggi K, Diegoli TM, Just R, Irwin J (2011) Evaluation of automatable silica-based extraction methods for low quantity samples. Forensic Sci Int 3:e504–e505. https://doi.org/10.1016/j.fsigss.2011.10.003

SWGDAM (2016) Scientific Working Group on DNA analysis methods validation guidelines for DNA analysis methods. https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_813b241e8944497e99b9c45b163b76bd.pdf. Accessed 3 Oct 2020

SWGDAM (2017) Scientific working group on DNA analysis methods interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories. https://1ecb9588-ea6f-4feb-971a-73265dbf079c.filesusr.com/ugd/4344b0_50e2749756a242528e6285a5bb478f4c.pdf. Accessed 5 Oct 2020

Texas Forensic Science Commission (2018) Final report on complaint by Lisa Gefrides against the Houston Forensic Science Center. https://www.txcourts.gov/media/1442270/17-04-final-report-lisa-gefridies-against-houston-forensic-science-center-07202018.pdf. Accessed 7 Nov 2020

Zimmermann P, Vollack K, Haak B, Bretthauer M, Jelinski A, Kugler M, Loidl J, Pflug W (2009) Adaptation and evaluation of the PrepFiler™ DNA extraction technology in an automated forensic DNA analysis process with emphasis on DNA yield, inhibitor removal and contamination security. Forensic Sci Int 2:62–63. https://doi.org/10.1016/j.fsigss.2009.08.162

Quality Control Measures in Short Tandem Repeat (STR) Analysis

51

Heather Miller Coyle

Contents

Abstract

Quality control is essential to the integrity and scientific validity of the DNA test result in human identification. In all scientific methods, there is room for error, and a process must be in place to guarantee that reagents, equipment, and methods are functioning properly. Four main areas are highlighted in this chapter: DNA collection methods, polymerase chain reaction (PCR) factors, electrokinetic injection considerations, and DNA data interpretation. Forensic science laboratories are provided guidance documents by the DNA Advisory Board and the Organization of Scientific Area Committees designed to improve uniformity in DNA test results nationwide. As a path forward, this is necessary since much of past policy created a situation of variation in methodology and data interpretation. It is highly desirable to have a simple uniform methodology that can be applied globally for sharing of data and consistency in the courts.

Keywords

Human identification · DNA · PCR · DAB · OSAC · Quality control · Forensic science

H. Miller Coyle (✉)
Forensic Science Department, Henry C. Lee College of Criminal Justice and Forensic Sciences, University of New Haven, West Haven, CT, USA
e-mail: Hcoyle@newhaven.edu

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_53

Introduction

The current oversight for forensic DNA analysis is provided through the National Institute of Standards and Technology (NIST) via review by OSAC, the Organization of Scientific Area Committees. This group facilitates implementation of forensically sound science and technical methods by drafting and evaluating standards that are to be used by forensic science laboratories (www.nist.gov). Previously, the Technical Working Groups sponsored by the Federal Bureau of Investigation (FBI) were performing this role by providing recommendations for implementation of new and current technology in the forensic science workplace. In 1989, the Guidelines for a Quality Assurance Program for DNA Analysis was prepared by TWGDAM and then revised three times after the TWGDAM working group of private and public sector forensic scientists was renamed to SWGDAM, the Scientific Working Group for DNA Analysis Methods in 1999. Prior to 1989, there were no national recommendations for forensic science laboratories to follow.

The DNA Advisory Board (DAB) began its 5-year project to create a set of federal quality standards for forensic laboratories in 1995. These standards include the quality manual areas of forensic science laboratory organization, personnel, facilities, evidence tracking and storage, validation, test procedures, equipment calibration and maintenance, technical review and reports, proficiency testing and corrective action, laboratory audits, safety guidelines, and use of subcontractors for testing of backlogged samples (https://strbase.nist.gov/validation/Intro_to_DAB_Standards.pdf). Two sets of standards were created by the DNA Advisory Board. In October 1998, The Quality Assurance Standards for Forensic DNA Testing Laboratories (Forensic Standards) became active to provide oversight to the activities of DNA laboratories that analyze criminal casework (https://strbase.nist.gov/QAS/Final-FBI-Director-Forensic-Standards.pdf). The Quality Assurance Standards for Convicted Offender DNA Databasing Laboratories was initiated in April 1999 for the analysis of DNA known reference samples collected from convicted offenders (https://strbase.nist.gov/QAS/Final-FBI-Director-Databasing-Standards.pdf).

DNA is a molecule that defines the function of genes but also contains noncoding regions that vary from individual to individual that are useful for human identification. Deoxyribonucleic acid (DNA) is composed of two polynucleotide chains carrying genetic instructions for all known organisms and their functions. DNA is a double helix comprised of two antiparallel strands that form the backbone of the helix by creating a sugar-phosphate structure that are bound together by complementary nucleotide base pairing, adenine (A) to thymine (T) and guanine (G) to cytosine (C). The regions of interest for short tandem repeat analysis have been prescreened and preselected for genetic variability in human populations. The estimated frequency of each allele at any given chromosome locus has been calculated in sample ethnic populations and have been shown to vary slightly in frequency based on original founder populations. The high discrimination power or ability to individualize comes from both the great number of allele possibilities at any given locus and the larger number of loci used in combination to calculate estimated relative frequency of any given DNA profile compared to the unrelated random

individuals in a population. The level of STR analysis is now so sophisticated that a human DNA profile is unique to an individual unless they have a genetic identical twin.

The integrity of DNA data is essential for establishing a link to a correct DNA source. Deoxyribonucleic acid or DNA is the biological molecule used to identify potential donors to biological evidence. It is useful as a forensic tool to indicate associations of individuals to biological evidence such as blood, semen, saliva, urine, and epithelial cells. Although an association can be identified by DNA testing, the timing of deposit and the context to the case often is in question. After evidence collection, DNA is extracted, quantified, amplified, and detected by capillary electrophoresis in the forensic science laboratory. Analysis software is used for the viewing and interpretation of DNA fragments to make a conclusion about a potential inclusion or exclusion of a DNA source. Additional DNA mixture interpretation software may be used as an aid to interpret complex DNA samples. The basic steps used to generate a human DNA profile are carefully quality controlled to be sure that a scientifically accurate result can be achieved. This leaves the court debate to the meaning of the DNA in the case rather than with the scientific validity of the DNA result itself. The basic steps of this process can be broken down into the component parts: (a) DNA collection, (b) DNA extraction, (c) DNA quantitation, (d) DNA amplification, (e) capillary electrophoresis, (f) data interpretation, and (g) statistical assessment.

DNA Methods

DNA collection. DNA evidence collection refers to the recognition and collection of biological evidence that may be pertinent to the case. Forms of DNA evidence include blood, saliva, semen, urine, touch DNA, bone, tissue, hair, vomit, feces, perspiration, and tears. These forms of evidence may be found as part of property crimes, homicides, sexual assaults, child abuse cases, robbery, kidnapping, etc. DNA collection strategies include the use of personal protective equipment (PPE) to avoid contamination, and therefore error, in the final DNA results. Saliva can contain droplets with DNA molecules that can fall on evidence or a surface and be mixed inadvertently into the crime scene sample (Aparna and Shanti Iyer 2020; Pandeshwar and Das 2014). Therefore, masks and face shields are essential for prevention of DNA contamination by evidence collection personnel. Likewise, shed epithelial cells from the skin and hairs may be falsely included in DNA evidence should it be deposited on the evidence, and full disposable crime scene suits, gloves, and head coverings are required to prevent this phenomenon (Zajac et al. 2019). The key to high-quality DNA results in criminal casework relies on best practices for quality control at the crime scene for evidence recognition, collection, and preservation (Goray et al. 2012; Szkuta et al. 2017; Pilli et al. 2013).

DNA collection methods and techniques vary for DNA recovery rates as does the expected amount of DNA yield from different tissue and fluid sources (Tan and Yiap 2009). The estimated DNA yield from various tissue sources is the following: blood

(200 uL, 4–12 ug), cultured cells (5 million, 15–20 ug), liver (25 mg, 10–30 ug), heart (25 mg, 5–10 ug), and spleen (10 mg, 5–30 ug) (www.qiagen.com). The quantity of DNA recovered from biological evidence is affected by the environment that it is exposed to, and therefore, a range of DNA yield is presented in most forensic studies. The estimated DNA content for forensic biological samples in liquid blood is 20–40 ug/mL, in liquid semen is 150–300 ug/mL, and in liquid saliva is 1–10 ug/mL (Lee and Ladd 2001). Collection method recovery rates vary but include the following estimated percent recovery for epithelial cells from polyester blend fabric: cotton swab (80%), blotting paper (55%), foam-tipped swab (50%), scotch tape (45%), gauze (20%), fingerprint lift tape (15%), Whatman filter paper (15%), Sirchie lift tape (10%), duct tape (5%), positive nylon membrane (1%), and Post-it Note (0%) (https://www.ncjrs.gov/pdffiles1/nij/grants/236826.pdf). Since the method of recovery of DNA molecules varies from forensic science laboratory to laboratory, there may be some level of error introduced by the sampling method or subsequent processing that could lead to suboptimal or no DNA results dependent on the facility or analyst preference (Garvin et al. 2013; Hebda et al. 2014; Singh et al. 2018).

Case documentation. Submission form(s) are included in the case folder to document the requesting agency evidence submissions for testing. On occasion, the incorrect test can be requested for the type of evidence, and one quality control step in both standard DNA and postconviction DNA testing is to evaluate the evidence, the test requests, and the unsubmitted items for potential error. If evidence was never submitted for testing but could have been probative and relevant, there may be the opportunity to reevaluate the case circumstances. Other quality control measures include ensuring pages are labeled with the case number, initials, and date; photographs and diagrams are included and labeled correctly; the proper worksheets are used and filled out completely; and communication and correspondence forms are included in the case folder. For each and every step of the DNA testing process, there should be a quality worksheet that is signed and dated and records the process for the case folder. Peer review internal to the forensic laboratory where analysts review each other's case documentation and supervisors and quality managers also review documentation is designed as a quality control measure to reduce sloppy and incomplete documentation or typographical errors.

Reagent preparation. Quality and technical manuals are required in forensic science laboratories to show analysts how reagents are to be made and to reduce variability in preparation techniques. Receipt of reagents and chemicals by the forensic biology section of the laboratory is documented in a chemical log. The tracking of reagent quality includes catalogue numbers, manufacturer lot numbers, suppliers, and testing of new reagents with old reagent results to ensure that a comparable result is achieved. Once the reagent has been tested and is appropriate for laboratory use, it can be implemented in the general preparation of forensic test reagents thereafter. Expiration dates are checked regularly as well (e.g., https://www1.nyc.gov/assets/ocme/downloads/pdf/technical-manuals/qaqc-procedures-manual/reagents.pdf).

DNA extraction. The DNA extraction process requires either the manufacturer of the kit or the forensic laboratory (or both) quality controls the reagents used to purify

DNA. All of the reagents and disposable plastics are required to be of molecular biology grade quality indicating they are DNase, RNase, and protease-free (e.g., https://www.mt.com/dam/RAININ/PDFs/TechPapers/test_protocol.pdf). These enzymes, if present, will damage the recovery of the DNA and, therefore, must be removed from all surfaces, disposable plastics, and reagents to ensure DNA of high quality and purity is recovered from the biological evidence. Even the water used in reagent preparation must be of molecular biology grade to be quality enough to use in the DNA extraction process. DNA is a large polymeric molecule that is too large to enter the cell membrane. In order to utilize external DNA, some bacteria secrete DNases (enzymes that biodegrade DNA by hydrolysis) outside of the cell to digest DNA into nucleotides. The nucleotides can then move into the bacterial cell membrane via transport proteins. The bacteria use nucleotides to make nucleic acids and as a source of nitrogen, phosphate, and carbon (https://microbiologyinfo.com/deoxyribonuclease-dnase-test/). In addition, a positive DNA sample of known quality and identity is used as a calibration control or reference standard for the DNA extraction process. A negative reagent control is processed simultaneously with the samples and the positive control to test the purity of the reagents and indicate no contaminant DNA is present in the reagents (Morton and Collins 1995). The purpose of DNA extraction is to chemically remove any inhibitory contaminants in the evidentiary sample and purify the DNA molecules away from other cellular components such as proteins, lipids, and carbohydrates so that the template is available for PCR replication.

DNA quantitation. This step is to determine the yield of the DNA from the evidentiary sample (Arya et al. 2005). qPCR assays vary, but in general they are a two-step PCR method that amplifies DNA target loci and includes a fluorescent dye accumulation step to visualize amplification in real time. This allows the analyst to determine both the quantity and quality of the DNA sample prior to generating a full human DNA profile using STR analysis kits and a three-step PCR cycling process called end point PCR. This step requires a manufacturer certification for qPCR kit reagents as well as internal control samples to the assay to check for proper functionality. Positive and negative PCR amplification controls for qPCR assays are commonly included in the kits or can be purchased separately. Human DNA of known concentration is purchasable from molecular biology suppliers and can be purified DNA from American Type Culture Collection (ATCC) certified human cell lines (www.atcc.org); negative controls can be purchased as purified DNA from other animal and plant species to test for PCR amplification specificity. A negative reagent control is performed for the assay by addition of kit reagents to a sample well minus the DNA template to test for purity and no contamination of reagents. In addition, synthetic DNA (internal positive control, IPC) is added to the reagents in the kit as a PCR amplification and pipetting control to indicate that specific DNA template is able to be replicated with the test reagents regardless of whether the evidentiary DNA is amplifiable. The calibration of reagents across the plate is simultaneously performed with the synthetic DNA as the results should be comparable from sample well to sample well (Ewing et al. 2016; Swango et al. 2007; Raymaekers et al. 2009).

DNA amplification. The PCR process is a molecular biology technique to replicate DNA synthetically in a sterile tube. The components of a PCR reaction include buffer, free nucleotides, short complementary PCR primers to the region of interest for amplification, purified genomic DNA template, and magnesium chloride as a cofactor for the Taq polymerase enzyme and Taq polymerase itself. In the PCR process, the DNA double helix is denatured by heat, the PCR primers bind to the complementary regions during the annealing step, and then the final new strand synthesis step occurs. This three step thermal cycling process continues for 25–32 cycles until millions of copies of the original target sequence are represented. The specificity of primer binding is due to DNA sequence homology on the template strand; mutations can result in primer mispairing or absence of replication (Zhu et al. 2020; Green and Sambrook 2019). STR analysis kit details and the history of their development are reviewed in a later section in this chapter.

PCR amplification of genetic loci is performed by using manufacturer kits. The evolution of these kits to those with expanded multiplexes has not changed the quality control measures. A positive human DNA sample is included in each kit to verify the kit functions properly, and the positive control genotype is consistent from run to run. A negative control is recommended as a "blank" sample well containing reagents but lacking DNA template and, therefore, should produce no DNA profile. PCR replication can be difficult to achieve with low quantity and low quality DNA templates due to variability in template sampling and preferential amplification of higher-quantity target sequences. This results in stochastic effects of the minor component in low level samples and minor components in a mixture (Miller Coyle 2015). Internal validation studies of new DNA typing kits are critical to show that the manufacturer claims are true regarding efficiency and accuracy. The internal validation studies are performed "in-house" in the current laboratory setting to determine how it performs in the hands of the analyst. Validation studies include performance checks on method reliability, reproducibility, sensitivity, and specificity. For each new kit that is developed, forensic laboratories need to validate and peer-review validation studies to show forensic community consensus as part of meeting court admissibility criteria (Ewing et al. 2016; Zhou et al. 2016; Gopinath et al. 2016).

Capillary electrophoresis. Capillary electrophoresis is a semi-automated method for selective size separation of DNA fragments based on size and charge. DNA is heat denatured and placed in purified deionized formamide to maintain a single-strand configuration to the DNA fragment. Only one strand of DNA has been fluorescently labeled in the PCR amplification process, so that is the strand that can be visualized in the capillary electrophoresis step. A liquid polymer is used as the sieving medium to separate the DNA fragments into a fluorescently labeled and visible "barcode" that fills the glass capillaries in the capillary electrophoresis system. Allelic ladders and internal lane size standards control for the accurate sizing of DNA fragments by providing systematic sized patterns to compare the unknown DNA fragments against. Capillaries are sensitive to heat and mobility of DNA fragments can vary per capillary; thus, an internal lane size standard (or comigration control) corrects for mobility shift. Beyond the use of these standards, data may not be reproducible from PCR reaction to PCR reaction due to slight

differences in efficiency rates at each locus and due to template sampling differences in the electrokinetic injection step (Krivácsy et al. 1999; Opekar et al. 2016).

Data interpretation. A technical review is performed to check that the reports and conclusions are correct. The review includes that proper controls were used; the controls gave appropriate results, and the conclusions are in agreement with data. The allele calls and statistical calculations are then verified by a second, qualified individual. Supervisor oversight includes review of a percentage of the casework that is produced at the forensic science laboratory. The challenges to DNA reside in complex DNA mixture interpretation, and there is considerable variability in data interpretation from forensic laboratory to forensic laboratory (Butler et al. 2018; Buckleton et al. 2018). Each laboratory has set its own policy to follow for data interpretation above and below an analytical instrument threshold, for establishing the true number of contributors to a DNA mixture, and for assessment of contaminant alleles.

Analyst bias has been recognized as a subconscious or conscious attempt to include an individual in a DNA result simply because an individual has been identified as a candidate suspect. Bias is represented by selective inclusion and/or nondisclosure of other candidates observed in the data either due to error, intent, or laboratory policy. Most importantly for trial, scientific accuracy is needed to provide effective interpretation of the data and effective counsel. Improvements in data interpretation with mixtures have been made with probabilistic genotyping software analysis systems (e.g., TrueAllele, STRmix) (Greenspoon et al. 2015; Perlin et al. 2015; Bauer et al. 2020; Perlin et al. 2011; Bright et al. 2019; Buckleton et al. 2019; Moretti et al. 2017). While computer software can assist in making statistical inferences, they do also need to be fully understood to recognize the benefits, limitations, and inherent error rates associated with each program. Precision is the ability to obtain the same result every time a test is run; however, the test result may not be accurate. So the ability to obtain both a precise and scientifically accurate result is the desired objective, and finding the optimal method is key while realizing technology may go through several iterations before arriving to the point of optimum.

A case in point, here, is the use of probabilistic genotyping software that is designed to eliminate confirmation bias when analyzing data by eye; however, the analyst may have bias in establishing the number of contributors to specify for the software analysis: a situation of concern. Less has been written about confirmation bias in DNA analysis; however, some good forensic studies have been done (Skellern 2015; Mattijssen et al. 2016; Dror 2012, 2015; Dror et al. 2015; Nakhaeizadeh et al. 2014; Brauner 2012; Dror and Hampikian 2011). Certain aspects such as the effect of adjusting the analytical threshold (an invisible line above which alleles are reported and below which alleles are not reported) and determining the effect on reporting number of contributors appear absent in many forensic laboratory validation studies and would affect the interpretation of whether or not an individual may be included in a DNA mixture (Kirby et al. 2017). The setting of the analytical threshold is established in historical policy when forensic laboratories first established their DNA units; however, this will be worth revisiting as probabilistic

genotyping software programs are installed and validated now, with new and better technology.

An article by Gill et al. (2006) clarifies for the DNA analyst factors for identifying the number of contributors in the mixture. "The number of alleles observed per locus, circumstances of the case, and the possibility of related contributors go into deciding how many contributors to condition on." The allele counting method (assuming a heterozygote as donor) yields a minimum estimate of number of contributors per genetic locus. Scientifically speaking, the maximum alleles observed at any given locus should yield the minimum estimated of detected number of contributors to the DNA mixture. Given the allele counting method, this article goes further to explain the issue of number of contributors and how that assessment relates to determining a likelihood ratio (LR) for probabilistic genotyping. "It is not always easy to specify hypotheses in complex cases where multiple perpetrators or victims may be present. The DNA result itself may indicate that different explanations may be possible. Furthermore, it is possible that Hp (prosecutor hypothesis) and Hd (defense hypothesis) could be very different from each other. For example, under Hp we might consider (victim and suspect); whereas with Hd we might examine more complex scenarios with 3 contributors (3 unknowns, $U_0 + U_1 + U_2$). There is a common misconception that the number of contributors under Hp and Hd should be the same. There is no requirement for this." The article suggests "the smallest number of unknown contributors needed to explain the evidence are usually the ones to maximize the respective likelihoods." In courtroom testimony, often a DNA analyst will state that a DNA result is conservative ("an assignment for the weight of the evidence that is believed to favor the defense"); however, the maximum number of contributors may not be assigned correctly to the DNA mixture due to the policy of consensus profiling.

Evidence of additional contributors in replicate PCR amplifications, alleles detected below the analytical threshold, spurious alleles, contamination, and detection of minor DNA elements by differential electrokinetic injection all amount to the same thing: additional alleles are present that cannot be accounted for by the standards submitted and should be disclosed in reports, testimony, and statistics. The effect of including the maximum possible contributors based on allele counts in any one of the PCR replicates (duplicates or triplicates) is to, first, acknowledge the scientific observation and, second, to generate the best possible probability estimate when using probabilistic genotyping methods. Historically, forensic laboratories were greatly concerned over reporting out trace levels of DNA contamination. The development of new DNA standards has been recently completed to address the validation of DNA probabilistic genotyping software systems and for DNA mixture interpretation to improve consistency in results determination from forensic science laboratory to forensic science laboratory. While PCR amplification and capillary electrophoresis artifacts and parameters still present a problem for data interpretation to DNA analysts, the new standards should be an aid to increase uniformity in DNA interpretation of mixtures from biological evidence.

History of STR analysis. Approximately three million nucleotide bases (noncoding regions) with multiple copies of short tandem repeat sequences construct the

DNA backbone (e.g., CAGTCAGTCAGT; three repeats). These regions are called "variable number of short tandem repeats (VNTRs)." If a sufficient number of STR loci are tested (profiled), then the evidence of a person's identity (and unique STR identifier) is enhanced because the likelihood of two unrelated people having the same number of repeated sequences in these regions becomes vanishingly small (https://nij.ojp.gov/topics/articles/what-str-analysis).

The original form of STR analysis used single PCR reactions to test per locus information, multiplexing to combine loci rapidly followed. The first multiplexes had three or four loci per PCR reaction. The primary commercial suppliers for the early systems were (a) Promega Corporation (Madison, WI) that used a silver staining method to visualize loci (CTT: CSF1PO, TPOX, TH01; CTTV: CSF1PO, TPOX, TH01, VWA; FFV: F13A1, FESFPS, VWA; FFFL: F13A1, FESFPS, F13B, LPL; GammaSTRTM: D16S539, D7S820, D13S317, D5S818) and (b) Applied Biosystems (Foster City, CA) that used a fluorescent dye technology AmpF*l*STR Green I (Amelogenin, TH01, TPOX, CSF1PO and AmpF*l*STR Blue: D3S1358, VWA, FGA (https://strbase.nist.gov/multiplx.htm)).

Expanded multiplexes quickly followed with the Promega Powerplex series: PowerPlex ES: D3S1358, TH01, D21S11, D18S51, SE33, Amelogenin, VWA, D8S1179, FGA; PowerPlex 16 HS: D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, Amelogenin, VWA, D8S1179, TPOX, FGA; PowerPlex 18D: D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, Amelogenin, VWA, D8S1179, TPOX, FGA, D19S433, D2S1338; PowerPlex ESX 16: Amelogenin, D3S1358, TH01, D21S11, D18S51, D10S1248, D1S1656, D2S1338, D16S539, D22S1045, VWA, D8S1179, FGA, D2S441, D12S391, D19S433; PowerPlex ESX 17: Amelogenin, D3S1358, TH01, D21S11, D18S51, D10S1248, D1S1656, D2S1338, D16S539, D22S1045, VWA, D8S1179, FGA, D2S441, D12S391, D19S433, SE33; PowerPlex ESI 16: Amelogenin, D3S1358, D19S433, D2S1338, D22S1045, D16S539, D18S51, D1S1656, D10S1248, D2S441, TH01, VWA, D21S11, D12S391, D8S1179, FGA; PowerPlex ESI 17: Amelogenin, D3S1358, D19S433, D2S1338, D22S1045, D16S539, D18S51, D1S1656, D10S1248, D2S441, TH01, VWA, D21S11, D12S391, D8S1179, FGA, SE33; and PowerPlex 21: Amelogenin, D3S1358, D1S1656, D6S1043, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, VWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433, FGA. Also, Applied Biosystems "filer" series: AmpF*l*STR Profiler Plus: D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820; AmpF*l*STR Profiler Plus ID: D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820; AmpF*l*STR COfiler: D3S1358, D16S539, Amelogenin, TH01, TPOX, CSF1PO, D7S820; AmpF*l*STR Sinofiler (available only in China): D8S1179, D21S11, D7S820, CSF1PO, D3S1358, D5S818, D13S317, D16S539, D2S1338, D19S433, VWA, D12S391, D18S51, Amelogenin, D6S1043, FGA; AmpF*l*STR Profiler: D3S1358, VWA, FGA, Amelogenin, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820; AmpF*l*STR SEfiler: D3S1358, VWA, D16S539, D2S1338, Amelogenin, D8S1179, SE33, D19S433, TH01, FGA,

D21S11, D18S51; and AmpF/STR SEfiler Plus: D3S1358, VWA, D16S539, D2S1338, Amelogenin, D8S1179, SE33, D19S433, TH01, FGA, D21S11, D18S51. These kits were a series integrated for the Combined DNA Indexing System (CODIS) core 13 loci required for comparisons with the National DNA Index System (NDIS) DNA database of convicted offender samples.

The latest generation kits are large megaplexes and specialty application kits. Included in the Promega series are PowerPlex 16: D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, Amelogenin, VWA, D8S1179, TPOX, FGA; PowerPlex Fusion (includes 22 loci, amelogenin for gender identification, and a Y chromosome locus): Amelogenin, D3S1358, D1S1656, D2S441, D10S1248, D13S317, Penta E, D16S539, D18S51, D2S1338, CSF1PO, Penta D, TH01, VWA, D21S11, D7S820, D5S818, TPOX, DYS391, D8S1179, D12S391, D19S433, FGA, D22S1045; and the Y chromosome specific STR kits: PowerPlex Y: DYS391, DYS389I, DYS439, DYS389II, DYS438, DYS437, DYS19, DYS392, DYS393, DYS390, DYS385a/b; and PowerPlex Y23: DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456, Y_GATA_H4. Applied Biosystems megaplexes include AmpF/STR Identifiler: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, Amelogenin, D5S818, FGA; AmpF/STR Identifiler Direct: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, Amelogenin, D5S818, FGA; AmpF/STR Identifiler Plus: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, Amelogenin, D5S818, FGA; AmpF/STR NGM: D10S1248, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D22S1045, D19S433, TH01, FGA, D2S441, D3S1358, D1S1656, D12S391; AmpF/STR NGM SElect: D10S1248, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D22S1045, D19S433, TH01, FGA, D2S441, D3S1358, D1S1656, D12S391, SE33; AmpF/STR GlobalFiler: D3S1358, VWA, D16S539, CSF1PO, TPOX, Yindel, Amelogenin, D8S1179, D21S11, D18S51, DYS391, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391, D2S1338; AmpF/STR VeriFiler: D10S1248, D1S1656, Amelogenin, D2S1338, D22S1045, D19S433, TH01, D2S441, D6S1043, D12S391; AmpF/STR MiniFiler: D13S317, D7S820, Amelogenin, D2S1338, D21S11, D16S539, D18S51, CSF1PO, FGA; AmpF/STR Yfiler: DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y_GATA_H4, DYS437, DYS438, DYS448. Promega Corporation and Applied Biosystems are no longer the only STR kit suppliers; manufacturer's now include QIAGEN N.V. (Venlo, Netherlands) and Biotype (Dresden, Germany) (https://strbase.nist.gov/multiplx.htm).

The STRBase website compiles relevant information on current scientific literature for autosomal and Y chromosome STR marker systems, fact sheets from various

kit manufacturers, statistics on tri-allele patterns, mutation rates observed for each chromosome locus, the sequence information for each short tandem repeat for motifs/nucleotide base combinations, genomic map positions, and allele size ranges (https://strbase.nist.gov). This website of compiled useful STR marker systems has been available since 1997. Each chromosome locus is fully described and links to the original publications, and authors are conveniently provided. PCR primer sequences for each kit are published, and concordant studies comparing various kit results have been compared for consistency between genotyping STR kit manufacturers. Where non-concordance was detected, NIST worked with the manufacturers to refine the science so that all data sets would be comparable for uploading into the NDIS DNA database system.

Specialty STR kits have been optimized for the detection of alleles from highly degraded DNA samples. The ability to type degraded DNA specimens was improved by redesigning the STR marker amplicons so that a smaller-sized polymerase chain reaction (PCR) product was created at each locus. This kit was called the AmpF*l*STR MiniFiler PCR Amplification Kit. The kit contains reagents for the amplification of eight miniSTRs which are the largest-sized loci in the AmpF*l*STR Identifiler PCR Amplification Kit (D7S820, D13S317, D16S539, D21S11, D2S1338, D18S51, CSF1PO, and FGA) (Mulero et al. 2008; Bright et al. 2011). The MiniFiler kit was validated for casework use (Luce et al. 2009; Hill et al. 2007) and has been used on degraded skeletal remains and cigarette butts (Ip et al. 2014) and for war remains identifications successfully (Marjanović et al. 2009). A Chinese forensic STR kit called Sinofiler was specifically released for the Chinese forensic science laboratories (Shuqin Huang et al. 2010). The kit includes the STR loci: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, D13S317, D16S539, D2S1338, D19S433, vWA, D18S51, D6S1043, D12S391, D5S818, and FGA. After evaluation for the Chinese Han population, the kit was deemed suitable. There was no statistical departure from expectation of Hardy–Weinberg Equilibrium (HWE) for all loci but D6S1043. There was no linkage disequilibrium in all pairs of loci examined. This kit was validated for this particular population group and country (Liu et al. 2014).

Future directions for STR technology. Two new trends are present in STR analysis. The first is to revive kinship analysis using DNA test methods and place it in a larger category called forensic genealogy. Interesting genealogy studies include a study of surnames and the founder male lineages (27 total) for the Old Order Amish males in Pennsylvania (Pollin et al. 2008). This study could account for 98% of the male lineages associated with the Lancaster area. Also, surname analysis and YSTR tests have been successfully applied for predictive value in Chinese Han populations to the level of 65% accuracy and increases to 80% accuracy for top four surnames (Shi et al. 2018). This approach can be useful as a "dragnet" concept where surnames can be predicted from Y chromosome haplotypes. Geography and restricted populations fare better with this approach as Y chromosome ancestral "footprints" are detectable in some populations to the level that they stretch back to 500 AD to an ancestral founder population; others are more difficult to use predictively but could still have usefulness for DNA investigative leads (Whiting and Miller Coyle 2019).

Familial DNA search methodology is different than ancestral surname searches since surname searches rely on the inheritance of paternal lineages by Y chromosome ancestry. Y chromosome ancestry is not one hundred percent accurate as there are some common Y chromosome haplotypes that are not apparently traceable back in historical records to the same family group. While seeking ancestors to your family genealogical tree can be useful with this method, it still requires some verification by investigation of property records, birth, marriage, and death certificates and other genealogical records to confirm a likely family relationship. The concept of a coincidental match in haplotype analysis using STR markers has to be considered as a fortuitous occurrence or the association was not documented back many generations ago in the family tree (https://www.ancestry.com). The coincidental match can be challenging to interpret using YSTR technology, but the DNA method is still highly valuable as an exclusionary tool.

Familial search, however, is a search of DNA databases by law enforcement for similar but not exact DNA profiles indicating a genetic relative that can provide a possible investigative lead in the case (https://criminal.findlaw.com/criminal-rights/familial-dna-searches.html). This DNA database search technique was able to identify through a DNA relative and investigation the Golden State Killer in 2018, approximately 30 years after his homicides began. This was the only effective manner in which he was identified and apprehended. Traditional DNA searches are performed to identify an exact high stringency DNA match, but with familial search techniques, partial matching at reduced stringency is permitted in the software-driven search to provide a candidate list of possible DNA matches. These individuals are then further evaluated and traditionally investigated to establish their possible role in the crime. Most state and federal authorities continuously collect known reference DNA samples from convicted offenders that are continuously uploaded into searchable DNA databases. If there is no familial search match, it may be due to the fact that there are no genetic relatives apprehended and convicted; therefore, there are no leads in that particular database. Law enforcement, however, does have the opportunity to access private ancestry-related DNA databases and uses those reference populations to also search for candidate leads. There are some justifiable concerns by civil liberties advocates that criticize familial DNA search technology as an invasion of personal privacy. The argument is that familial DNA searching affects the privacy of unconvicted genetic relatives, who are innocent of the crime, and this violates the Fourth Amendment in the Unites States Constitution which protects individuals against unreasonable searches. For this reason, some states such as Maryland have banned the use of familial search technology from being used; others like California have passed state legislation to allow familial searching in violent crime with policies in place for correct application and usage.

The second trend in human identification technology is to update and revise new DNA standards to fit with current technology and quality management strategies. The newest policies for year 2020 can be found at this website: https://www.nist.gov/news-events/news/2020/05/two-new-forensic-dna-standards-added-osac-registry. The Organization of Scientific Area Committees or OSAC is an organization with professional and scientific members that have expertise in 25 different forensic disciplines and

have expertise in the areas of peer-reviewed scientific research, measurement science, statistics, and legal policy. The role of OSAC is to build the foundational science of DNA methods and promote the use of DNA standards for evaluating and implementing new technology.

In 2015, OSAC began by drafting new DNA standards that were then submitted to the Academy Standards Board (ASB) of the American Academy of Forensic Sciences (AAFS) and evaluated again by OSAC prior to posting on the registry. The new standards are (1) ANSI/ASB Standard 020, Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol and (2) ANSI/ASB Standard 040, Standard for Forensic DNA Interpretation and Comparison Protocols. The verification of a laboratory's mixture interpretation protocol must demonstrate that a laboratory's protocols produce consistent and reliable conclusions with DNA samples different from the ones used in the initial validation studies. This implies that a test period is required on adjudicated casework samples to validate that the research samples and the authentic casework samples are similar and function appropriately with the new protocol. The new standards also limit forensic science laboratories and prevent them from interpreting DNA mixtures that exceed validated methods especially regarding use on increasingly complex mixtures with additional contributors. The majority of current DNA units in forensic science laboratories still only interpret three person DNA mixtures from biological evidence. The new OSAC DNA standards are still complementary to the FBI's DNA Quality Assurance Standards. They also build upon the Scientific Working Group on DNA Analysis Methods work, the SWGDAM guidelines. There still may be forensic science laboratory compliance issues as implementation of the OSAC DNA standards remains voluntary. Many forensic science laboratories are already working toward proper implementation, however, as it is best practice to follow the latest edition of DNA standards by OSAC to promote quality and scientific integrity in laboratory practices for forensic human DNA identification methods.

Conclusions

Quality control preserves the integrity of the DNA test result in STR analyses. Without reagent, equipment, and method checks and balances, there may be "protocol drift" that allows the analyst to deviate from the written document or method protocol thus creating variance from case to case. Part of public confidence building in forensic science laboratories is to convince the public that the scientific method is valid, accurate, and consistent from case to case regardless of who is on trial. Some of the most spectacular forensic fraudulent cases stem from this problem, where laboratory administration review and oversight is not sufficient to catch incorrect documentation or inconsistent DNA results. The Swecker and Wolf report (http://wwwcache.wral.com/asset/news/state/nccapitol/2016/09/07/15994768/254822-Swecker_Report.pdf) documents the type of quality assurance and quality control issues that can arise in presumptive and confirmatory blood identification tests and

also created discrepancies in blood serological and DNA reports that became evident in the courtroom to defense counsel and science experts alike. The exhaustive fifteen thousand or more case reviews by outside experts and a three judge panel identified two hundred and thirty cases that resulted in retrials, acquittals, and exonerations simply based on poorly written and poorly enforced protocols. A key element to this issue was the fact that "inconclusive" was not written into the blood identification procedure as a test result option; when forced to choose, sometimes the analyst selected the incorrect result. The Inspector General of New York State issued an executive summary of a leading forensic science serology and DNA unit as well that found fraud in sexual assault kit evaluations for semen evidence. Along with poor quality screening of evidence by a fraudulent serologist, the DNA unit was inconsistent in the manner in which DNA mixtures were interpreted (https://ig.ny.gov/sites/g/files/oee571/files/2016-12/OCMEFinalReport.pdf). The manner in which the analyst decided the true number of contributors (NOC) to the DNA mixture was at issue; some were calculating NOC based on alleles called by the analysis software; others were including evidence of additional contributors below the analytical threshold.

All scientific methods have an error rate, and a process must be in place to guarantee that reagents, equipment, and methods are functioning properly to reduce the error rate. The four main areas highlighted in this chapter (DNA collection methods, polymerase chain reaction (PCR) factors, electrokinetic injection considerations, and DNA data interpretation) are detailed to indicate there are checks and balances to the procedures that are built into the forensic science laboratory procedures. Forensic science laboratories are also provided the guidance documents by the DNA Advisory Board and the Organization of Scientific Area Committees designed to enhance DNA test result consistency. Still, there are some criticisms to the quality control process in forensic DNA testing. "Negative controls also can't rule out contamination of individual samples" (http://www.injusticeinperugia.org/viewfromwilmington.html). This statement is true. Most of the quality control in the DNA unit is designed to detect gross contamination events or reagent and equipment failures, but it is possible to have a single independent tube become contaminated, and it may go undetected in the surveillance system of the quality manager. However, there are internal forensic science laboratory DNA databases of the laboratory personnel that DNA results are screened against prior to providing a DNA report to a submitting agency to attempt to screen out accidental DNA contamination during the laboratory processing steps. Crime scene personnel and emergency services workers must be subpoenaed to provide an elimination known reference DNA sample for comparison to casework samples.

In addition to OSAC standards, enforcement by laboratory administration, and the personal ethics of the individual analyst, laboratory auditing is another method for reviewing the quality of a forensic laboratories work product. The American Society of Crime Laboratory Directors Laboratory Accreditation Board (ASCLAD-LAB) (www.ascld.org) is a not-for-profit professional society of crime laboratory directors and forensic science managers whose function is "to foster professional interests, assist the development of laboratory management principles and

techniques; acquire, preserve, and disseminate forensic based information; maintain and improve communication among crime laboratory directors; and to promote, encourage, and maintain the highest standards of practice in the field." ASCLD inspects forensic laboratories and accredits their practices and facilities every 5 years to review and maintain the integrity of the scientific processes used to evaluate forensic evidence in criminal casework. The ASCLD auditing procedure reviews the forensic laboratory process as a whole and makes recommendations for quality management improvement; however, the reviews are randomized and based on whatever becomes evident during the audit as the two week process can only provide a "spot check" on the complexities and total volume of casework that is processed in all of the units of the forensic science laboratory.

Many forensic science laboratories have also applied for and been granted ISO17025 accreditation status which is an external auditing process that is provided to all clinical and manufacturing laboratories. ISO/IEC 17025 standardization enables "laboratories to show that they operate competently and generate valid results, thereby promoting confidence in their work both nationally and around the world. It also helps facilitate cooperation between laboratories and other bodies by generating wider acceptance of results between countries. Test reports and certificates can be accepted from one country to another without the need for further testing, which, in turn, improves international trade" (https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html). These claims are true, and the need for global standardization is very helpful for sharing DNA data between country borders. If the DNA data is of high quality and can be shared, increased casework can be solved and more missing persons identified through quality controlled global DNA databases and STR analysis methods.

References

Aparna R, Shanti Iyer R (2020) Tears and eyewear in forensic investigation – a review. Forensic Sci Int 306:110055. https://doi.org/10.1016/j.forsciint.2019.110055

Arya M, Shergill IS, Williamson M, Gommersall L, Arya N, Patel HR (2005) Basic principles of real-time quantitative PCR. Expert Rev Mol Diagn 5(2):209–219. https://doi.org/10.1586/14737159.5.2.209

Bauer DW, Butt N, Hornyak JM, Perlin MW (2020) Validating TrueAllele® interpretation of DNA mixtures containing up to ten unknown contributors. J Forensic Sci 65(2):380–398. https://doi.org/10.1111/1556-4029.14204

Brauner P (2012) RE. Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice 52 (2):131; author reply 132. https://doi.org/10.1016/j.scijus.2012.04.001

Bright JA, Huizing E, Melia L, Buckleton J (2011) Determination of the variables affecting mixed MiniFiler™ DNA profiles. Forensic Sci Int Genet 5(5):381–385. https://doi.org/10.1016/j.fsigen.2010.08.006

Bright JA, Cheng K, Kerr Z, McGovern C, Kelly H, Moretti TR, Smith MA, Bieber FR, Budowle B, Coble MD, Alghafri R, Allen PS, Barber A, Beamer V, Buettner C, Russell M, Gehrig C, Hicks T, Charak J, Cheong-Wing K, Ciecko A, Davis CT, Donley M, Pedersen N, Gartside B, Granger D, Greer-Ritzheimer M, Reisinger E, Kennedy J, Grammer E, Kaplan M, Hansen D, Larsen HJ, Laureano A, Li C, Lien E, Lindberg E, Kelly C, Mallinder B, Malsom S, Yacovone-Margetts A, McWhorter A, Prajapati SM, Powell T, Shutler G, Stevenson K,

Stonehouse AR, Smith L, Murakami J, Halsing E, Wright D, Clark L, Taylor DA, Buckleton J (2019) STRmix™ collaborative exercise on DNA mixture interpretation. Forensic Sci Int Genet 40:1–8. https://doi.org/10.1016/j.fsigen.2019.01.006

Buckleton JS, Bright JA, Cheng K, Budowle B, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX13): a modern analysis. Forensic Sci Int Genet 37:172–179. https://doi.org/10.1016/j.fsigen.2018.08.014

Buckleton JS, Bright JA, Gittelson S, Moretti TR, Onorato AJ, Bieber FR, Budowle B, Taylor DA (2019) The probabilistic genotyping software STRmix: utility and evidence for its validity. J Forensic Sci 64(2):393–405. https://doi.org/10.1111/1556-4029.13898

Butler JM, Kline MC, Coble MD (2018) NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): variation observed and lessons learned. Forensic Sci Int Genet 37:81–94. https://doi.org/10.1016/j.fsigen.2018.07.024

Dror I (2012) Letter to the editor–combating bias: the next step in fighting cognitive and psychological contamination. J Forensic Sci 57(1):276–277. https://doi.org/10.1111/j.1556-4029.2011.01940.x

Dror IE (2015) Cognitive neuroscience in forensic science: understanding and utilizing the human element. Philos Trans R Soc Lond Ser B Biol Sci 370(1674). https://doi.org/10.1098/rstb.2014.0255

Dror IE, Hampikian G (2011) Subjectivity and bias in forensic DNA mixture interpretation. Sci Justice 51(4):204–208. https://doi.org/10.1016/j.scijus.2011.08.004

Dror IE, Thompson WC, Meissner CA, Kornfield I, Krane D, Saks M, Risinger M (2015) Letter to the editor- context management toolbox: a linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. J Forensic Sci 60(4):1111–1112. https://doi.org/10.1111/1556-4029.12805

Ewing MM, Thompson JM, McLaren RS, Purpero VM, Thomas KJ, Dobrowski PA, DeGroot GA, Romsos EL, Storts DR (2016) Human DNA quantification and sample quality assessment: developmental validation of the PowerQuant(®) system. Forensic Sci Int Genet 23:166–177. https://doi.org/10.1016/j.fsigen.2016.04.007

Garvin AM, Holzinger R, Berner F, Krebs W, Hostettler B, Lardi E, Hertli C, Quartermaine R, Stamm C (2013) The forensiX evidence collection tube and its impact on DNA preservation and recovery. Biomed Res Int 2013:105797. https://doi.org/10.1155/2013/105797

Gill P, Brenner CH, Buckleton JS, Carracedo A, Krawczak M, Mayr WR, Morling N, Prinz M, Schneider PM, Weir BS (2006) DNA Commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures. Forensic Sci Int 160(2–3):90–101

Gopinath S, Zhong C, Nguyen V, Ge J, Lagacé RE, Short ML, Mulero JJ (2016) Developmental validation of the Yfiler(®) plus PCR amplification kit: an enhanced Y-STR multiplex for casework and database applications. Forensic Sci Int Genet 24:164–175. https://doi.org/10.1016/j.fsigen.2016.07.006

Goray M, van Oorschot RA, Mitchell JR (2012) DNA transfer within forensic exhibit packaging: potential for DNA loss and relocation. Forensic Sci Int Genet 6(2):158–166. https://doi.org/10.1016/j.fsigen.2011.03.013

Green MR, Sambrook J (2019) Polymerase chain reaction. Cold Spring Harb Protoc 2019(6). https://doi.org/10.1101/pdb.top095109

Greenspoon SA, Schiermeier-Wood L, Jenkins BC (2015) Establishing the limits of TrueAllele® casework: a validation study. J Forensic Sci 60(5):1263–1276. https://doi.org/10.1111/1556-4029.12810

Hebda LM, Doran AE, Foran DR (2014) Collecting and analyzing DNA evidence from fingernails: a comparative study. J Forensic Sci 59(5):1343–1350. https://doi.org/10.1111/1556-4029.12465

Hill CR, Kline MC, Mulero JJ, Lagacé RE, Chang CW, Hennessy LK, Butler JM (2007) Concordance study between the AmpF/STR MiniFiler PCR amplification kit and conventional STR typing kits. J Forensic Sci 52(4):870–873. https://doi.org/10.1111/j.1556-4029.2007.00491.x

Ip SC, Lin SW, Li C, Lai KM (2014) Forensic DNA typing strategy of degraded DNA on discarded cigarette ends using the AmpFℓSTR® Identifiler®, Identifiler® Plus and MiniFiler™ PCR amplification kits. Sci Justice 54(4):311–315

Kirby A, Chance-Johnson J, Miller Coyle H (2017) The effect of the analytical threshold on the loss and gain of data from single source and mixed DNA samples. Int Res J Comput Sci 4(8):50–55

Krivácsy Z, Gelencsér A, Hlavay J, Kiss G, Sárvári Z (1999) Electrokinetic injection in capillary electrophoresis and its application to the analysis of inorganic compounds. J Chromatogr A 834(1–2):21–44. https://doi.org/10.1016/s0021-9673(98)00861-9

Lee HC, Ladd C (2001) Preservation and collection of biological evidence. Croat Med J 42(3):225–228

Liu QL, Luo H, Zhao H, Huang XL, Cheng JD, Lu DJ (2014) Recombination analysis of autosomal short tandem repeats in Chinese Han families. Electrophoresis 35(6):883–887. https://doi.org/10.1002/elps.201300417

Luce C, Montpetit S, Gangitano D, O'Donnell P (2009) Validation of the AMPF/STR MiniFiler PCR amplification kit for use in forensic casework. J Forensic Sci 54(5):1046–1054. https://doi.org/10.1111/j.1556-4029.2009.01099.x

Marjanović D, Durmić-Pasić A, Kovacević L, Avdić J, Dzehverović M, Haverić S, Ramić J, Kalamujić B, Lukić Bilela L, Skaro V, Projić P, Bajrović K, Drobnic K, Davoren J, Primorac D (2009) Identification of skeletal remains of Communist Armed Forces victims during and after World War II: combined Y-chromosome (STR) and MiniSTR approach. Croat Med J 50(3):296–304. https://doi.org/10.3325/cmj.2009.50.296

Mattijssen EJ, Kerkhoff W, Berger CE, Dror IE, Stoel RD (2016) Implementing context information management in forensic casework: minimizing contextual bias in firearms examination. Sci Justice 56(2):113–122. https://doi.org/10.1016/j.scijus.2015.11.004

Miller Coyle H (2015) Quality control and duplication for concordance in forensic DNA samples: implications for interpretation in mixtures. IRJCS 6(2):16–18

Moretti TR, Just RS, Kehl SC, Willis LE, Buckleton JS, Bright JA, Taylor DA, Onorato AJ (2017) Internal validation of STRmix™ for the interpretation of single source and mixed DNA profiles. Forensic Sci Int Genet 29:126–144. https://doi.org/10.1016/j.fsigen.2017.04.004

Morton NE, Collins AE (1995) Statistical and genetic aspects of quality control for DNA identification. Electrophoresis 16(9):1670–1677. https://doi.org/10.1002/elps.11501601277

Mulero JJ, Chang CW, Lagacé RE, Wang DY, Bas JL, McMahon TP, Hennessy LK (2008) Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA. J Forensic Sci 53(4):838–852. https://doi.org/10.1111/j.1556-4029.2008.00760.x

Nakhaeizadeh S, Dror IE, Morgan RM (2014) Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias. Sci Justice 54(3):208–214. https://doi.org/10.1016/j.scijus.2013.11.003

Opekar F, Nesměrák K, Tůma P (2016) Electrokinetic injection of samples into a short electrophoretic capillary controlled by piezoelectric micropumps. Electrophoresis 37(4):595–600. https://doi.org/10.1002/elps.201500464

Pandeshwar P, Das R (2014) Role of oral fluids in DNA investigations. J Forensic Legal Med 22:45–50. https://doi.org/10.1016/j.jflm.2013.12.007

Perlin MW, Legler MM, Spencer CE, Smith JL, Allan WP, Belrose JL, Duceman BW (2011) Validating TrueAllele® DNA mixture interpretation. J Forensic Sci 56(6):1430–1447. https://doi.org/10.1111/j.1556-4029.2011.01859.x

Perlin MW, Hornyak JM, Sugimoto G, Miller KW (2015) TrueAllele(®) genotype identification on DNA mixtures containing up to five unknown contributors. J Forensic Sci 60(4):857–868. https://doi.org/10.1111/1556-4029.12788

Pilli E, Modi A, Serpico C, Achilli A, Lancioni H, Lippi B, Bertoldi F, Gelichi S, Lari M, Caramelli D (2013) Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains. PLoS One 8(1):e52524. https://doi.org/10.1371/journal.pone.0052524

Pollin TI, McBride DJ, Agarwala R, Schäffer AA, Shuldiner AR, Mitchell BD, O'Connell JR (2008) Investigations of the Y chromosome, male founder structure and YSTR mutation rates in the Old Order Amish. Hum Hered 65(2):91–104. https://doi.org/10.1159/000108941

Raymaekers M, Smets R, Maes B, Cartuyvels R (2009) Checklist for optimization and validation of real-time PCR assays. J Clin Lab Anal 23(3):145–151. https://doi.org/10.1002/jcla.20307

Shi CM, Li C, Ma L, Chi L, Zhao J, Yuan W, Zhou Z, Yan JW, Chen H (2018) Inferring Chinese surnames with Y-STR profiles. Forensic Sci Int Genet 33:66–71. https://doi.org/10.1016/j.fsigen.2017.11.014

Shuqin H, Yunliang Z, Xintang S, Xiaoping L, Hongtao Y (2010) Genetic variation analysis of 15 autosomal STR loci of AmpFℓSTR® Sinofiler™ PCR Amplification Kit in Henan (central China) Han population. Legal Med 12(3):160–161. https://doi.org/10.1016/j.legalmed.2010.02.004

Singh UA, Kumari M, Iyengar S (2018) Method for improving the quality of genomic DNA obtained from minute quantities of tissue and blood samples using Chelex 100 resin. Biol Proced Online 20:12. https://doi.org/10.1186/s12575-018-0077-6

Skellern C (2015) Minimising bias in the forensic evaluation of suspicious paediatric injury. J Forensic Legal Med 34:11–16. https://doi.org/10.1016/j.jflm.2015.05.002

Swango KL, Hudlow WR, Timken MD, Buoncristiani MR (2007) Developmental validation of a multiplex qPCR assay for assessing the quantity and quality of nuclear DNA in forensic samples. Forensic Sci Int 170(1):35–45. https://doi.org/10.1016/j.forsciint.2006.09.002

Szkuta B, Oorschot RAHV, Ballantyne KN (2017) DNA decontamination of fingerprint brushes. Forensic Sci Int 277:41–50. https://doi.org/10.1016/j.forsciint.2017.05.009

Tan SC, Yiap BC (2009) DNA, RNA, and protein extraction: the past and the present. J Biomed Biotechnol 2009:574398. https://doi.org/10.1155/2009/574398

Whiting R, Miller Coyle H (2019) Haplotype analysis for Irish ancestry. Forensic Sci Res. https://doi.org/10.1080/20961790.2019.1639881

Zajac GJM, Fritsche LG, Weinstock JS, Dagenais SL, Lyons RH, Brummett CM, Abecasis GR (2019) Estimation of DNA contamination and its sources in genotyped samples. Genet Epidemiol 43(8):980–995. https://doi.org/10.1002/gepi.22257

Zhou H, Bi G, Zhang C, Liu Y, Chen R, Li F, Mei X, Guo Y, Zheng W (2016) Developmental validation of forensic DNA-STR kits: expressmarker 16+10Y and expressmarker 16+18Y. Forensic Sci Int Genet 24:1–17. https://doi.org/10.1016/j.fsigen.2016.05.011

Zhu H, Zhang H, Xu Y, Laššáková S, Korabečná M, Neužil P (2020) PCR past, present and future. BioTechniques 69(4):317–325. https://doi.org/10.2144/btn-2020-0057

DNA Phenotyping: The Technique of the Future

52

Kamayani Vajpayee and Ritesh Kumar Shukla

Contents

K. Vajpayee (✉)
DNA Fingerprinting Unit, Forensic Science Laboratory, Bhopal, Madhya Pradesh, India

School of Arts and Sciences, Ahmedabad University, Ahmedabad, India

Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, Gujarat, India

R. K. Shukla
Biological and Life Sciences, School of Arts and Sciences, Ahmedabad University, Ahmedabad, Gujarat, India
e-mail: ritesh.shukla@ahduni.edu.in

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_54

Abstract

Identity is a set of individual characteristics. It is crucial to criminal justice system as it links the crime to the perpetrator and the victim. Since long there have been many developments in shaping out the systems of identification like bertillonage, serological methods of identification, etc. Later with the advent of DNA profiling technique using Short Tandem Repeat (STRs), the process of individualization became more robust. With the advancements in the field of genetics, a novel technique was introduced to Forensic DNA technology called DNA Phenotyping. It is the technique where the phenotypic traits are translated from the genotype of the individual and a "Snapshot" of a suspect is created. Although the technology is new, it is growing at a faster rate. The present chapter unfolds the science behind this novel technique. It talks about the DNA-based studies on the prediction of phenotypic characters, legislations, and the notable case studies.

Keywords

DNA Phenotyping · SNPs · SNaPshot Technology · Parabon NanoLabs · IDentify

Inception of DNA Technology

Identification is crucial to criminal justice system. Identity could be defined as a set of characteristics or attributes, physical or functional, which are unique to an individual. Knowing these unique features helps in identification and individualization of the miscreant. Identification helps in establishing the relationship between the crime, victim, and the criminal, aiding in criminal trial. It also helps in tracing the victims of mass disaster.

Since long the criminal justice system has been working to develop a system of identification which is easy to understand, scientifically and statistically sound, saves time, and is cost-effective. Sooner, the very first identification system called the **Bertillon System** was developed by Sir Alphonse Bertillon in 1879. The system considered measurements of 11 body parts, photography, and individual's descriptive features, collectively called as Bertillonage System. Advancements in science and technology kept on improving the existing identification systems. It can be seen clearly that one system of identification succeeded the other. Following this, the anthropometric system of identification was replaced later by the fingerprint system of identification (Pierre Piazza 2006). Parallel to this, forensic biology which includes forensic serology and forensic DNA technology was also developing.

It is well known that the deoxyribonucleic acid (DNA) is the informational basis of life. The twentieth century observed the development of DNA profiling technology (1984) by Sir Alec Jeffreys where the polymorphism in DNA was taken as markers to establish individuality (Matheson 2016). Since then it has become a robust tool for forensic community. Initially Restriction Fragment Length Polymorphism (RFLP) technology was developed which was highly discriminatory and

Table 1 Extended set of 20 CODIS core loci

CSF1PO	FGA	THO1	TPOX
VWA	D3S1358	D5S818	D7S820
D8S1179	D13S317	D16S539	D18S51
D21S11	D1S1656	D2S441	D2S1338
D10S1248	D12S391	D19S433	D22S1045

scientifically sound. It could analyze the single and multilocus markers for individualization. But the analysis would require large amount of DNA sample (Murnaghan 2019). Due to many such limitations the RFLP methodology of DNA profiling failed. It was then taken over by more sound and advanced technology which used Variable Number of Tandem Repeats (VNTRs) as markers. The technique consumed comparatively less amount of sample DNA. Using VNTRs analysis of degraded DNA also became possible (Gaensslen et al. 2008). But this was not the end. The experts of forensic community tried to introduce a system where even lesser amount of DNA could be used without hampering the reliability of results.

With the invention of PCR by Kary Mullis in 1983, PCR-based markers for DNA analysis came into practice. This created the sense of enthusiasm among the researchers. Sooner, Short Tandem Repeats (STRs) became markers of choice by the experts (Goodwin et al. 2007). The DNA profiling technique using STR markers became popular across the laboratories of the world. The method proved to be relatively sensitive, less time-consuming, and required almost one tenth of the nanogram of sample. Further the analysis of Y-STR markers, mtDNA (mitochondrial DNA) markers, familial searching developed making the DNA profiling technology more robust and reliable (Laan 2017).

In the year 1994, the US congress brought a bill named DNA Identification Act (Public Law 103, 322) under which Combined DNA Index System (CODIS) was established. With the immense success of UK National DNA Database (UK NDNAD), the concept of databasing the DNA profiles emerged. FBI created its own (US) DNA database National DNA Index System (NDIS) later in 1998. Hence, DNA profiling technology found a new height with the development of DNA databases throughout the world. The formation of CODIS led to the standardization of 13 Core loci which was then extended to the set of 20 core loci (Butler 2010; Li 2008; Goodwin et al. 2008). These CODIS core loci are given in Table 1.

Thus, we can see that the identification systems have been evolved gradually from traditional anthropometric methods to technically more advanced, sensitive, and reliable DNA profiling technology.

Although the DNA profiling or the DNA fingerprinting technique is being practiced worldwide for individualization, it has also got few limitations.

DNA profile generated from the questioned biological sample is taken for comparison with a reference sample in accordance with the *law of comparison* to look for a possible match. The reference profile is either generated from the sample collected from the suspects or the reference profile drawn out from the DNA databases. The problem with DNA profiling arises when the reference profile is unavailable. In such cases DNA Dragnets method is used where the whole population residing in the area under investigation is screened for DNA. The culprit is expected to participate in the

mass screening, but the probability is far less than expected. If not the perpetrator then his/her close relative may have participated and could be investigated through the familial searching method using Y-STR and mtDNA technology. But then there are cases where these DNA Dragnets too fail (Kayser 2015).

These limitations of relative DNA technology (DNA Profiling) gave birth to a new domain in DNA technology called Forensic DNA Phenotyping.

Forensic DNA Phenotyping

Introduction

Forensic DNA Phenotyping (FDP) is a novel technique introduced to Forensic DNA technology. It is still developing and growing at faster rate. FDP can be understood as the technique where the phenotypic traits are translated from the genotype of the individual and a "Snapshot" of a suspect is created. Since the information predicted from the genotype cannot be exact, the inference is, however, made by considering the likelihood ratio and presenting it with significant certainty (Samuel and Prainsack 2018).

The main objective of the DNA Phenotyping technique is to identify and create a "probable" picture of an unknown suspect, a missing individual, or a victim of mass disaster where the traditional DNA identification systems fail due to unavailability of ante-mortem DNA profile or profile of any relative for reference. The technology uses the molecular genetic markers (DNA isolated from the biological evidences collected from the crime scene) to assist investigators in characterization and investigation process (Kayser and de Knijff 2011). Thus, it can be rightly said that the FDP technology acts as "biological witness" – more reliable than traditional eye witnesses.

Besides helping the criminal justice system in identifying the unknown perpetrators, FDP is also expected to aid anthropologists and paleo-genetic researchers in reconstructing old human remains using ancient DNA analysis. Moreover, FDP is considered to answer questions related to bio-geographic ancestry; however, there are cases of mixed ancestry which cannot be solved by only considering the physical traits (King et al. 2014).

The notion of applicability of molecular biology – DNA technology – to forensic science came relatively late and progressed slowly. It was in the late 1990s and early 2000 when experts started applying the concepts of DNA technology to forensic caseworks. Forensic DNA Phenotyping as compared to other DNA methodologies was introduced even later because the knowledge about the predicting appearance still remains limited. Although there have been many researches in the field of human genetics, unfortunately most of them were directed toward the gene-disease relationships (Stranger et al. 2010). Scientists worked in several projects; even the milestone project "The Human Genome Project" was intended to decode the human disease genetics by sequencing the whole genome. Very less has been studied about the genes responsible for a particular visible trait/physical appearance. Possible reasons could be the funding for genetic disease-related projects is high compared to the researches on genetic influence and human variation (Kayser 2015).

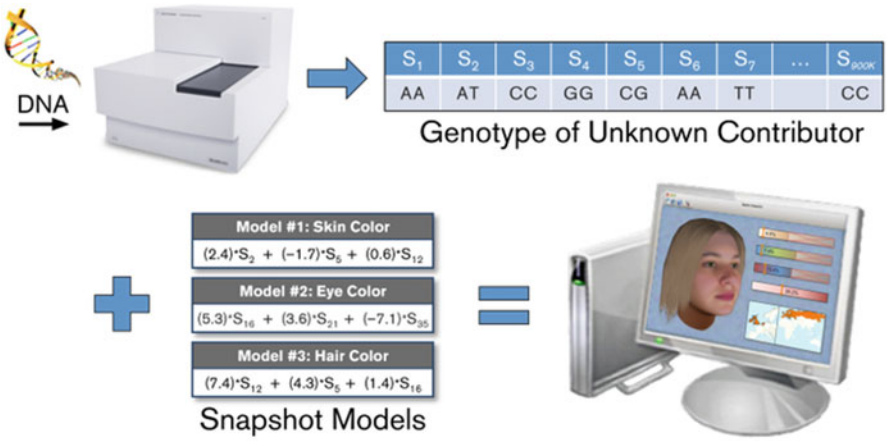


Fig. 1 The basic working methodology of Forensic DNA Phenotyping, where DNA is genotyped and processed by applying statistical models to formulate a probable snapshot of the unknown individual ("Parabon® Snapshot® DNA Analysis Service – Powered by Parabon NanoLabs," 2020)

Apart from this it is well known that these visible characteristics or the phenotypic traits are multi-faceted. They aren't only controlled by the genetic markers alone but are the result of coordinated mutual interaction between the environmental factors and genetic markers (Kayser 2015).

DNA Phenotyping is basically of two types – indirect and direct. Indirect approach to DNA Phenotyping infers the external features of individual by accessing his/her geographic origin. However, the direct DNA Phenotyping analyses the genome/genotype of the individual to infer external physical characteristics of the individual for identification. For Forensic DNA Phenotyping to be concerned both the approaches are followed. The technology utilizes PCR-STR and identifies the SNPs associated with the EVCs (Fig. 1).

Science Behind DNA Phenotyping Technology

To understand how a human population varies in terms of physical appearance, it is important to first understand few basic concepts of genetics.

In a diploid organism for a given gene, there exist two allelic forms which interact with each other and produce a particular physical trait. The expression of these alleles producing the physical trait is called as its *Phenotype*, whereas the genetic makeup of the organism which includes all the expressed and non-expressed groups of genes is termed as *Genotype* (Bartee et al. n.d.). Another term associated with the DNA Phenotyping is SNPs. SNPs or the single nucleotide polymorphism can be understood as a type of (most common) genetic variation in which there exists a change at the level of single nucleotide. A person's DNA can have 4.5 to 5 million SNPs. They tend to occur after every 1000 nucleotides and act as markers. They can

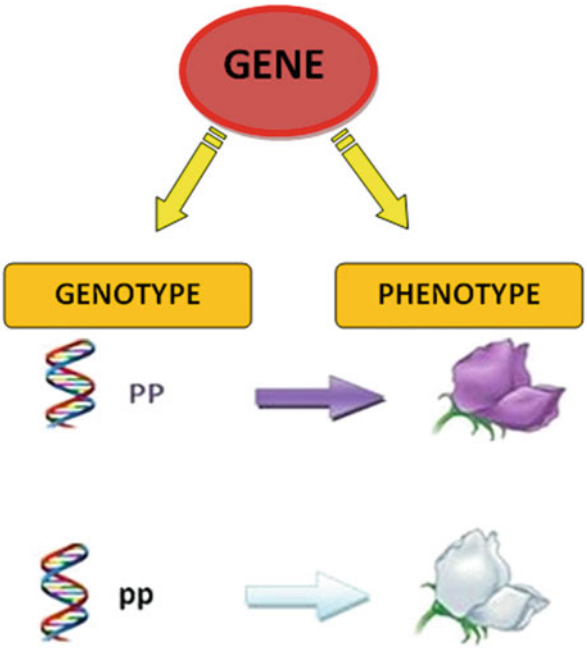


Fig. 2 Figure showing the difference between Genotype and Phenotype

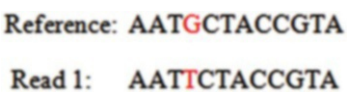


Fig. 3 Figure showing the SNP at 4th Nucleotide position (shown in red)

occur within a gene and may play a direct role in gene function ("Genomic Research," 2020) (Figs. 2 and 3).

As per the studies, scientists have found that the different kinds of variations like SNPs and insertions and deletions commonly called as InDels depending upon their position on regulatory gene directly influence the protein translated. They do so by altering the sequence of amino acid chain. If in case this altered protein is associated with some phenotypic character, then a changed/varied external feature is observed. Thus, these variations in genome of an individual (genotype) tend to influence the physical characteristics (phenotype) (Marano and Fridman 2019).

Presently, the scientifically and statistically sound methods of genotyping for the Pigmented Physical traits, *eye color*, *hair color*, and *skin color*, are under practice for accurate prediction (Mehta et al. 2016; Phillips 2015). Additionally, studies have shown that traits like biological age, morphology of face and hair, freckling, body

Table 2 Group of researchers who were behind the development of FDP technique (Matheson 2016; Pollack 2015; Murphy 2013; Enserink 2011)

S. no.	Name of the scientist	Institution	Research area
1.	Dr. Manfred Kayser	Forensic molecular biology department, Erasmus University Medical Center in Rotterdam, Netherlands	Laid the foundation of many advances in biological analysis, which includes identification of male DNA using the Y chromosome
2.	Dr. Tim Spector	Genetic Epidemiology, Kings College, London, England, and the Director of the TwinsUK Registry	The Inheritability of visible traits as demonstrated in biological twin populations is recognized as a milestone work. His work has given a new direction toward the study of relationship between genetic markers and EVCs
3.	Dr. Susan Walsh	Forensic geneticist, Indiana University-Purdue University, Indianapolis, Indiana	Working upon DNA intelligence tools to aid criminal justice system in identifying the unidentified deceased body
4.	Dr. Mark D. Shriver	Professor (anthropology and genetics), Pennsylvania State University	Dr. Shriver along with Dr. Claes developed a complex mathematical method to represent face which was based on measuring the 3–D coordinates of more than 7,000 points on the face. It is the basis for FDP imaging computer software
5.	Dr. Peter Claes	Researcher in morphometrics at the Medical Image Computing laboratory at KU Leuven in Belgium	Worked along with Dr. Shriver on developing a complex mathematical method to represent faces
6.	Dr. Wojciech Branicki	Researcher, Department of Genetics and Evolution, Jagiellonian University, and a DNA expert at the Institute of Forensic Research, located in Krakow, Poland	Identified the genetic markers which are responsible for the expression of EVCs

height, baldness in males, and genetic diseases can also be predicted using the genotype analysis (Murphy 2013).

The panel of eminent scientists whose contribution to FDP is cherished has been summed up in Table 2. Their studies and observations have laid the foundation and gave the strength to the field. Although the field is new, the immense work of these scientists have put together so much knowledge already. It has drawn the attention of many researchers and scientists toward the utility of phenotyping traits.

DNA-Based Prediction of Externally Visible Characteristics

Eye Color

The very first publication on DNA-based prediction of eye/iris color came out in the year 2007. The year saw studies from two well-known expert groups – Frudakis et al. and Sulem et al.

Frudakis et al. (2007) worked upon OCA2 gene and used 33 SNPs. With their observations they were able to classify eye colors (8%) of more than 1000 samples considered. At the same time Sulem et al. (2007) conducted GWAS studies on pigmentation traits. He used 9 SNPs from SLC24A4, KITLG, 6p25.3, TYR, OCA2–HERC2, and MC1R genomic regions. Although out of all the individuals (Europeans) chosen as samples, 60% were brown eyed, and the observation was made with 0.5 probability.

In 2008, there were many experts who were simultaneously working on the DNA-based eye color prediction. The major groups were Sturm et al., Eiberg et al., and Kayser et al. Sturm et al. and Eiberg et al. inferred that HERC2 rs12913832 plays an important role in color prediction of eye (Sturm et al. 2008; Eiberg et al. 2008). However, Kayser et al. in their GWAS studies included other HERC2 SNPs like rs916977 as major SNP in color prediction (Kayser et al. 2008). Their group further performed studies on DNA-based prediction of eye color and used 3 SNPs: HERC2 rs916977, OCA2 rs11855019, and OCA2 rs7495174. Upon statistically (receiver characteristic operating curve) analyzing, they found that the accuracy reaches 0.8, where 0.5 stands for random prediction level and 1.0 shows complete accurate prediction. Moreover, HERC2 rs916977 alone provided the eye color predictive value mostly.

Later in 2009 major work in the area was done by Liu et al. (2009). Liu et al. did a thorough and extensive study on DNA-based eye color prediction. In their study they took SNPs from previous studies and did exhaustive study on their prediction ability. The number of SNPs was taken to be 37 from 8 pigmentation genes, and more than 6100 Dutch Europeans were sampled for study. They also created and validated a model based upon 24 SNPs from 8 pigmentation genes. The results, however, provided the accuracy level of 0.93, brown; 0.91, blue eye color; and 0.73 for intermediate eye color. The study also shows that as predicted by earlier studies SNP HERC2 rs12913832 to be more informative about the eye color, alone achieved the predictive values of 0.899 and 0.877 for brown and blue, respectively. Thus, after validation of the model and analyzing the results from the prediction curve, Liu et al. suggested following SNPs from 6 pigmentation genes: HERC2 rs12913832, OCA2 rs1800407, SLC24A4 rs12896399, SLC45A2 rs16891982, TYR rs1393350, and IRF4 rs12203592 (Liu et al. 2009). This set of SNPs provided the values of 0.93 for brown, 0.91 for blue, and 0.72 for intermediate eye color when tested among more than 2300 Dutch Europeans.

In Year 2010, Valenzuela et al. (2010) published a study where they tested 24 pigmentation genes for 75 SNPs with regard to their effect on pigmentation traits skin, hair, and eye. The study considered European as well as non-European

population. However, mixing these European and non-European subjects in analysis questioned the authenticity of the study – since ancestry too had a role to play when it comes to two different populations. Although the study on one hand got success in proving that the SLC24A5 rs1426654 could not be considered to have its effect on pigmentation trait (19, 20). Lately, Mengel-From et al. (2010) suggested the value of HERC2 rs12913832 in influencing the eye color pigmentation trait.

Based upon the earlier studies and exhaustive observations of Liu et al., Walsh et al. (Walsh et al. 2011a) developed the first DNA-based eye color prediction system. It was intended to be used by forensic community in identification purpose. The system was named as IrisPlex. The IrisPlex system is sensitive and compatible to the SWGDAM guidelines (Walsh et al. 2011b). The system works by multiplex genotyping all the 6 SNPs demonstrated by Liu et al. (HERC2 rs12913832, OCA2 rs1800407, SLC24A4 rs12896399, SLC45A2 rs16891982, TYR rs1393350, and IRF4 rs12203592). The system was then crossed checked for its validity under different situations and was approved for use. The application uses interactive user experience providing the Excel sheet for user input of data. Thus, being validated at all possible levels, the system was said to be accurate worldwide, and the biogeographic origin does not have much influence to it. In the year 2014, an advanced and technically sound version of the IrisPlex was developed. The advanced technology was even accessed by EDNAP and ISFG, and as per them the module is reliable and easy to use (Chaitanya et al. 2014). For the IrisPlex a commercial tool was invented named as Identitas V1 Forensic Chip. It allowed inference of all the 6 SNPs IrisPlex, ancestry analysis, etc. from the genomic sample (Keating et al. 2012). The tool can be searched under the link http://identitascorp.com/. Studies related to hair color/eye color prediction can be done using the IrisPlex model (Keating et al. 2012). The development of such system has never failed to invite debates on SNP sets included in the system especially regarding the prediction of the intermediate eye color pigmentation. There were groups of experts, Spichenok et al. and Pneuman et al., who insisted inclusion of few more SNPs (Spichenok et al. 2011; Pneuman et al. 2012).

One such attempt was done by Allwood and Harbison (2013) who proposed 19 SNPs and emphasized on including 4 SNPs (SLC24A4 rs12896399, OCA2 rs1800407, TYR rs1393350, and HERC2 rs1129038) for eye color prediction. Later a 23 SNP-based system of eye color was proposed by Ruiz et al. (2013); the authors highlighted the addition of HERC2 and OCA2 SNPs.

Finally a set of 13 SNPs – 6 IrisPlex SNPs, 4 HERC2 SNPs (rs1129038, rs11636232, rs7183877, and rs1667394), and 3 OCA2 SNPs (rs4778241, rs4778232, and rs8024968) – were accepted on the basis of accuracy level (Kayser 2015).

Currently, there are many scientific groups who are still working for predicting eye color pigmentation with accuracy. Some are emphasizing upon limiting DNA-based eye color prediction only to HERC2 rs12913832, while others are debating over effect of gender on influencing the accurate prediction (Kayser 2015).

Conclusively it can be said that the field concerning with the eye color pigmentation is still expanding and a lot of studies are required to establish the markers influencing the trait with accuracy.

Hair

Initially studies on hair color were centered to red hair only. Hair color as an EVC has been studied in detail by several expert groups, and the major ones are Grimes et al., Branicki et al., Sulem et al., and Valenzuela et al.

Grimes et al. (2001) are known for their contribution in introducing the very first DNA test which could predict the hair color. The DNA test was based upon 12 MC1R DNA variants (Grimes et al. 2001). Branicki et al. in the year 2007 sequenced the MC1R gene in more than 180 individuals with hair color variations which included 40 red-haired ones and 36 blond hair individual. They, thus, developed red hair prediction assay on the basis of 5 MC1R DNA variants (Branicki et al. 2007).

Sulem et al. (2007) in the year 2007 started a GWAS on pigmentation and published the first DNA prediction assay for hair colors belonging to all category. The authors predicted the red hair individual with >0.5 probability by using 2 MC1R SNPs (rs1805008 and rs1805007). In another study the authors predicted all the other pigments except red using 9 SNPs from 6 genes. But they fail to produce accurate result. Year 2010 saw publications from Valenzuela et al. (2010) who claimed to report 3 SNPs (SLC45A2 rs16891982, SLC24A5 rs1426654, and HERC2 rs12913832) for total hair melanin.

Later in the year 2011, Branicki et al. used 46 SNPs associated with hair color to evaluate their predictive power in Europeans of Poland. Further they developed a model which involved 22 SNPs having AUC values of 0.93 for red hair, 0.87 for black, 0.82 for brown, and 0.81 for blond, respectively (Branicki et al. 2011). However, they failed to confirm the role of SLC24A5 rs1426654 in predicting hair color.

On the basis of all the previous studies especially the SNP prediction ranking by Branicki et al., a DNA test system for predicting hair and eye colors was developed (2013). This test system is known as HIrisPlex system and is known to have a single multiplex genotyping assay for SNPs associated with eye and hair color. The system thus includes all 6 SNPs from IrisPlex two prediction models, one for hair color and one IrisPlex model for eye color. The 24 SNPs included in the HIrisPlex system are MC1R SNPs (one indel, Y152OCH, N29insA, rs1805006, rs11547464, rs1805007, rs1805008, rs1805009, rs1805005, rs2228479, rs1110400, and rs885479), two from SLC45A2 (rs28777 and rs16891982, one from KITLG (rs12821256), one from EXOC2 (rs4959270), one from IRF4 (rs12203592), two from TYR (rs1042602 and rs1393350), one from OCA2 (rs1800407), two from SLC24A4 (rs2402130 and rs12896399), one from HERC2 (rs12913832), one from ASIP/PIGU (rs2378249), and one from TYRP1 (rs683) (Walsh et al. 2013).

For model-based hair color prediction, SNPs TYR rs1393350 and SLC24A4 rs12896399 were used, whereas 6 IrisPlex SNP were used for model-based eye color prediction.

HIrisPlex was then validated for the forensic purpose in 2014 and has been described to be compatible with SWGDAM guidelines. Furthermore, the assay can also be applied to access the degraded DNA samples (Draus-Barini et al. 2013; Walsh et al. 2014).

Presently, HIrisPlex system is being commercialized as Identitas V1 Forensic Chip. It includes 22 SNPs for hair color prediction used in the HIrisPlex system. This tool is available at http://identitascorp.com/. It also provides access to other information relevant to forensic investigations like eye color/hair color prediction. However, 4 MC1R SNPs (N29insA, Y152OCH, rs1805007, and rs1805009) are not included in the system due to several other reasons (Keating et al. 2012).

Skin

Presently, there is little knowledge about the skin color variation trait as compared to what we have for other EVCs. This lack of knowledge could be associated with the fact that the global distribution of the skin color variation trait is heterogenous in European population which is unfavorable for GWAS studies, whereas the other traits – eye color and hair color –are homogenous among the population. Hence an inference for these traits can be drawn easily (Kayser 2015).

Research on skin pigmentation trait using the same traditional multi-ethnic study group was conducted by Valenzuela et al. (2010). The authors found accuracy level of only 45.7% for skin reflectance. The study emphasized 3 SNPs (SLC45A2 rs16891982, SLC24A5 rs1426654, and ASIP rs2424984) influence the skin pigmentation. Later, on the basis of 7-SNP set (the 6 SNPs for eye color, SLC24A5 rs1426654) and 2 SNPs as described for skin color by Valenzuela et al. Spichenok et al. (2011) predicted the non-white and non-dark skin pigmentation.

Pneuman et al. (2012) on their detailed study assessed and evaluated the above-stated 7-SNP set and found 1% error. On a similar study Hart et al. (2013) emphasized on 6 SNPs which included the 7-SNP set by Spichenok et al. except IRF4 rs12203592. They concluded that there occurred no error.

Maroñas et al. (2014) articulated their first comprehensive study in which they included 59 SNPs which were also associated with skin, eye (Ruiz et al. 2013), and hair color in European and non-European populations. In this study they framed questions and measured the skin reflectance. These sets of SNPs differentiated white skin-colored individuals from intermediate/black skin-colored individuals. The authors highlighted the following SNP set, SLC45A2 rs16891982, SLC24A5 rs1426654, KITLG rs10777129, ASIP rs6058017, TYRP1 rs1408799, and OCA2 rs1448484, for skin pigmentation prediction. They further gave the additional SNP set including SLC45A2 rs13289, SLC24A4 rs2402130, TPCN2 rs3829241, and ASIP rs6119471.

Since previous studies have included only a small number of samples, it becomes essential to include more data from the individuals so as to form a reliable opinion about the predictive accuracy. It could also be suggested that as with other EVCs skin pigmentation trait should also be considered for quantitative prediction.

DNA-Based Inference of Non-pigmented Traits

Face

Identifying the person from the facial sketch is easy. One can expect predicting the facial characteristics from DNA using FDP technique. And if possible this will be regarded as a gold standard technique in the field of forensic science. But the reality of predicting facial features using DNA markers is still a far reaching dream.

Liu et al. (2012) studied about the markers influencing the facial traits/shapes in their GWAS. At the same time, Paternoster et al. in their GWAS (Paternoster et al. 2012) too were studying the same markers. The GWAS published by Liu et al. (2012) and by Paternoster et al. (2012) is the only study where the genes involved in expression of the facial characteristics have been described in brief. Liu et al. (2012) in his genome-wide study worked upon 10,000 Europeans and finally came up with a set of 5 genes PAX3, PRDM16, TP63, C5orf50, and COL17A1. Their study stated that these genes are involved in human facial shape variation. Using 3D magnetic resonance images (MRI) of the head and of 2D portrait pictures, they automated the facial landmarks and had their measures (facial distancing). Of these 5 gene set three (PAX3, PRDM16, and TP63) have already been recognized in a study related to craniofacial development and disease. The only gene to be identified for influencing a facial character (nasion) is PAX3 (Paternoster et al. 2012).

Both the GWAS studies revealed that for determining a facial trait, a large number of DNA variants are involved. Liu et al. (2012) further suggested that for TP63 rs17447439 the largest effect can be seen, where the heterozygote carriers had a 0.9 mm and the homozygote carriers a 1.8 mm reduced eye-to-eye distance.

Later, Claes et al. (2014) used complex but advance approach to infer facial traits. They studied SNPs from craniofacial genes in three populations (US Americans, Brazilians, and Cape Verdeans). The researchers highlighted 20 genes and 24 SNPs. Of all, more stress was given on three SNPs, namely, SLC35D1, FGFR1, and LRP6. Recently a group of researchers have questioned the statistical approach used by Claes et al. (Hallgrimsson et al. 2014).

Since very little is known about the facial variation in humans, the field needs more involvement and research. It seems there is a long way for the technique to get established and to provide significant predictive markers to be used by the forensic experts.

Hair Structure

As per the independent studies by Fujimoto et al. (2007) and Medland et al. (2009) in Asian and European population, respectively, three genes have been reported to be associated with the human hair morphology (Fujimoto et al. 2008). In a genomic scan study on Asian population, Fujimoto et al. described FGFR2 and the EDAR gene to be involved in hair thickness (Fujimoto et al. 2009). In similar studies

conducted by Medland et al. (2009), the gene associated with the straight hair variant in Australian Europeans is found to be TCHH.

Since the hair structure in Asian population remains uniform, the significance of applying Human Hair Structure in identification process remains null. However, in European population the gene TCHH explained the 6% variation observed. The authors briefed that the minor T allele, rs11803731 (a coding, nonsynonymous variant in exon 3 of TCHH), is absent in the populations of East Asia, Oceania, sub-Saharan Africa, and in Native Americans, but is found across Europe and neighboring regions such as North Africa, the Middle East, and West Asia (Medland et al. 2009a). Medland et al. also studied the candidate genes selected by Fujimoto et al. and found that they have an association with WNT10A (Fujimoto et al. 2007).

Both the studies on hair morphology suggested that there exists a heritability rate for curly variant, but the genes responsible were not identified (Medland et al. 2009b).

Since there is a great diversity for hair morphology among European (straight, wavy, or curly), FDP will have a significant impact only when there will be more predictive genes for this particular trait.

Hair Loss/Baldness

Presently as per the studies conducted over the phenomenon of baldness, 12 genomic regions have been identified. These genes are found to be associated with the early onset of androgenic alopecia (AGA). Androgenic alopecia is commonly found in male and is associated with baldness. The genes found to be associated with this particular trait are AR/EDA2R, TARDBP, HDAC9, AUTS2, SETBP1, PAX1/FOXA2, WNT10A, 17q21.31, 3q25, 5q33.3, and 12p12.1 (Richards et al. 2008; Li et al. 2012; Heilmann et al. 2013). A study has highlighted AR/EDA2R to be strongly associated with the trait and is located on X chromosome. The female pattern hair loss (early onset) is a rare phenomenon. But it has been observed that it shares some of the genetic basis of early baldness with male (as regard to the X chromosome). But no studies have shown the association of the other genes. Thus, the etiology still remains unclear (Redler et al. 2012).

The hair loss trait seems to be promising, but since all the studies have been done on early onset pattern and the general population have late onset pattern, the genetic study becomes less vocal. Thus, there is need to establish studies on late onset hair loss pattern which is more likely to have strong connection with the environmental effects too.

Age

There are several traits which indicate age indirectly, but as it seems the age in itself is an externally visible characteristic. There have been several studies in predicting

the genes responsible for aging. Many studies have established the fact of decrement in the T-cell numbers with age (sjTREC). Based upon this knowledge Zubakov et al. in 2010 introduced a method of age estimation by quantifying the sjTREC (Zubakov et al. 2010).

The process of DNA Methylation is also considered to be one of the processes behind aging. The field of epigenetics has evolved in recent years and has expanded its knowledge system about DNA methylation and dependency of aging on it. It has also enhanced our understanding over highly promising CpG candidate markers.

Expert groups like Bocklandt et al. (2011) studied these CpG markers for age estimation and concluded that promoters of EDARADD, TOM1L1, and NPTX2, two CpG markers, are responsible for 70–73% of variation in age. Further authors predicted the age of an individual with average accuracy of 5 years.

Later, Garagnani et al. (2012) published a study on CpG sites in 3 genes ELOVL2, FHL2, and PENK. Out of these three genes, the authors highlighted the ELOVL2 gene as a strong age prediction marker. In a similar study, Weidner et al. (2014) established a relationship between DNA methylation with chronological age. Further, they introduced age estimation calculator. As we all know chronological and biological age of an individual can differ, the DNA methylation is seen to be associated more with biological age. Thus, before applying the theory of DNA methylation in forensic case works, it becomes important to understand the underlying principles behind them. Thus, in the near future one can see the application of age prediction based on DNA methylation in forensic science (Weidner et al. 2014).

Parallel to this Yi et al. (2014) studied eight loci and their relation with aging process. The authors included a small sample set (65 individuals) but failed to provide any validation of the markers' involvement in biological age. Later several studies have provided the evidence that ELOVL2 gene is a better marker till date for age estimation.

Body Height/Stature

There is a large data available for the body height. There have been several genome-wide studies which had used body height as a common complex trait. Previously, there have been several GWAS studies by expert groups like Aulchenko et al. and Liu et al.; Genetic Investigation of ANthropometric Traits (GIANT) consortium on human height trait. GAINT consortium investigated >183,000 individuals and identified hundreds of SNPs at 180 genetic loci significant height association (Lango Allen et al. 2010).

Aulchenko et al. (2009) did the very first systematic study on body height in 2009. The authors predicted top 5% of tall individuals from >1000s of Dutch Europeans using 54 height-associated SNPs. Liu et al. (2013) in 2010 proposed that the 180 genetic loci identified by GIANT have significant association with the normal height.

However, the knowledge bank on body height signifies that a lot of information is still missing and there are more than a thousand of SNPs associated with the body height trait in humans.

Currently Available Forensic DNA Phenotyping Test Systems

Currently only a handful of software has been developed by the research groups which upon analyzing the data provide reliable results on physical traits. Apart from physical traits, these softwares have been developed to infer genetic ancestry, freckling, and facial shape. The milestone softwares aiding the criminal justice system are **IDentify** by **Identitas** and **SNaPshot Software** by **Parabon NanoLabs.**

IDentify is an advance software. Using around 800,000 SNP database and applying rules of genetics upon them generates the external visible traits. It also infers the biogeographic ancestry. The predictions are highly reliable. Forensic community many a times faces the issue related to mixed DNA sample. Taking this into consideration Identitas is working toward developing a software that could discriminate between the mixed samples having 10 sources ("Identitas – Forensic Phenotyping," 2020).

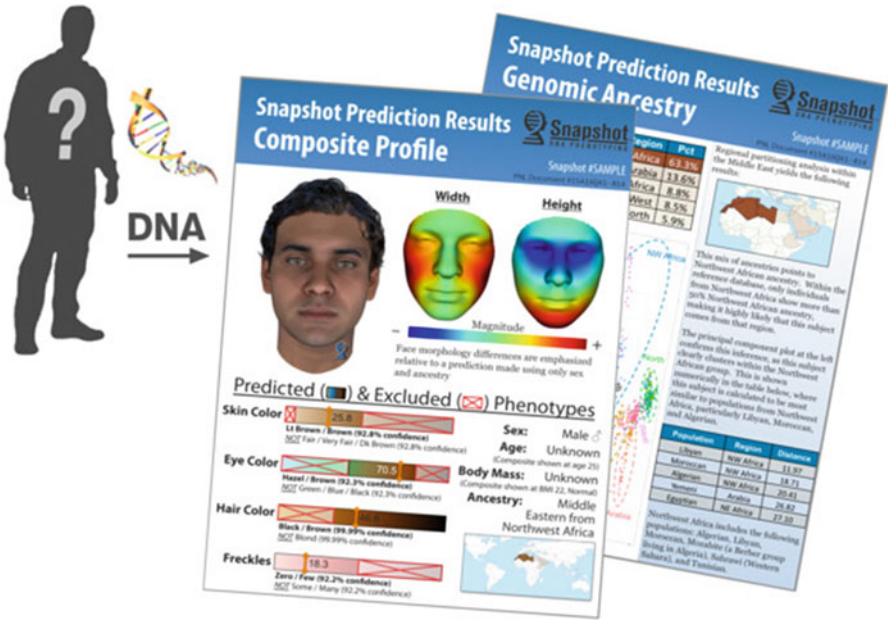


Fig. 4 Figure showing a sample SNaPshot prediction composite profile ("Parabon® Snapshot® DNA Analysis Service – Powered by Parabon NanoLabs," 2020)

Parabon NanoLabs situated in Virginia has released a software that works on the programming developed by Drs. Shriver and Claes (Pollack 2015; Mehta et al. 2016). The software is called SNaPshot and is taken care by the US Department of Defense. The SNaPshot utilizes the Next Generation Mini-Sequencing technology and processes 20,000 DNA markers, applies complex algorithms to finally create a file, and compares it with the donor profiles. It then uses the reverse engineering to prepare a snapshot of the suspect out of the raw data (Laan 2017). The SNaPshot system also has provisions which uses the Forensic Facial Reconstruction technique to create 3-D model of the suspect and then adds the non-biological characteristics like tattoos, eye glasses, hair style, etc. with the help of an artist (Laan 2017; Mehta et al. 2016). Like IDentify, SNaPshot analysis can infer genetic ancestry apart from phenotypic characteristics ("Parabon® Snapshot® DNA Analysis Service – Powered by Parabon NanoLabs," 2020) (Fig. 4).

Legislations for Forensic DNA Phenotyping

With the advancements in science and technology, the law keeps on amending itself. It can be seen that with the introduction of new technology, the new legislations are being formed or a new bill is framed to guard the use/misuse of the invented technology. Thus, scientific technology and law evolve simultaneously with their effect on each other.

Since the development of DNA technology, there have been rules and legislations framed in different parts of the world. Every country has its own legislation regarding application of DNA by criminal justice system.

With respect to the legislations for forensic DNA Phenotyping, not every country has got its own legislations framed, or otherwise some countries have a limited provision for regulation of forensic DNA Phenotyping.

The countries with common law systems – the UK and the USA – have provisions that allow the DNA testing unless it is not specifically prohibited by law (Kayser and de Knijff 2011). The UK has legislations that allow the analysis of the physical appearance of the perpetrator, but the same technology cannot be used for red hair testing and ethnic inference (MacLean 2013).

On similar lines, the federal statute of the USA has put a restriction on DNA Phenotyping technology only to identification purpose, although identification purpose has not been clearly defined in the statute (Murphy 2013; Kayser and de Knijff 2011). Since, in the USA law enforcement agencies have the tendency to slowly adapt the new technologies as compared to the other countries, thus the USA lags in regulating forensic DNA Phenotyping technology (Matheson 2016). Meanwhile some States do have a DNA processing regulation in their jurisdiction. The only state authorizing difference in effective Phenotyping testing which includes physical characteristics and genetic diseases is Texas (MacLean 2013). States like Florida, Michigan, South Dakota, Vermont, Utah, and Washington don't have any restrictions on FDP or external visible characters, but they do prohibit DNA testing in case of genetic disorders or other medical-based studies. As a representative legislation,

these five states have Vermont statute. The statute states that "[the] analysis of DNA samples is authorised... to type the genetic markers from the DNA samples for law enforcement identification process...[but] analysis of DNA samples obtained pursuant to [the statute] is not authorised for identification of any medical or genetic disorder" (MacLean 2013). The States prohibiting the determination of physical traits from DNA include Indiana, New Mexico, Rhode Island, and Wyoming (Laan 2017). As per Indiana statute the DNA profiles stored in DNA database are prohibited for obtaining information about external visible characteristics (MacLean 2013).

Similarly Rhode Island's legislation theory states that the DNA samples are the DNA profiles which have been recorded as per the statute and will only be used by the law enforcement agencies for identification purpose. This identification purpose could include the identification of missing persons, but the submitted DNA samples or the profiles shall not be used for obtaining information about physical traits (MacLean 2013). The Wyoming legislation has been built on similar lines where obtaining information about EVCs is restricted whereas the DNA genotyping is permitted (MacLean 2013).

Countries like Canada, Belgium, Spain, and South Africa only allow non-coding regions of DNA for typing meanwhile genotyping is allowed in Australia.

Germany has the legislations where DNA testing is only permitted for determining gender source parent age; however, they forbid the application of FDP for other purposes (MacLean 2013; Laan 2017).

The only country having regulations regarding forensic DNA Phenotyping is the Netherlands. The legislation of the Netherlands restricts the testing only to physical traits. The Netherlands has the Dutch act "determining externally perceptible personal characteristics from cell material" which defines the DNA investigation. As per the legislation "DNA investigation [as] the research of cell material which is only targeted at comparing DNA profiles or determining externally perceptible personal characteristics of the unknown suspect." Further it states that "if it is uncertain that the source knows about the trait it may not be investigated...[and] any physical characteristic must be (1) externally perceptible; (2) visible; (3) present at the time of and since birth; and (4) publicly perceptible" (Enserink 2011; MacLean 2013; Kayser and de Knijff 2011).

In India, since "The DNA technology (Use and Regulation) Bill, 2017" is still pending, the legal position of DNA testing remains dicey.

Legal and Ethical Aspects of Forensic DNA Phenotyping

Since the DNA Phenotyping technology assesses the genomic content of an individual, it triggers the legal and some ethical issues with itself. Few ethical issues concerned with forensic DNA Phenotyping are:

1. **The slippery slope theory**: Since the FDP technology is very new, a lot of information is fed into the knowledge bank daily. This undoubtedly increases

the scope of forensic DNA Phenotyping technology in the near future. There is a group of scientists who fear that this advancement in capabilities of FDP will lead to the misuse of information inferred. According to them if with advancements assessing the information from one form of FDP will be allowed, then eventually all the forms of FDP restricted/unrestricted will be used. Thus, controlling only a part of FDP to be used for information purposes will become a tough task (Kayser and Schneider 2009).

2. **Right not to know**: This is a universal right and is also a part of universal declaration of the human genome and Human Rights declaration. Within this an individual enjoys his rights "no to know" his own that medical information- all the information gathered from genetic testing. As per the experts, the information like hidden parentage or genetic diseases could have drastic effects on the individual. Thus, the investigation must be limited to analysis of physically visible traits and all the associated information must not be told to the individual himself. Advancements in the technology have revealed that there exists a relationship between the genetic markers and the behavioral traits of the individual. The genomic testing could thus reveal the behavioral aspect of the individual relating it directly or indirectly to the person's criminal behavior, for example, aggression, anxiety, pedophilia, etc. The experts are concerned that if such information will be revealed to the person or be accepted by the court of law, the aspects like parole eligibility and preventive detention would be affected to a greater extent. A person may take a benefit of this information to make an excuse for his deeds. These are the characteristics of which an individual in himself is aware of, and hence analysts must refrain themselves from publication of such traits so as to avoid any future harm (mentally and socially) to the individual and/or his family (Murphy 2013; MacLean 2013; Kayser and Schneider 2009).
3. **Individual freedoms and privacy**: Breach of right to privacy is the foremost concern related to forensic DNA Phenotyping technology. DNA testing infers a lot more information about the concerned individual. This information can be abused in later period of time leading to hamper his right to privacy. Does the information that is released in public should be limited, and loan additional information that is beyond requirement should be derived from FDP testing; the testing such as genetic disease behavior propensities should remain confidential, or otherwise the donor individual should be given a choice of requesting the full profile. To avoid the privacy breach FDP data should be disposed as soon as the identification process has been completed (Murphy 2013; Toom 2012).
4. **Racial profiling**: Since FDP technique can be used to access the geographic ancestry, this clearly means that while identifying the individual from its EVCs his/ her ancestry can also be revealed. This would, in turn, create sense discrimination among the population. The society could label the respective race or the community with a criminal tag, thus encouraging the prejudice. This will be an injury to the complete society and could create fear among minor communities. A belief would be reinforced that a specific community is of a criminal mindset. Hence, not limiting the access of FDP for investigation can cause racial discrimination (Matheson 2016; MacLean 2013; Enserink 2011; Koops and Schellekens 2006).

However, the solution to these emerging issues would be restricting the use of FDP technology. The legislations put forward by the Netherlands could be followed at global level. The FDP technique should only be implied where the identification cannot be achieved by using traditional means of identification. Last but not the least the advancements in the field of forensic DNA Phenotyping technique should never be blanketed due to emerging concerns; rather a solution to such concerns should be put forward (Laan 2017).

Case Studies: DNA Phenotyping

Currently forensic community is using advanced technologies and working in collaboration with software companies like Identitas and Parabon NanoLabs in analyzing the evidences having limited leads (Laan 2017). Pilot studies are being carried out in the Netherlands, Poland, and Australia (Matheson 2016). Further, the active cases in Ontario are being investigated using the novel Forensic DNA Phenotyping technique (Laan 2017). A milestone case in the field of identification by FDP technique was of Delroy Easton Grant (Marcus and Monique 2015).

FDP technique has also played an important role in predicting the ancestry of the Baton Rouge Serial Killer, Derrick Todd Lee, who before getting arrested claimed the lives of seven women in 2003. In the year 2004, the case of Eric Copple in Napa, California, was also assisted by FDP technique (MacLean 2013). The infamous case of Robert Barnes of 2008 – the person behind the rape and murder of Meghan Landowski – in Virginia was also solved using the biogeographical ancestry inferred from the DNA collected from the body of the victim ("'48 Hours: NCIS:' To Catch a Killer," 2020).

The law enforcement agencies in the USA have been using SNaPshot technology since long. Using SNaPshot technique in combination with Familial analysis using DNA and Y-STR technique aided the identification of José Alvarez, Jr., the person who murdered Troy and LaDonna French (February 2012, North Carolina) ("Parabon NanoLabs: Engineering DNA for Next-Generation Therapeutics and Forensics," 2016). Moreover, cases like the Bennett family in Aurora, Colorado, in 1984; Lisa Ziegert in Springfield, Massachusetts, in 1992; Sierra Bouzigard in Calcasieu Parish, Louisiana, in 2009; Candra Alston and Malaysia Boykin in Columbia, South Carolina, in 2011; and identification of an unidentified male victim in Glen Burnie, Maryland, in 1985 were resolved using facial reconstruction technique provided by SNaPshot systems (Augenstein 2016; Greenwood 2016; Pollack 2015).

The results from the SNaPshot and IDentify analysis assist the investigation where traditional techniques fail to produce any leads. There are numerous cases where the DNA database could not find a match and the case was held stalled. Such cases have been re-examined/investigated using SNaPshot systems where the probable picture of the culprit was created and resumed the investigation. It can be said that the FDP technology is a promising technique to forensic community ("Parabon NanoLabs: Engineering DNA for Next-Generation Therapeutics and Forensics," 2016).

The Future of FDP

Since the Forensic DNA Phenotyping technology is novel, many research groups are working together to improve and further develop the technique. In such effort The US Department of Defense has granted around $2 million to Parabon NanoLabs (Laan 2017). The US Department of Defense has also signed an agreement upon development of a new technology called **Keystone** with Parabon NanoLabs. The Keystone module is being created with an aim to formulate a system which being the first platform will be able to integrate NGS tools for analyzing the DNA sample under a common umbrella. It is thought to be a platform where all the essential DNA analytical tools will be available to the forensic community. Keystone technology will soon be commercialized as a product for laboratories (Laan 2017).

Apart from this there are the groups of researchers who are continuously decoding the human genome to broaden the horizons of Forensic genetics. Studies are being done in the areas where the additional traits like hair morphology, age determination, facial structure, and adult height could be inferred from DNA analysis (Laan 2017). Furthermore, for the holistic development of the FDP technique, there is an urgent need of developing the analytical domain of research. It is now essential to improve and invent advanced new generation tools and methodologies which can offer parallel sequencing and DNA Genotyping in a short span of time while maintaining the quality of results (Phillips 2015; Kayser and de Knijff 2011).

Researchers are also gaining interest over other EVCs like handedness, chin and cheek dimpling, and earlobe attachment (MacLean 2013). In most of the cases only the teeth and blood are being taken as sample; experts today are more focused toward using other biological evidences (semen, saliva, etc.) as sample (Kayser and de Knijff 2011).

The main goal of FDP technique is identification of an unknown individual – victim or suspect. Determination of the biological age, as a factor for identification, from the DNA molecule is also achieving interest (Hamano et al. 2016). Since studies related to the FDP technology is still under process, one can imagine how powerful the technology would be in the coming years when all these projects will get commercialized. Surely, the technology would enhance the working capacity of the forensic laboratories around the world.

Therefore, there is no doubt in accepting the fact that the FDP technology is a technology of the future.

Conclusion

Identification is crucial to the criminal justice system. The science of identification in the past years has seen tremendous advancements and developments. It has evolved from the use of anthropometric methods of identification (Bertilionage) to the very advanced DNA profiling technology, developed by Sir Alec Jeffreys in1984. Since then, the DNA profiling technique proved to be the gold standard technique in establishing individuality across the globe. Since science is ever evolving, a very

novel technique called forensic DNA phenotyping (FDP) was introduced to forensic genetics. This technique can thus identify and create a "probable" picture of an unknown individual. From its inception, the technique has attracted several scientific experts from the forensic community. Studies are now being conducted on investigating the SNP sets for skin, hair, and eye color pigmentation genes. Systems like IrisPlex, IrisPlex model, and HIrisPlex system have been developed in this regard for forensic purposes. FDP technology is also being applied to the nonpigmented traits like hair structure, baldness, face, body height/stature, and age. The milestone softwares like IDentify by Identitas and SNaPshot Software by Parabon NanoLabs are now available in the market along with few upcoming proposed projects for commercial purposes. As the technology is expanding, countries are formulating new laws with regards to the latest technology in use. Similar to other genetic research works, FDP also poses some ethical issues which are being considered before applying the methodology in forensic case works. Thus, the FDP technique can be seen as a sensitive technology, aiding the forensic community in all its possible forms. It can be foreseen as a promising technology of the future.

References

"48 Hours: NCIS:" To Catch a Killer (2020) Retrieved 26 July 2020, from https://www.cbsnews.com/news/48-hours-ncis-meghan-landowski-marilyn-allen-murders-to-catch-a-killer/

Allwood J, Harbison S (2013) SNP model development for the prediction of eye colour in New Zealand. Forensic Sci Int Genet 7(4):444–452. https://doi.org/10.1016/j.fsigen.2013.03.005

Augenstein S (2016) DNA phenotyping recreates the face of an alleged serial killer. Forensic Magazine. Retrieved from https://www.forensicmag.com/article/2016/08/dna-phenotyping-recreates-face-allegedserial-killer on 15 Jan 2017

Aulchenko Y, Struchalin M, Belonogova N, Axenovich T, Weedon M, Hofman A et al (2009) Predicting human height by Victorian and genomic methods. Eur J Hum Genet 17(8):1070–1075. https://doi.org/10.1038/ejhg.2009.5

Bartee L, Shriner W, Creech C (n.d.) Phenotypes and Genotypes. Retrieved 17 July 2020, from https://openoregon.pressbooks.pub/mhccmajorsbio/chapter/phenotypes-and-genotypes/

Bocklandt S, Lin W, Sehl M, Sánchez F, Sinsheimer J, Horvath S, Vilain E (2011) Epigenetic predictor of age. PLoS One 6(6):e14821. https://doi.org/10.1371/journal.pone.0014821

Branicki W, Brudnik U, Kupiec T, Wolañska-Nowak P, Wojas-Pelc A (2007) Determination of phenotype associated SNPs in the MC1R gene. J Forensic Sci 52(2):349–354. https://doi.org/10.1111/j.1556-4029.2006.00361.x

Branicki W, Liu F, van Duijn K, Draus-Barini J, Pośpiech E, Walsh S et al (2011) Model-based prediction of human hair color using DNA variants. Hum Genet 129(4):443–454. https://doi.org/10.1007/s00439-010-0939-8

Butler J (2010) Fundamentals of forensic DNA typing, 10th edn. Academic/Elsevier, Amsterdam

Chaitanya L, Walsh S, Andersen J, Ansell R, Ballantyne K, Ballard D et al (2014) Collaborative EDNAP exercise on the IrisPlex system for DNA-based prediction of human eye colour. Forensic Sci Int Genet 11:241–251. https://doi.org/10.1016/j.fsigen.2014.04.006

Claes P, Liberton D, Daniels K, Rosana K, Quillen E, Pearson L et al (2014) Modeling 3D facial shape from DNA. PLoS Genet 10(3):e1004224. https://doi.org/10.1371/journal.pgen.1004224

Draus-Barini J, Walsh S, Pośpiech E, Kupiec T, Głąb H, Branicki W, Kayser M (2013) Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. Investig Genet 4(1):3. https://doi.org/10.1186/2041-2223-4-3

Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer K, Hansen L (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet 123(2):177–187. https://doi.org/10.1007/s00439-007-0460-x

Enserink M (2011). Can This DNA Sleuth Help Catch Criminals? Science 331(6019):838–840. https://doi.org/10.1126/science.331.6019.838

Frudakis T, Terravainen T, Thomas M (2007) Multilocus OCA2 genotypes specify human iris colors. Hum Genet 122(3–4):311–326. https://doi.org/10.1007/s00439-007-0401-8

Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L et al (2007) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet 17(6):835–843. https://doi.org/10.1093/hmg/ddm355

Fujimoto A, Ohashi J, Nishida N, Miyagawa T, Morishita Y, Tsunoda T et al (2008) A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia. Hum Genet 124(2):179–185. https://doi.org/10.1007/s00439-008-0537-1

Fujimoto A, Nishida N, Kimura R, Miyagawa T, Yuliwulandari R, Batubara L et al (2009) FGFR2 is associated with hair thickness in Asian populations. J Hum Genet 54(8):461–465. https://doi.org/10.1038/jhg.2009.61

Gaensslen R, Harris H, Lee H (2008) Introduction to forensics & criminalistics. McGraw-Hill Higher Education, New York

Garagnani P, Bacalini M, Pirazzini C, Gori D, Giuliani C, Mari D et al (2012) Methylation ofELOVL2gene as a new epigenetic marker of age. Aging Cell 11(6):1132–1134. https://doi.org/10.1111/acel.12005

Genomic Research (2020) Retrieved 15 July 2020, from https://ghr.nlm.nih.gov/primer/genomicresearch/snp

Goodwin W, Linacre A, Hadi S (2007) An introduction to forensic genetics. Wiley, pp 1–5

Goodwin W, Linacre A, Hadi S (2008) An introduction to forensic genetics. Chichester: Wiley

Greenwood V (2016) How science is putting a new face on crime solving. Retrieved 26 July 2020, from https://www.nationalgeographic.com/magazine/2016/07/forensic-science-justice-crime-evidence/

Grimes E, Noake P, Dixon L, Urquhart A (2001) Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. Forensic Sci Int 122(2–3):124–129. https://doi.org/10.1016/s0379-0738(01)00480-7

Hallgrimsson B, Mio W, Marcucio R, Spritz R (2014) Let's face it – complex traits are just not that simple. PLoS Genet 10(11):e1004724. https://doi.org/10.1371/journal.pgen.1004724

Hamano Y, Manabe S, Morimoto C, Fujimoto S, Ozeki M, Tamaki K (2016) Forensic age prediction for dead or living samples by use of methylation-sensitive high resolution melting. Legal Med 21:5–10. https://doi.org/10.1016/j.legalmed.2016.05.001

Hart K, Kimura S, Mushailov V, Budimlija Z, Prinz M, Wurmbach E (2013) Improved eye- and skin-color prediction based on 8 SNPs. Croat Med J 54(3):248–256. https://doi.org/10.3325/cmj.2013.54.248

Heilmann S, Kiefer A, Fricker N, Drichel D, Hillmer A, Herold C et al (2013) Androgenetic alopecia: identification of four genetic risk loci and evidence for the contribution of WNT signaling to its etiology. J Investig Dermatol 133(6):1489–1496. https://doi.org/10.1038/jid.2013.43

Identitas – Forensic Phenotyping (2020) Retrieved 28 July 2020, from https://www.identitascorp.com/

Kayser M (2015) Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet 18:33–48. https://doi.org/10.1016/j.fsigen.2015.02.003

Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12(3):179–192. https://doi.org/10.1038/nrg2952

Kayser M, Schneider P (2009) DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Sci Int Genet 3(3):154–161. https://doi.org/10.1016/j.fsigen.2009.01.012

Kayser M, Liu F, Janssens A, Rivadeneira F, Lao O, van Duijn K et al (2008) Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet 82(2):411–423. https://doi.org/10.1016/j.ajhg.2007.10.003

Keating B, Bansal A, Walsh S, Millman J, Newman J, Kidd K et al (2012) First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. Int J Legal Med 127(3):559–572. https://doi.org/10.1007/s00414-012-0788-1

King T, Fortes G, Balaresque P, Thomas M, Balding D, Delser P et al (2014) Identification of the remains of King Richard III. Nat Commun 5(1). https://doi.org/10.1038/ncomms6631

Koops B, Schellekens M (2006) Forensic DNA Phenotyping: Regulatory Issues. SSRN Electronic Journal 9(1):158–202. https://doi.org/10.2139/ssrn.975032

Laan M (2017) The genetic witness: forensic DNA phenotyping. Retrieved 26 July 2020, from https://jefsr.uwindsor.ca/index.php/jefsr/article/view/5005

Lango Allen H, Estrada K, Lettre G, Berndt S, Weedon M, Rivadeneira F et al (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467 (7317):832–838. https://doi.org/10.1038/nature09410

Li R (2008) Forensic biology. CRC Press/Taylor & Francis, Boca Raton

Li R, Brockschmidt F, Kiefer A, Stefansson H, Nyholt D, Song K et al (2012) Six novel susceptibility loci for early-onset androgenetic alopecia and their unexpected association with common diseases. PLoS Genet 8(5):e1002746. https://doi.org/10.1371/journal.pgen.1002746

Liu F, van Duijn K, Vingerling J, Hofman A, Uitterlinden A, Janssens A, Kayser M (2009) Eye color and the prediction of complex phenotypes from genotypes. Curr Biol 19(5):R192–R193. https://doi.org/10.1016/j.cub.2009.01.027

Liu F, van der Lijn F, Schurmann C, Zhu G, Chakravarty M, Hysi P et al (2012) A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genet 8 (9):e1002932. https://doi.org/10.1371/journal.pgen.1002932

Liu F, Hendriks A, Ralf A, Boot A, Benyi E, Sävendahl L et al (2013) Common DNA variants predict tall stature in Europeans. Hum Genet 133(5):587–597. https://doi.org/10.1007/s00439-013-1394-0

MacLean CE (2013) Creating a wanted poster from a drop of blood: using DNA phenotyping to generate an artist's rendering of an offender based only on DNA shed at the crime scene. Hamline Law Rev 36(3):1. Available at: http://digitalcommons.hamline.edu/hlr/vol36/iss3/1

Marano L, Fridman C (2019) DNA phenotyping: current application in forensic science. Res Rep Forensic Med Sci 9:1–8. https://doi.org/10.2147/rrfms.s164090

Marcus S, Monique M (2015) Recent developments in DNA evidence | Australian Institute of Criminology. Retrieved 26 July 2020, from https://www.aic.gov.au/publications/tandi/tandi506

Maroñas O, Phillips C, Söchtig J, Gomez-Tato A, Cruz R, Alvarez-Dios J et al (2014) Development of a forensic skin colour predictive test. Forensic Sci Int Genet 13:34–44. https://doi.org/10.1016/j.fsigen.2014.06.017

Matheson S (2016) DNA phenotyping: snapshot of a criminal. Cell 166(5):1061–1064. https://doi.org/10.1016/j.cell.2016.08.016

Medland S, Nyholt D, Painter J, McEvoy B, McRae A, Zhu G et al (2009a) Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 85(5):750–755. https://doi.org/10.1016/j.ajhg.2009.10.009

Medland S, Zhu G, Martin N (2009b) Estimating the heritability of hair curliness in twins of European ancestry. Twin Res Hum Genet 12(5):514–518. https://doi.org/10.1375/twin.12.5.514

Mehta B, Daniel R, Phillips C, McNevin D (2016) Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. Int J Legal Med 131(1):21–37. https://doi.org/10.1007/s00414-016-1490-5

Mengel-From J, Børsting C, Sanchez J, Eiberg H, Morling N (2010) Human eye colour and HERC2, OCA2 and MATP. Forensic Sci Int Genet 4(5):323–328. https://doi.org/10.1016/j.fsigen.2009.12.004

Murnaghan I (2019) RFLP Analysis. Retrieved 15 July 2020, from http://www.exploredna.co.uk/rflp-analysis.html

Murphy E (2013) Legal and ethical issues in forensic DNA phenotyping. New York University School of Law, Public Law & Legal Theory Research Paper Series Working Paper No. 13–46, pp 1–36

Parabon NanoLabs: Engineering DNA for Next-Generation Therapeutics and Forensics (2016). Retrieved 26 July 2020, from https://www.parabon-nanolabs.com/

Parabon® Snapshot® DNA Analysis Service – Powered by Parabon NanoLabs (2020). Retrieved 28 July 2020, from https://snapshot.parabon-nanolabs.com/

Paternoster L, Zhurov A, Toma A, Kemp J, St. Pourcain B, Timpson N et al (2012) Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with Nasion position. Am J Hum Genet 90(3):478–485. https://doi.org/10.1016/j.ajhg.2011.12.021

Phillips C (2015) Forensic genetic analysis of bio-geographical ancestry. Forensic Sci Int Genet 18: 49–65. https://doi.org/10.1016/j.fsigen.2015.05.012

Pierre Piazza (2006) Alphonse Bertillon and the Identification of Persons (1880–1914), Musée Criminocorpus. Published on 26 Aug 2016. Consulted on 13 July 2020 from https://criminocorpus.org/en/ref/25/18341/

Pneuman A, Budimlija Z, Caragine T, Prinz M, Wurmbach E (2012) Verification of eye and skin color predictors in various populations. Legal Med 14(2):78–83. https://doi.org/10.1016/j.legalmed.2011.12.005

Pollack A (2015) Building a Face, and a Case, on DNA. The New York Times. Retrieved from https://www.nytimes.com/2015/02/24/science/building-face-and-a-caseon-dna.html?_r=1 on 17 July 2020

Redler S, Brockschmidt F, Tazi-Ahnini R, Drichel D, Birch M, Dobson K et al (2012) Investigation of the male pattern baldness major genetic susceptibility loci AR/EDA2R and 20p11 in female pattern hair loss. Br J Dermatol 166(6):1314–1318. https://doi.org/10.1111/j.1365-2133.2012.10877.x

Richards J, Yuan X, Geller F, Waterworth D, Bataille V, Glass D et al (2008) Male-pattern baldness susceptibility locus at 20p11. Nature Genetics 40(11):1282–1284. https://doi.org/10.1038/ng.255

Ruiz Y, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares de Cal M, Cruz R et al (2013) Further development of forensic eye color predictive tests. Forensic Sci Int Genet 7(1):28–40. https://doi.org/10.1016/j.fsigen.2012.05.009

Samuel G, Prainsack B (2018) Forensic DNA phenotyping in Europe: views "on the ground" from those who have a professional stake in the technology. New Genet Soc 38(2):119–141. https://doi.org/10.1080/14636778.2018.1549984

Spichenok O, Budimlija Z, Mitchell A, Jenny A, Kovacevic L, Marjanovic D et al (2011) Prediction of eye and skin color in diverse populations using seven SNPs. Forensic Sci Int Genet 5(5):472–478. https://doi.org/10.1016/j.fsigen.2010.10.005

Stranger B, Stahl E, Raj T (2010) Progress and promise of genome-wide association studies for human complex trait genetics. Genetics 187(2):367–383. https://doi.org/10.1534/genetics.110.120907

Sturm R, Duffy D, Zhao Z, Leite F, Stark M, Hayward N et al (2008) A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet 82(2):424–431. https://doi.org/10.1016/j.ajhg.2007.11.005

Sulem P, Gudbjartsson D, Stacey S, Helgason A, Rafnar T, Magnusson K et al (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39(12):1443–1452. https://doi.org/10.1038/ng.2007.13

Toom V (2012) Bodies of science and law: forensic DNA profiling, biological bodies, and biopower. J Law Soc 39(1):150–166. https://doi.org/10.1111/j.1467-6478.2012.00575.x

Valenzuela R, Henderson M, Walsh M, Garrison N, Kelch J, Cohen-Barak O et al (2010) Predicting phenotype from genotype: normal pigmentation. J Forensic Sci 55(2):315–322. https://doi.org/10.1111/j.1556-4029.2009.01317.x

Walsh S, Lindenbergh A, Zuniga S, Sijen T, de Knijff P, Kayser M, Ballantyne K (2011a) Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5(5):464–471. https://doi.org/10.1016/j.fsigen.2010.09.008

Walsh S, Liu F, Ballantyne K, van Oven M, Lao O, Kayser M (2011b) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5(3):170–180. https://doi.org/10.1016/j.fsigen.2010.02.004

Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, Kosiniak-Kamysz A et al (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7(1):98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L et al (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet 9:150–161. https://doi.org/10.1016/j.fsigen.2013.12.006

Weidner C, Lin Q, Koch C, Eisele L, Beier F, Ziegler P et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol 15(2):R24. https://doi.org/10.1186/gb-2014-15-2-r24

Yi S, Xu L, Mei K, Yang R, Huang D (2014) Isolation and identification of age-related DNA methylation markers for forensic age-prediction. Forensic Sci Int Genet 11:117–125. https://doi.org/10.1016/j.fsigen.2014.03.006

Zubakov D, Liu F, van Zelm M, Vermeulen J, Oostra B, van Duijn C et al (2010) Estimating human age from T-cell DNA rearrangements. Current Biology 20(22):R970–R971. https://doi.org/10.1016/j.cub.2010.10.022

Swift DNA Technologies and Their Usefulness for Law Enforcement Agencies

53

James Simpson

Contents

Abstract

Recent advancements in swift forensic DNA technologies have allowed for faster convictions and exonerations within the criminal justice systems. One particular technology, known as "Rapid DNA," allows law enforcement to generate a DNA profile in less than 2 h from a buccal (cheek) swab. These devices have a number of potential applications for law enforcement agencies, as well as for military, forensic, homeland security, and intelligence purposes. While the prospective use of Rapid DNA in criminal investigations has garnered major interest, the increased use of DNA analysis has also led to an increase in DNA profiles uploaded into databases, which has surfaced concerns around privacy and the reliability of the machines being developed by private companies. This chapter will broadly outline the costs and benefits of Rapid DNA as a forensic tool within a law enforcement context.

J. Simpson (✉)
Victoria University of Wellington, Tirohanga, New Zealand
e-mail: simpsojame1@myvuw.ac.nz

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_62

Keywords

Rapid DNA identification · Forensic analysis · Crime-scene analysis · Short tandem repeat · Forensic laboratories

Introduction

The use of Rapid DNA machines in law enforcement is a growing trend internationally. Also referred to as "magic boxes" (Murphy 2019) or "swab in–profile out systems" (Buscaino et al. 2018), this technology operates by producing a DNA profile from blood, saliva, or other biological matter in as little as 90 min, using a simple "sample in–answer out" method (Mapes et al. 2016). Due to their portability and ongoing success, Rapid DNA machines today can be found in a wide variety of places, including at border crossings, embassies, and in traditional forensic laboratories (Butler and Willis 2020). Until recently, conventional laboratory testing conducted by highly trained and qualified personnel was the only option available for analyzing a DNA sample (Della Manna et al. 2016). Today, by some accounts, Rapid DNA machines have the potential to lead to the most significant change in forensic DNA processing since the commencing of the Combined DNA Index System in 1998 (Carney et al. 2019). With the ability to produce swift results, be operated outside the physical boundaries of the traditional laboratory, and be used without any technical expertise, Rapid DNA machines present themselves as an attractive deployment tool for law enforcement personnel. Rapid DNA in the USA received a significant boost in 2017, with the passing of the Rapid DNA act which has fueled law enforcement's enthusiasm for using this technology to analyze the growing backlog of forensic evidence (Murphy 2019). The goals of this enactment were to better inform decisions about pretrial release or detention, solve and prevent crimes, exonerate the innocent, and prevent DNA analysis backlogs (Rapid DNA Act of 2017, Public Law 115-50). At present, Rapid DNA machines can be introduced in three locations within a law enforcement context. These include inside forensic laboratories, within crime scene units, and at police stations (Wilson-Wilde and Pitman 2017). While there is growing evidence that Rapid DNA might play a crucial role in improving the efficiency of law enforcement procedures, there are several important issues raised by this new form of testing which cannot be ignored.

Growing Demand on Forensic Laboratories

For many law enforcement agencies, conducting DNA analysis requires a major commitment of finite resources. While the swift, autonomous process of Rapid DNA testing is still less sensitive than standard laboratory procedures, the growing demand for police investigative units is driving a need for a more efficient process. The turnaround time of urgent crime scene DNA samples is often far longer than desired, which leads to increased pressure on police investigatory units to produce a

turning point in the investigation of the sample (Buscaino et al. 2018). Similarly, laboratories around the world are experiencing a "flood of sample submissions" (Butler 2015, p. 370) in the wake of the growing success of DNA analysis in solving crimes. As a result, backlogs and extended waiting periods have become routine throughout most DNA typing laboratories in the USA (Dash et al. 2020). This alone raises an important issue around unsubmitted samples, with a growing amount of evidence stockpiling in police storage facilities that have not been sent to labs for testing. Because of this, many states have passed laws requiring police to review their evidence rooms for unsubmitted evidence, further increasing laboratory workloads. Illustrating this, the number of untested cases throughout the USA has grown by 85% in the past 6 years (Jackman 2019). The recent uptake of Rapid DNA machines in forensic laboratories and police stations has shown the potential to help fix this growing problem. A recent example of success can be seen throughout Illinois with State Police Director Brendan Kelly recently reporting a 33% overall backlog reduction in DNA samples (Governor's Task Force on Forensic Science Report 2020). It is believed that Rapid DNA implementation played a significant role in achieving this reduction (Spinelli 2020).

Rapid DNA Testing as an Investigative Tool

In response to the growing demand for criminal justice systems, there has unsurprisingly been a strong drive for efficient, portable forensic technology (Morrison et al. 2018). It is believed that the main drivers of demand for this growing need for service have been the increased amount of DNA evidence that is collected in criminal cases, as well as the expanded effort to collect DNA samples from convicted felons and arrested persons (National Institute of Justice 2010). Perhaps the most salient and documented benefit of Rapid DNA systems is the speed at which they can analyze a sample. Rapid DNA drastically cuts down wait times which could otherwise be weeks to a period of under 2 h. Fast results are of particular importance in homicide cases, where the first 24-h period is crucial to an investigation (Wilson-Wilde and Pitman 2017). The value of real-time forensic information produced by Rapid DNA is well summarized by Van Asten, who argues that the availability of fast data directs the investigation and allows police to use their "scarce resources efficiently and effectively, ultimately solving more crimes" (p. 33).

Turnaround Time

Rapid DNA machines allow law enforcement officials to generate rapid intelligence, which provides great benefits to solving investigations. Such machines have been specifically developed in a way to combine the four stages of extraction, amplification, separation, and detection of a DNA sample, all without the need for human input. This autonomous process operates in contrast to the conventional, labor-intensive processes of laboratory analysis procedures, which can often take

months to produce a profile (Steward 2016). In many cases, a single sample can last 5 months or longer in storage before being turned around (Neuhauser 2019). In response to this growing issue, Rapid DNA devices have become quickly adopted with an overarching objective to help reduce this backlog of samples. So far, Rapid DNA systems have received praise for their efficiency from law enforcement officials and academics alike, including the US National Association of Police Organizations. The speed of Rapid DNA analysis is beneficial for law enforcement in several ways. As well as decreasing backlogs, Rapid DNA systems also have the potential to provide an early lead in the investigation, which is particularly beneficial for police who are constantly limited by their finite resources during investigations. This is especially relevant for New Zealand's current situation, as failure to investigate was the number one complaint made against New Zealand Police in recent times. Processing a sample in real time during an investigation is additionally beneficial because stations typically have a 24–96 h holding window on suspects before having to formally charge them of a crime, or otherwise release them. The testing can work to help ensure the apprehended individual is not linked to any unsolved crimes before their release.

Another benefit of Rapid DNA testing within law enforcement is its applicability to a wide range of different crime types. Although originally touted for its potential to revolutionize rape investigations (Schuppe 2019), Rapid DNA's applicability to lower-level crimes has also been noted by academics who argue that the need for fast DNA analysis is actually of less importance for significant crimes. This is because a suspect of a serious crime would likely be remained in police custody, regardless of available DNA evidence. For less significant crimes, however, there is the possibility that a cold link to a serious matter could be obtained, which better allows investigators to link crimes and could influence decisions around whether to retain the person in custody. Studies from Wilson et al. (2010) also allude to the benefits of Rapid DNA testing in burglary cases by highlighting the noteworthy percentage increase of a successful identification across routinely low-clearance types of offending, such as burglary. This has also been demonstrated by a series of randomized controlled trials by Roman et al. (2008) across the USA, showing a two- to threefold increase in the percentage of a successful resolution for property crimes when analyzing collected DNA from the crime scene.

Transportability

Transportability is another key advantage of Rapid DNA devices. Historically speaking, forensic DNA identification methods have suffered from high latency and low portability (Zaaijer et al. 2017), meaning that analysis has been limited to specialist laboratories and often suffers from long delays in producing information. With the ongoing success of forensic DNA analysis, however, it is not surprising that the use of Short Tandem Repeats is broadening to more efficient measures beyond the laboratory (Grover et al. 2017). Khanna et al. (2020) note that the turnaround time for processing a sample is further reduced by the capacity for Rapid DNA

devices to be stored at police stations, or even be taken at the scene, as opposed to having to send evidence away to a laboratory. About the size of a large desktop printer (Jackman 2018) and weighing 60 pounds (Tucker 2015), their portability factor means that the devices can also be used by first responders in the field for a wide range of disasters, such as earthquakes or other mass casualty events to identify victims. Indeed, this is an area that Rapid DNA machines have already proven to be highly effective in. In particular, Rapid DNA played a significant role in victim identification during the 2018 Californian wildfires. Nevertheless, despite their wider successes in a broad array of circumstances, some express concern that the devices could be used for purposes beyond their intended means. For example, FBI chief biometric scientist Thomas Callaghan believes Rapid DNA machines are currently best suited to analyzing large amounts of DNA from one person, as opposed to crime scene use which commonly contains the DNA of multiple people (Jackman 2018).

Contamination

Contamination of a DNA sample is a major issue in forensics, as it can result in greatly reduced genotype quality for sequencing studies (Jun et al. 2012). The limited existing literature has shown mixed results regarding the risks of contamination when using Rapid DNA devices. Scholars like Busciano et al. (2018) have alluded to the diminished risk of contamination from the use of these devices, although this is contested (see Dolan 2019). On the one hand, the autonomous processing of a sample decreases the likelihood of external contamination, as it requires no human input. Zhang et al. endorse this line of thought, arguing that by autonomously combining the steps of DNA analysis, the risk of analysts accidentally introducing their DNA is eliminated. On the other hand, the reality that non-laboratory professionals are now responsible for gathering and processing samples through the devices increases the risk of polluting or damaging the sample. This is because they lack the extensive education and training with forensic equipment that laboratory professionals have. We know that crime-related traces may easily be confused with traces produced by events before and after the crime. Further to this, the first extraction phase is the step where the DNA sample is most susceptible to contamination. If a sample is not installed correctly and/or becomes contaminated beforehand, this runs the risk of leading police to the wrong person in an investigation or jeopardizing the case as a whole. In one instance, incorrectly processed DNA led to the incrimination of an impossible suspect, a man who had been dead for 2 years (Murphy 2015). Such inaccuracies have also charged at least one innocent man with capital murder (Worth 2018). It is also crucial to note that many Rapid DNA machines are currently stored in offices as opposed to sterile laboratories (Balk 2015), which in itself heightens the risk of contaminating the sample. One criticism pointed out by Salceda et al. (2017) is that most Rapid DNA machines can only process one sample at a time, which suggests that their applicability to crime scenes may be limited. This is because such environments often

contain a mixture of DNA from several individuals, thus increasing the likelihood of sample contamination or a false match. This is noted by Romsos et al. (2020), who advocates for Rapid DNA's use strictly for single-source samples only. In sum, the limited set of existing literature is inconclusive regarding the risk of contamination when using Rapid DNA devices. However, even traditional DNA analysis undertaken by professionals in recognized laboratories have led to drastic errors. Whether or not the expansion of DNA testing beyond the laboratory increases the number of inaccuracies remains uncertain. Future success and accuracy rates of Rapid DNA will largely depend on the operators of these devices receiving high-quality training in evidence collection methodology.

Key Concerns

Privacy

At the heart of most concerns with DNA in criminal investigations are issues around privacy. Academics such as Murphy (2018) and Butler (2015) have raised concerns about privacy, equality, and abusive government surveillance stemming from autonomous forensic DNA testing. In the USA, recent law changes have allowed police officers to obtain DNA samples from arrestees without the need for a warrant (Maryland vs King 2013). Due to the high degree of law-abiding citizens who have encounters with police, the increasing use of Rapid DNA systems puts innocent people at risk of prosecution based on their DNA, eroding the presumption of innocence (Quick 2019). Showcasing their impact in recent times, it is estimated that forensic DNA analysis has identified more than 440,000 previously unknown suspects over the past 20 years (Jackman 2019). Although evidence exists that the growth of DNA databases does have a net deterrent effect on convicted offenders (Doleac 2017), several concerns are raised by Rapid DNA in law enforcement and the subsequent expansion of forensic databases. As the growing number of profiles in databases increases, the likelihood of finding an offender, the chances of inaccuracies and false matches also rise. Some have commented that the enthusiastic, increasing deployment of Rapid DNA machines threatens the tradition of protecting the privacy and rights of individuals who would not otherwise be under suspicion (Flaus 2013). While the goal of clearing backlogs has achieved some success already, there is concern around the increasing use of DNA within the criminal justice system resulting in a high number of innocent people's information being stored in databases. With record-high reports of cybercrime in recent times (Healy and Mcgrath 2019), the threat of this data being hacked could lead to serious privacy breaches. In Samuel and Prainstack's (2019) study around stakeholder perceptions of forensic DNA profiling, around half of the interviewees described FDP as "intrusive" or "invasive." Furthermore, it has been pointed out that these machines are also likely to incentivize the growth of rogue DNA databases, which are maintained with far fewer quality, privacy, and security controls than federal databases (Eidelman and Stanley 2019). Also, others have expressed worry that providing genetic material to

the government allows unscrupulous analysts to reveal health conditions and other sensitive information (Doleac 2017). The increasing use of forensic DNA testing raises important ethical and social policy discussion points around how much authority we give law enforcement officials in these types of scenarios.

Cognitive Challenges

Rapid DNA is widely considered to be accurate, though imperfect (Kloosterman et al. Kloosterman et al. 2014). At present, it is still in the early stages of being accepted as a reliable alternative to standard DNA tests (Cino 2017). In 2018, testing found that Rapid DNA devices were accurate approximately 85% of the time, with the potential for increased accuracy when operated by laboratory experts. Its applicability to a law enforcement context has faced some criticism for the "time/success rate trade-off" that Police and Scene of Crime Officers face when using this technology. In law enforcement agencies that use Rapid DNA machines, Scene of Crime Officers are confronted with a dilemma of whether to utilize this fast-acting technology or forward the crime sample to the laboratory, where the rate of accuracy will be greater. It remains the case that this technology is less still sensitive than traditional methods carried out at the forensic laboratory (Mapes et al. 2016). Because of this, Rapid DNA instruments are not, at present, sufficiently matured for use in analyzing crime scene evidence for automatic submission to CODIS (Hares et al. 2020). In one previous study, the sensitivity rate between laboratory and Rapid DNA testing differentiated by 5% in favor of laboratory testing (Schroeder and White 2009) when examining a sample from a ski mask. Rapid DNA profiling has, however, proved to be especially accurate in certain scenarios, such as trace detection of bloodstains at crime scenes (Vittori et al. 2016). Cino (2017) advises that the devices are currently best suited to quickly rule out or further investigate an individual suspected of a crime. She further recommends that a prosecutor seeking to use DNA evidence during a trial should be required to revert to traditional laboratory analysis to argue his/her case in the courtroom. This is because the validity of Rapid DNA still could be brought into question during a courtroom case, simply due to its novelty (Seaone 2019). Although this technology has made considerable improvements since its conception, it is not expected that rapid analyses will reach the precision of laboratory testing any time soon (Mapes et al. 2019), meaning that this cognitive dilemma faced by law enforcement officials is likely to persevere. However, due to the urgent nature of criminal investigations, we can expect frontline officers to favor rapidity over sensitivity when considering the high demand for their services (Mapes 2017). Helsloot and Groenendaal (2011) highlight that the internal factor of emotion can greatly impact an officer's discretionary powers in deciding to employ a Rapid DNA test, especially in instances where the "desire to see justice done" (p. 897) is stronger. Furthermore, early studies from Tversky and Kahneman (1974) have argued that people are inclined to act quickly when faced with uncertain situations, which can lead to errors when an officer's categorizations are incorrect (Gigerenzer et al. 1999).

Function Creep

Ease of use combined with affordability is a recipe for overuse or "function creep." In a society that favors certainty over accuracy (Stevens 2013), there is a high potential for this technology to be overused by law enforcement officials, or be used for reasons beyond their intended purpose. Unlike skilled laboratory technicians, frontline officers who use Rapid DNA machines are trained for a couple of hours and lack any distinct protocols when using the devices (Hawkins 2020). Although Rapid DNA testing is autonomous, the human factor plays an important role even when complete trace analyses are conducted by machines (De Gruijter and de Poot 2019). This is especially relevant as DNA is often considered the "gold standard" of a forensic investigation, making officers even more eager to obtain incriminating evidence from a suspect. Furthermore, as Gruijter et al. (2016) suggest, the use of Rapid DNA devices at crime scenes could potentially cause a shift in investigatory focus toward finding and analyzing alleged perpetrator-related traces, resulting in other crime-related traces being overlooked. The direction of an investigation could be influenced if crime scene officers prioritized the search for specific traces that are compatible with Rapid DNA machines. This narrow focus would mean that noncompatible traces might be labeled as irrelevant (Wyatt 2014) if they do not fit the requirements for a Rapid DNA device. De Gruijter and colleagues make note of several miscarriages of justice stemming from the negative consequences of "tunnel vision" (p. 850) during an investigation. It remains the case that crime scenes can and should be examined in multiple ways (De Gruijter et al. 2017). The measures taken to protect against tunnel vision or contamination mean that some work processes may be set up less efficiently, for example, more protocolled, to avoid the risks of bias (Groenendaal and Helsloot 2015).

Another key concern of Rapid DNA's use in law enforcement is that database-matches may have too much influence on an investigation, leading to scenarios being based on database-matches without considering the relevance of the traces that did provide a match. This is especially problematic considering the high volume of nonoffenders who may have a profile in Combined DNA Index Systems. If used improperly, the potential for error and breach of public trust is extremely high (Hoey 2019). Weiss (2020) summarizes that the future may be bright for Rapid DNA systems, but only if quality and integrity are maintained at both organizational and frontline levels. While Rapid DNA by design is geared toward efficiency and simplicity, they do present new cognitive challenges for users around whether to employ the machines and to what extent they should direct an investigation. Little research has been conducted on CSI decision-making (De Gruijter et al. 2016), yet it is arguably more important now than ever to understand the growing development of potentially behavior-changing technologies. The accuracy of Rapid DNA is expected to increase with time, yet it should not be assumed that the benefits of forensic DNA will necessarily override the social and ethical costs (Roewer 2013). Looking into and addressing some of the cognitive challenges faced by frontline employees at a training level should not be ignored as a route to future success.

Going Forward

To weigh these issues against the potential value of Rapid DNA machines, several approaches should be considered. Firstly, quality of training is essential as it is crucial for law enforcement officials to recognize the full capabilities of Rapid DNA testing, including its potential downsides. Operators of these devices should be competent not only in a technical manner but also in ensuring the nondisclosure of private information from the results produced. Revisiting the IACP (2001) study on the implementation of police in-car cameras, researchers argued that "as with any new technology, failure to properly train officers in its use, operation, and legal implications of improper use can result in disaster" (p. 19). While Scene of Crime Officers (SOCOs) have strict procedures and protocols in place for collecting evidence, Wyatt (2014) observes that such investigators routinely go beyond the provided guidelines and utilize their discretion in decision-making. Mapes et al. (2019) similarly note that the majority of SOCOs' decisions in practice are likely based on intuition. Therefore, it is crucial to recognize that policy or protocol alterations do not guarantee changes in operational behavior. Instead, positive changes should look to be driven by proper and regular training procedures.

The increasing use of swift DNA technologies could also facilitate the growth of other forensic DNA techniques such as familial searching, which similarly has shown the potential to be a valuable investigative tool. Perhaps in the future, we could see the development of machines that can combine rapid technologies with other forensic analysis results, such as familial searching and phenotyping. However, growth in any form of forensic DNA analysis further expands databases to include innocent individuals. If forensic DNA systems are employed more frequently, public trust in the technology is vital (Samuel and Prainsack 2019). It is also necessary that the high levels of trust extend to law enforcement officials, who in recent times have come under intense public scrutiny by citizens, politicians, and advocacy groups due to allegations of misconduct (Ufford 2019). Today, growing calls to "defund the police" have emerged, largely due to ongoing incidents of excessive use of force and alleged racial bias of officers on duty. This has spotlighted the perceptions of many that police officers are untrustworthy, with consistent media portrayals of tense confrontations between police officers and citizens. Stemming from the high levels of distrust in recent times, especially from minority groups, there have been greater calls for oversight and review of police decisions while on duty. To be effective in their duties, law enforcement officials are highly dependent on public support and cooperation (Skogan and Frydl 2004). Concerns around differential policing and officer misconduct when using Rapid DNA could be addressed by body-worn camera use for officers. As frontline officers largely work without direct supervision, there is potential for their high levels of discretion to be used in a discriminatory way, which could undermine police legitimacy (Lum 2011). With the absence of empirical evidence, it is unlikely that such officers will be held accountable for their actions. However, studies have demonstrated that the presence of video recording technology in law enforcement may have a "civilizing effect" on police-citizen encounters (Ariel et al. 2015). Others report behavioral improvements in police

officers from their deployment (Ufford 2019). Furthermore, officers equipped with this technology are shown to be significantly less likely to have a complaint sustained against them (Katz et al. 2015; Ariel 2016). Such findings fundamentally relate to the Hawthorne Effect Theory, which argues that individuals behave differently when they are being watched (Adair 1984). The adoption of body-worn cameras could be a beneficial accountability tool to ensure that law enforcement officials are using swift DNA technologies in a legitimate manner and context. Unfortunately, many law enforcement agencies are hindered by financial constraints in adopting the latest technologies. Despite this, the costs of running a sample through a Rapid DNA machine are expected to reach all-time lows in the near future. For instance, Weiss (2020) estimates that Rapid DNA will soon be estimated at $100 per sample, compared to $500 in a forensic laboratory. However, it is important to recognize that the cost of swift DNA technologies does not just relate to the equipment and consumables but also the infrastructure (e.g., waste disposal) and logistical maintenance (Wilson-Wilde and Pitman 2017).

Conclusion

Rapid DNA, even in its developing state, is useful. In the past, successful DNA testing has relied on the skill of experienced laboratory analysts. Today, Rapid DNA is an exciting adaption to help law enforcement and criminal justice meet the growing needs of communities. The significance of this new technology to aid in criminal investigations largely derives from its efficiency, portability, and ease of use. However, research into its effectiveness in a law enforcement context is still in its infancy, with unanswered questions remaining around its accuracy and applicability to forensic investigations. Additional concerns lie with the cognitive challenges and bias that law enforcement officials may face when choosing to employ the technology. Perhaps the biggest key to the future success of Rapid DNA will involve ensuring that adequate training is provided for officers who may be charged with the operation of the devices. Over the next decade, it is expected that DNA testing will become more rapid, more informative, and more sensitive (Butler 2015). As the amount of time taken to analyze a sample continues to decrease, we might see the expansion of Rapid DNA machines to other areas of law enforcement, for instance, in police vehicles. To avoid future scrutiny around ethical and privacy concerns, it is of crucial importance that these technological advancements do not become used as a surveillance tool rather than for their original intended purpose as an aid for criminal investigations.

References

Adair JG (1984) The Hawthorne effect: A reconsideration of the methodological artifact. J Appl Psychol 69(2):334–345. https://doi.org/10.1037/0021-9010.69.2.334

Ariel B (2016) Increasing cooperation with the police using body worn cameras. Police Q 19(3): 326–362. https://doi.org/10.1177/1098611116653723

Ariel B, Farrar WA, Sutherland A (2015) The effect of police body-worn cameras on use of force and citizens' complaints against the police: A randomized controlled trial. J Quant Criminol 31: 509–535. https://doi.org/10.1007/s10940-014-9236-3

Balk C (2015) Reducing contamination in forensic science. Themis Res J Just Stud Forensic Sci 3: Article 12. 11

Buscaino J, Barican A, Farrales L, Goldman B, Klevenberg J, Kuhn M, Lin F, Nguyen P, Salceda S, Schueren R, Smith C, Troup C, Tsou D, Vangbo M, King D (2018) Evaluation of a rapid DNA process with the RapidHIT® ID system using a specialized cartridge for extracted and quantified human DNA. Forensic Sci Int Genet 34:116–127. https://doi.org/10.1016/j.fsigen.2018.02.010

Butler JM (2015) The future of forensic DNA analysis. Philos Trans R Soc Lond Ser B Biol Sci 370(1674):20140252. https://doi.org/10.1098/rstb.2014.0252

Butler JM, Willis S (2020) Interpol review of forensic biology and forensic DNA typing 2016–2019. Synergy, Forensic Science International

Carney C, Whitney S, Vaidyanathan J, Persick R, Noel F, Vallone PM, Romsos EL, Tan E, Grover R, Turingan RS, French JL, Selden RF (2019) Developmental validation of the ANDE™ rapid DNA system with FlexPlex™ assay for arrestee and reference buccal swab processing and database searching. Forensic Sci Int Genet 40:120–130. https://doi.org/10.1016/j.fsigen.2019.02.016

Cino J (2017) Tackling technical debt: managing advances in dna technology that outpace the evolution of law. Am Crim Law Rev 54(2):373–422

Dash H. R., Shrivastava P., & Das S. (2020). RAPID DNA technology: A boon to forensic DNA typing. In Principles and practices of DNA analysis: A laboratory manual for forensic DNA typing. Springer protocols handbooks. Humana, New York, pp 313–316.

De Gruijter M, de Poot CJ (2019) Cognitive challenges at the crime scene: the importance of social science research when introducing mobile technologies at the crime scene. Forensic Sci Int 297: e16–e18. https://doi.org/10.1016/j.forsciint.2019.01.026

De Gruijter M, de Poot CJ, Elffers H (2016) The influence of new technologies on the visual attention of CSIs performing a crime scene investigation. J Forensic Sci 61(1):43–51. https://doi.org/10.1111/1556-4029.12904

De Gruijter M, Nee C, De Poot CJ (2017) Rapid identification information and its influence on the perceived clues at a crime scene: an experimental study. Sci Justice 57(6):421–430

Della Manna A, Nye JV, Carney C, Hammons JS, Mann M, Al Shamali F, Vallone PM, Romsos EL, Marne BA, Tan E, Turingan RS, Hogan C, Selden RF, French JL (2016) Developmental validation of the DNAscan™ rapid DNA analysis™ instrument and expert system for reference sample processing. Forensic Sci Int Genet 25:145–156. https://doi.org/10.1016/j.fsigen.2016.08.008

Dolan M (2019) Rapid DNA' promises breakthroughs in solving crimes. So why does it face a backlash? Los Angeles Times. Retrieved 27 September, 2020 from https://www.latimes.com/california/story/2019-09-24/rapid-dna-forensics-crime-police

Doleac J (2017) The effects of DNA databases on crime. Am Econ J Appl Econ 9(1):165–201. Retrieved September 26, 2020, from http://www.jstor.org/stable/26156212

Eidelman V, Stanley J (2019) Rapid DNA machines in police departments need regulation. Retrieved September 28, 2020, from https://www.aclu.org/blog/privacy-technology/medical-and-genetic-privacy/rapid-dna-machines-police-departments-need

Flaus A (2013) Familial searches and the New Zealand DNA profile databank: the thin edge of the genetic wedge? University of Otago. Retrieved 26 September, 2020, from https://www.otago.ac.nz/law/research/journals/otago065282.pdf

Gigerenzer G, Todd PM, ABC Group (1999) Simple heuristics that make us smart. Oxford University Press, New York

Governor's Task Force on Forensic Science Report (2020) ISP. Retrieved 26 September 2020 from http://www.isp.state.il.us/docs/6-740%20-%20For%20Sci%20Draft%20Report%206-1-20.pdf

Groenendaal J, Helsloot I (2015) Tunnel vision on tunnel vision? A preliminary examination of the tension between precaution and efficacy in major criminal investigations in the Netherlands. Police Pract Res 16(3):224–238. https://doi.org/10.1080/15614263.2014.928622

Grover R, Jiang H, Turingan RS et al (2017) FlexPlex27—highly multiplexed rapid DNA identification for law enforcement, kinship, and military applications. Int J Legal Med 131:1489–1501. https://doi.org/10.1007/s00414-017-1567-9

Gruijter MD, Poot CD, Elffers H (2016) Reconstructing with trace information: does rapid identification information lead to better crime reconstructions? J Investig Psychol Offender Profiling 14(1):88–103. https://doi.org/10.1002/jip.1471

Hares DR, Kneppers A, Onorato AJ, Kahn S (2020) Rapid DNA for crime scene use: enhancements and data needed to consider use on forensic evidence for state and national DNA Databasing-an agreed position statement by ENFSI, SWGDAM and the rapid DNA Crime Scene Technology Advancement Task Group. Forensic science international. Genetics 48, 102349

Hawkins J (2020) Police use of rapid DNA machines unregulated. Criminal Legal News. Retrieved 26 September, from https://www.criminallegalnews.org/news/issue/3/1/

Healy D, Mcgrath J (2019) Simple rhetoric and complex punitiveness: federal criminal justice responses to White-Collar criminality. Justice Q 36(7):1258–1283. https://doi.org/10.1080/07418825.2019.1628289

Helsloot I, Groenendaal J (2011) Naturalistic decision making in forensic science: toward a better understanding of decision making by forensic team leaders. J Forensic Sci 56(4):890–897. https://doi.org/10.1111/j.1556-4029.2011.01714.x

Hoey BK (2019) Rapid DNA initiative law enforcement external. Forensic Sci Int Synergy 1. https://doi.org/10.1016/j.fsisyn.2019.02.021

International Association of Chiefs of Police (2001) The impact of video evidence on modern policing. Rep. No. 2001-CK-WX-0157. Washington, DC, pp 1–40. Retrieved from https://ric-zai-inc.com/Publications/cops-w0404-pub.pdfI

Jackman T (2018) FBI plans 'rapid DNA' network for quick database checks on arrestees. Washington Post. Accessed 18th Sep 2020. https://www.washingtonpost.com/crime-law/2018/12/13/fbi-plans-rapid-dna-network-quick-database-checks-arrestees/

Jackman T (2019) Nationwide DNA testing backlog has nearly doubled, despite $1 billion in federal funding. Retrieved September 26, 2020, from https://www.washingtonpost.com/crime-law/2019/03/23/nationwide-dna-testing-backlog-has-nearly-doubled-despite-billion-federal-funding/

Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M, Kang HM (2012) Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet 91(5):839–848. https://doi.org/10.1016/j.ajhg.2012.09.004

Katz M, Choate DE, Ready JR, Nuño L (2015) Evaluating the impact of officer worn body cameras in the Phoenix police department. Center for Violence Prevention & Community Safety, Arizona State University, Phoenix, pp 3–4

Khanna G, Mishra K, Singh R (2020) Collection and preservation of DNA evidence. Acad J Forensic Sci 3(1) ISSN: 2581-4273

Kloosterman A, Sjerps M, Quak A (2014) Error rates in forensic DNA analysis: definition, numbers, impact and communication. Forensic Sci Int Genet 12:77–85. https://doi.org/10.1016/j.fsigen.2014.04.014

Lum C (2011) The influence of places on police decision pathways: from call for service to arrest. Justice Q 28(4):631–665. https://doi.org/10.1080/07418825.2010.526130

Mapes A (2017) Rapid DNA technologies at the crime scene 'CSI' fiction matching reality (Unpublished doctoral dissertation). Proefschrift Universiteit van Amsterdam, pp 80–109

Mapes AA, Kloosterman AD, Poot CJ, van Marion V (2016) Objective data on DNA success rates can aid the selection process of crime samples for analysis by rapid mobile DNA technologies. Forensic Sci Int 264:28–33. https://doi.org/10.1016/j.forsciint.2016.03.020

Mapes AA, Stoel RD, de Poot C, Vergeer P, Huyck M (2019) Decision support for using mobile rapid DNA analysis at the crime scene. Sci Justice 59(1):29–45. https://doi.org/10.1016/j.scijus.2018.05.003

Maryland vs King, 569 (2013) U.S Supreme Court. No. 12–207. Retrieved 26 September, 2020 from https://www.law.cornell.edu/supremecourt/text/12-207

Morrison J, Watts G, Hobbs G, Dawnay N (2018) Field-based detection of biological samples for forensic analysis: established techniques, novel tools, and future innovations. Forensic Sci Int 285:147–160. https://doi.org/10.1016/j.forsciint.2018.02.002

Murphy E (2015) How DNA evidence incriminated an impossible suspect. The New Republic. Retrieved 25 September 2020 from https://newrepublic.com/article/123177/how-dna-evidence-incriminated-impossible-suspect

Murphy E (2018) Forensic DNA typing. Ann Rev Criminol 1(1):497–515. https://doi.org/10.1146/annurev-criminol-032317-092127

Murphy H (2019) Coming soon to a police station near you: the DNA 'Magic Box' – with rapid DNA machines, genetic fingerprinting could become as routine as the old-fashioned kind. But forensic experts see a potential for misuse. The New York Times. Retrieved 11 July 2020

National Institute of Justice (2010) Why the demand for DNA testing is increasing. Retrieved 26 September, 2020 from https://nij.ojp.gov/topics/articles/why-demand-dna-testing-increasing

Neuhauser A. (2019). DNA evidence backlog has soared by 85 percent. Retrieved September 26, 2020, from https://www.usnews.com/news/national-news/articles/2019-03-22/dna-evidence-backlog-has-soared-by-85-percent

Quick E (2019). Police departments adopt automated DNA testing. The regulatory review. Retrieved 26 September 2020, from https://www.theregreview.org/2019/06/18/quick-police-departments-adopt-automated-dna-testing/

Rapid DNA Act of 2017: Report (to accompany H.R. 510) (including cost estimate of the Congressional Budget Office). (2017). U.S. Government Publishing Office, Washington, DC

Roewer L (2013) DNA fingerprinting in forensics: past, present, future. Investig Genet 4(1):22. https://doi.org/10.1186/2041-2223-4-22

Roman JK, Reid S, Reid J, Chaflin A, Adams W, Knight C (2008) The DNA field experiment: cost-effectiveness analysis of the use of DNA in the investigation of high-volume crimes. Urban Instittue—Justice Policy Center, Washington, DC

Romsos EL, French JL, Smith M, Figarelli V, Harran F, Vandegrift G, Moreno LI, Callaghan TF, Brocato J, Vaidyanathan J, Pedroso JC, Amy A, Stoiloff S, Morillo VH, Czetyrko K, Johnson ED, de Tagyos J, Murray A, Vallone PM (2020) Results of the 2018 rapid DNA maturity assessment. J Forensic Sci 65(3):953–959. https://doi.org/10.1111/1556-4029.14267

Salceda S, Barican A, Buscaino J, Goldman B, Klevenberg J, Kuhn M, Lehto D, Lin F, Nguyen P, Park C, Pearson F, Pittaro R, Salodkar S, Schueren R, Smith C, Troup C, Tsou D, Vangbo M, Wunderle J, King D (2017) Validation of a rapid DNA process with the RapidHIT® ID system using GlobalFiler® express chemistry, a platform optimized for decentralized testing environments. Forensic Sci Int Genet 28:21–34. https://doi.org/10.1016/j.fsigen.2017.01.005

Samuel G, Prainsack B (2019) Civil society stakeholder views on forensic DNA phenotyping: balancing risk and benefits. Forensic Sci Int Genet 43:102157

Schroeder A, White D (2009) Exploring the use of DNA evidence in homicide investigations: implications for detective work and case clearance. Police Q 12(3):319–342

Schuppe J (2019) Rapid DNA testing could revolutionise rape investigations – if the science holds up. ABC News. https://www.nbcnews.com/news/us-news/rapid-dna-testing-could-revolutionize-rape-investigations-if-science-holds-n1096516

Seaone S (2019) What is rapid DNA? GovThink. Retrieved 29 September, 2020 from https://www.govthink.com/2019/11/what-is-rapid-dna/

Skogan WG, Frydl K (2004) Fairness and effectiveness in policing: the evidence. National Academies Press, Washington, DC

Spinelli C (2020) Illinois officials announce 33 percent reduction in state forensic backlog. Retrieved September 26, 2020, from https://www.kwqc.com/2020/07/01/illinois-officials-announce-33-percent-reduction-in-state-forensic-backlog/

Stevens A (2013) Telling policy stories: An ethnographic study of the use of evidence in policy-making in the UK. J Social Policy 40:237–255. https://doi.org/10.1017/S0047279410000723

Steward ER (2016) Discussion and evaluation: the legality and use of rapid DNA technologies. UMKC Law Rev 84(4):1133–1162

Tucker P (2015) Special operators are using rapid DNA readers. Defense One. Retrieved 26 September, 2020 from https://www.defenseone.com/technology/2015/05/special-operators-are-using-rapid-dna-readers/113383/

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. Science (New York, NY) 185(4157):1124–1131

Ufford SP (2019) Police officers' perceptions of changes in their behavior while being video recorded (Doctoral dissertation, Walden University), pp 1–22

Vittori ED, Barni F, Lewis SW, Antonini G, Rapone C, Berti A (2016) Forensic application of a rapid one-step tetramethylbenzidine-based test for the presumptive trace detection of blood-stains at the crime scene and in the laboratory. Forensic Chem 2:63–74. https://doi.org/10.1016/j.forc.2016.10.002

Weiss D (2020) Rapid DNA analysis a new tool for forensic laboratories. In Public safety. Retrieved 25 September from https://inpublicsafety.com/2020/08/rapid-dna-analysis-a-new-tool-for-forensic-laboratories/

Wilson DB, McClure D, Weisburd D (2010) Does forensic DNA help to solve crime? The benefit of sophisticated answers to naive questions. J Contemp Crim Justice 26(4):458–469. https://doi.org/10.1177/1043986210377231

Wilson-Wilde L, Pitman F (2017) Legislative and policy implications for the use of rapid DNA Technology in the Australian Context. Forensic Sci Policy Manage Int J 8(1–2):26–37. https://doi.org/10.1080/19409044.2017.1335809

Worth K (2018) Framed for murder by his own DNA. Wired. Retrieved 25 September from https://www.wired.com/story/dna-transfer-framed-murder/

Wyatt D (2014) Practising crime scene investigation: trace and contamination in routine work. Polic Soc 24(4):443–458. https://doi.org/10.1080/10439463.2013.868460

Zaaijer S, Gordon A, Speyer D, Piccone R, Groen SC, Erlich Y (2017) Rapid re-identification of human samples using portable DNA sequencing. eLife, 6, e27798. https://doi.org/10.7554/eLife.27798

Tracing of Human Migration and Diversity by Forensic DNA Analysis

54

Nithyanandam Mahalaxmi, Avinash Chand Puri, Pawan Kumar Chouhan, and Alka Mishra

Contents

Abstract

Human migrations have changed the face of lands and continents, racial, ethnic, and linguistic composition of their populations. Many people perish on perilous treks around the world in search of a better life. However, there is not much being done to track down these people. DNA profiling has made it possible to track these migrants with its advanced techniques and established protocols. DNA profiling is still a gold standard technology for human identification. In the last three decades, genetic research has become increasingly applicable in the study of human history. Several advancements in our knowledge of human migration have

N. Mahalaxmi
Forensic Sciences Department, Chennai, Tamil Nadu, India

A. C. Puri (✉) · P. K. Chouhan
Regional Forensic Science Laboratory, Indore, Madhya Pradesh, India

A. Mishra
Regional Forensic Science Laboratory, Gwalior, Madhya Pradesh, India

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7_65

been made possible by developments in the sequencing and study of both current and ancient peoples' genomes. They have helped to prove that anatomically modern humans initially arose in Africa around 250,000–350,000 years ago and then spread to other provinces of the world, and the humans' genetic adaptations to local environmental conditions are among them. Analyses of this original genetic data are transforming our understanding of humans' migration history and adaption. We may now collect whole-genome sequences from a variety of existing and prehistoric Africans using today's sequencing technologies. In the coming years, significant discoveries will be made in the study of human migration.

Keywords

Haplotypes · Mitochondrial DNA · Y chromosome marker · Human migration

Introduction

When Alec Jeffreys of the University of Leicester in the United Kingdom discovered remarkably variable and heritable patterns from repetitive DNA studies using multilocus probes, his "Eureka" cry shook England and was heard across the world. He named this technique "DNA fingerprinting," a first breakthrough that opened a new area of science (Jeffreys et al. 1985; Roewer 2013). Genetic analysis, in its most basic form, entails comparing DNA from different groups of people, such as those with or without a specific type of disease or people from different parts of the world. Since the sequencing of the complete human genome, our ability to sequence DNA has improved considerably (Johnston et al. 2019). We do not need to look at all of the three billion letters in the human genome because persons differ by only one letter in a thousand on average. Instead, we can compare persons who have known genetic differences, which are referred to as genetic markers. Millions of these markers have been identified, and this, along with genetic sequencing technology that allows us to look at these markers in a large number of people at a low cost, has resulted in a massive increase in the amount of data available to geneticists. Dr. Kary Mullis, a scientist at the Cetus Corporation in Emeryville, California, created the polymerase chain reaction (PCR) in 1985, which was the second major advance in molecular biology. By allowing for the amplification of specific sections of highly polymorphic DNA, this stunningly simple technology revolutionized forensic DNA analysis. This not only gave forensic analysts the sensitivity they needed to extract valuable DNA profiles from difficult forensic material, but it also laid the groundwork for the many generations of DNA-based typing procedures that are now employed in forensic laboratories.

Contemporary biology has given scientists strong tools for reforming the past of the Earth and its residents, including humans. Over the last 25 years, experts have endorsed the multiregional continuity hypothesis (Thorne and Wolpoff 1992) that modern humans left Africa around 350,000 years ago as a single developing species,

Homo sapiens, and expanded over the world (Schlebusch and Jakobsson 2018). The formation of recognized regional morphologies in the continents of Africa, Europe, and Asia is justified by some skeletal traits that originated and lasted for differing periods in distinct geographical regions. The "recent out of Africa" approach (Wilson and Cann 1992), on the other hand, claimed that since people began to radiate out of Africa, there have been emergent patterns. This scenario similarly claims that *H. sapiens* first appeared in Africa some 100,000 years ago and quickly spread over the world, displacing other species. During the previous one and a half decade, molecular evidence from populations from many ethnic groups around the world has made a significant contribution to the discussion. Rapid discoveries in DNA sequencing, on the other hand, have opened up a whole new window into the past, showing that human history is more complicated than previously thought. In fact, recent DNA investigations have revealed that humans have migrated and intermingled far more than previously assumed – particularly over the last 10,000 years. In this chapter, it is aimed to shed some light on the crucial associations of DNA profiling in human migrations around the world.

DNA as a Palimpsest of Human Migration

When Alec Jeffreys identified considerable individual variation between persons in certain areas of DNA using the variable number of tandem repeats (VNTRs) and restriction fragment length polymorphism (RFLP) analysis in 1985, he coined the term "DNA profiling" as a means of positive identification. While DNA testing was initially used to determine paternity, its use in solving human identification became clear when DNA examination positively identified the skeletal remains of Josef Mengele (Jeffreys et al. 1992). The capacity to examine creatures' genetic material has been critical to this progress. Today, scientists can extract DNA from microbes, animals, plants, and humans, including their fossils. We can sequence DNA and obtain crucial information about the evolutionary history from the sequences. To date, a substantial amount of human DNA sequence data has been acquired. These have been used to analyze our variety and how individuals from different geographical places or cultural groupings are genetically connected.

The Human Genome Project, which began in 1990 and ended in 2003, accompanied a new era of forensic DNA assessment. The Human Genome Project sequence aided in the discovery of a wide range of short tandem repeat (STR) sequences. PCR-based technique by Dr. Kary Mullis aided in rapid amplification of these polymorphic genetic markers, allowing population studies to examine their genetic variation. By the mid-1990s, DNA profiling effectively integrated high sensitivity provided by PCR-based technologies, higher discrimination power than previously achievable, and increased throughput provided by simultaneous amplification of many STR markers in a single multiplexed reaction. Since then, the field has seen a steady increase in the number of STR markers and marker classes used.

In human migration research, nuclear DNA, mitochondrial DNA (mtDNA), and Y chromosome DNA profiling have been utilized, and each type of DNA contains

Table 1 Analytical timeline of milestones in human evolutionary genomics. A great number of research have used genetic data to gain key insights into human history; those with the most clout in terms of the data or data analyses presented are highlighted (Nielsen et al., 2017; Orlando et al., 2021)

Year	Analytical timeline of milestones in human evolutionary genomics
2010	First ancient human genome First draft Neanderthal genome First draft Denisovan genome 1000 genomes project phase I data
2011	First aboriginal Australian sequenced from 90-year-old tuft of hair
2012	Neolithic Europeans sequenced
2014	12.6-kyr-old Clovis individual sequenced 23-kyr-old Malta individual sequenced 45-kyr-old Ust'ishim individual from Siberia sequenced 38-kyr-old upper Paleolithic European genome
2015	4 L-kyr-old European individual with recent Neanderthal introgression sequenced Large population genome studies of bronze age Europeans and Asians Large genomic studies of ancient and modern native Americans, polo-Eskimo people, and the Invit
2016	Large population genomic studies of Eurosians and Australians
2017	First automated capture of ancient DNA
2018	Tree-based selection scans kinship inference, READ
2019	Ancient metagenomic profiling (HOPS)
2020	High-accuracy phylogenetic assignation of metagenomic data Paleofeces host species identifier, coproHD

different genetic information (Holobinko 2012). Only nuclear DNA represents both the parental genetic materials as one copy is inherited from both the parents, while mtDNA and Y chromosome are nonrecombinant. MtDNA is maternally inherited and Y chromosome paternally, and hence both are traceable to a single female and male ancestral type, respectively (Cox and Hammer 2010). The unique identification abilities of each one are enhanced by the difference in copies per cell between nuclear DNA (2 copies) and mtDNA (more than 1000), which enables positive identification of the individual while mtDNA substantiates family relationships. Ahead of the technology horizon, forensic applications of next-generation sequencing (NGS) technologies have also been significantly improved to provide useful information on an individual's kinship, ancestry, and even phenotype. Analytical milestones of DNA profiling in human migration and evolution are given in Table 1.

General Markers Used in DNA Profiling

Short Tandem Repeat (STR) Markers

Microsatellites, which include simple sequence repeats, enlarged simple tandem repeats, and STRs are a type of genetic markers that are particularly valuable in DNA profiling. These markers are made up of 2–9 base pair DNA motifs that are

tandemly repeated 5–50 times. The human genome contains thousands of STRs loci (Collins et al. 2003), and they are distinguished for their high mutation rates caused by polymerase slippage. As a result, these markers are remarkably polymorphic, with varying numbers of tandem repeat units almost in every population. STRs have been the prime genetic marker that is used to individualize evidentiary material for DNA profiling due to their small size and highly variable nature. This large number of STR loci gives forensic laboratories a lot of options when it comes to incorporating them into commercial kits. Separating various size fragments, usually by capillary electrophoresis, and detecting fluorescently tagged products are required for analyzing amplified STRs. Profiles created from evidence can then be compared to reference samples to determine whether they match, exclude, or are inconclusive. The most useful STR loci (e.g., European Standard Set: FGA, D2S441, TH01, VWA, D1S1656, D10S1248, D12S391, D8S1179, D18S51, D3S1358, D21S11, and D22S1045), the UK core loci (FGA, D2S1338, D18S51, D3S1358, D8S1179, VWA, D16S539, TH01, D19S433, and D21S11), the German core loci (FGA, TH01, SE33, D8S1179, VWA, D18S51, D3S1358, and D21S11), and the Interpol Standard Set (FGA, TH01, VWA, D3S1358, D8S1179, D18S51, and D21S11) (McKiernan and Danielson 2017), and those that are commonly employed in the field, are those that have enough individual variability to allow an inferred linkage of an evidential DNA profile to a specific individual of interest. These STR loci are thought to have a high discrimination power, especially when many STR loci are used. In most cases, the best loci for human identification are those with the most variance in tandem repeats across human population groupings.

One key drawback of utilizing STRs as genetic markers is that they are subject to various degrees of stuttering (Walsh et al. 1996). Stutter is an amplification artifact that occurs as a result of strand slippage during the PCR process. As a result, the DNA polymerase fails to duplicate one or more repeating units or copies the same repeating unit several times. The former produces an amplification product that is one or more repeat units smaller than the genuine genomic allele, whereas the latter produces a product that is typically one repeat larger than the genuine genomic allele (Walsh et al. 1996). STRs, which comprise a four-base-pair core sequence motif (tetranucleotide repeats), are associated with low-stuttering rates and have thus become the most extensively used markers for human identity (Edwards et al. 1991).

Mitochondrial DNA (mtDNA)

Despite accounting for a small portion of an organism's total genome size, mtDNA has become one of the most common markers of genetic variation in animals during the last three decades (Galtier et al. 2009). It is a circular, double-stranded molecule of 16,569 bp size that contains 2 rRNA genes, 22 tRNA genes, and 13 structural genes that code for mitochondrial respiratory chain subunits (DiMauro and Schon 2003). The mutation rate in mitochondrial genomes is several times above in nuclear sequences (Saccone et al. 2000). Many distinct mtDNA variations are detected in one individual as a result of such a high rate of mutation events. The mitochondrial genome has several properties that make it an appealing subject for DNA profiling.

The circular structure of mtDNA and its subcellular confinement within the mitochondrion contribute to its stability. In many tissues, this results in a large mtDNA copy number per cell (i.e., hundreds to thousands of copies of the mtDNA genome per cell) rather than only one copy of the diploid nuclear genome per cell (McKiernan and Danielson 2017). Allen et al. (1998) and Robin and Wong (1988) found that this feature makes it easier to analyze extremely degraded and/or little amounts of initial material. Human mtDNA (both males and females) is entirely acquired through the mother, therefore an individual's mtDNA haplotype is a direct reflection of the up to 100,000 copies of the mitochondrial genome, present in the oocyte at conception, and has few key conserved coding regions (Galtier et al. 2009). As a result, the mtDNA haplotype is the female lineage's counterpart to the Y-STR haplotype.

The examination of mtDNA is a viable option in circumstances when analysis of biological material is problematic (e.g., badly decomposed or skeletonized remains) (Loreille et al. 2010) or nuclear DNA cannot be recovered (Budowle et al. 2003) due to the much higher number of mtDNA copies per cell. The possibility of generating an mtDNA profile in these circumstances makes it a particularly useful technique for historical investigations (e.g., Romanov children (Coble et al. 2009)), and modern forensic cases as mtDNA can be compared to any maternal relative if no immediate relatives are available for body identification or a nuclear DNA comparison (Decorte 2010).

Investigations on DNAs of mitochondria have aided in the development of hypotheses for significant human migration events. For example, analysis of mtDNA in combination with archaeological and climatological data revealed that prehistoric migrations and demographic expansions were linked with paleoclimate (Forster 2004). A number of key evolutionary findings have come from studies of mtDNA. Most crucially, the first study of worldwide human mtDNA sequence variation led to the adoption of the out-of-Africa theory, which proposes that modern humans began in Africa and spread outward from there (Vigilant et al. 1991). The out-of-Africa scenario is supported by phylogenetic trees of human mtDNA sequences, which have a root in Africa. L1, L2, and L3 were the first haplogroups to emerge from Africa. In Europe and Asia, haplogroup L3 gave rise to haplogroups M and N (Fig. 1). Haplogroup N spawned haplogroups H, I, J, N1b, T, U, V, W, and X, all of which can still be found in Europe today. In Asia, haplogroup M gave birth to haplogroups A, B, C, D, F, and G. Native Americans also have haplogroups A, B, C, and D (Fig. 1). mtDNA sequences are often aligned using the Cambridge reference sequence, which is the first fully sequenced mtDNA genome and is derived from European haplogroup H. (Nesheva 2014).

Y Chromosome Marker

The nonrecombining region of the Y chromosome has increasingly become an important tool for studying human evolutionary relationships. With only roughly

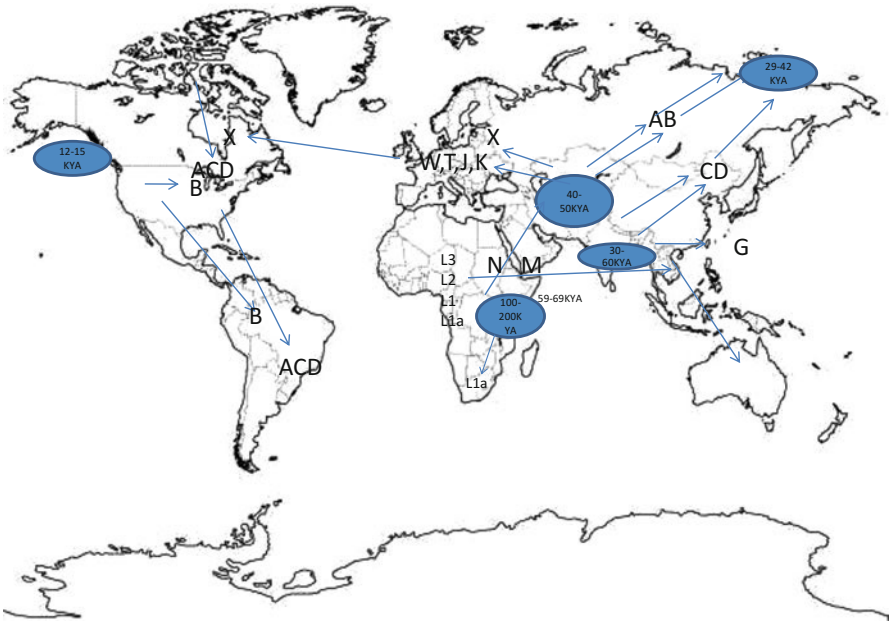


Fig. 1 Geographical distribution of human mtDNA haplotypes (letters in map). The image depicts the origin of anatomically modern humans and possible routes of human migrations "out of Africa" (Witas and Zawicki, 2004)

60 million base pairs of sequence, the Y chromosome is the smallest of the human chromosomes. Y chromosome DNA, like conventional STR markers, is found in the nucleus; nevertheless, there are some major distinctions between markers on the Y chromosome and autosomal STR loci (Kayser 2003). To begin with, the Y chromosome is inherited paternally, and nearly all of the DNA in the Y chromosome is nonrecombinant, with the exception of the Y chromosome's most distal parts (Skaletsky et al. 2003). As a result, a Y haplotype is normally identical among all male relatives of a given paternal family lineage (unless in rare cases of meiotic mutation). The Y chromosome's single nucleotide polymorphisms (SNPs) make these markers show human genetic variation. Based on the studies of global populations (Jobling and Tyler-Smith 2003), we now have a lot of information about the regional origins of Y-SNPs. This makes Y chromosome markers relevant in human identity and paternity cases, but it drastically reduces the power of discriminating obtainable when compared to autosomal STR loci. SNP haplogroups can be used directly to evaluate mixing among distinct populations without resorting to more complex admixture models (Bertorelle and Excoffier 1998) due to the high geographic specificity of Y-SNPs (Jobling and Tyler-Smith 2003). The current Y chromosome genotype-naming system has identified 20 primary haplogroups, designated A through T (Karafet et al. 2008). The genetic variations observed in the Y chromosome of Native American populations were

connected with the dispersal of Siberian groups demonstrating the presence of two separate populations in the New World (growth of Asian populations) (Bianchi et al. 1998).

The Human Migration: Origins in Africa

The oldest evidence for anatomically modern humans comes from Ethiopian fossils dating from 150,000 to 190,000 years (150–190 kyr) ago (McDougall et al. 2005) (Fig. 2). Beyond Africa, anatomically modern humans have been found in the Middle East as early as 100 kyr ago (Grün et al. 2005), and in southern China as early as 80 kyr ago (Liu et al. 2015). Neanderthals, who vanished from the fossil record about 40 kyr ago (Higham et al. 2014), have been discovered as far back as 400 kyr in Eurasia. Initial studies of genomic diversity revealed that Africans have the highest levels of diversity of any living population (Rosenberg et al. 2002). A study of microsatellite DNA variation in more than 3000 Africans discovered 14 ancestral population clusters that corroborated the evidence that the root of the human mtDNA phylogenetic tree is in Africa revealing the extensive population substructure.

These findings were substantially validated by genome-wide SNP genotyping studies (Busby et al. 2016). These and other research findings suggest that African

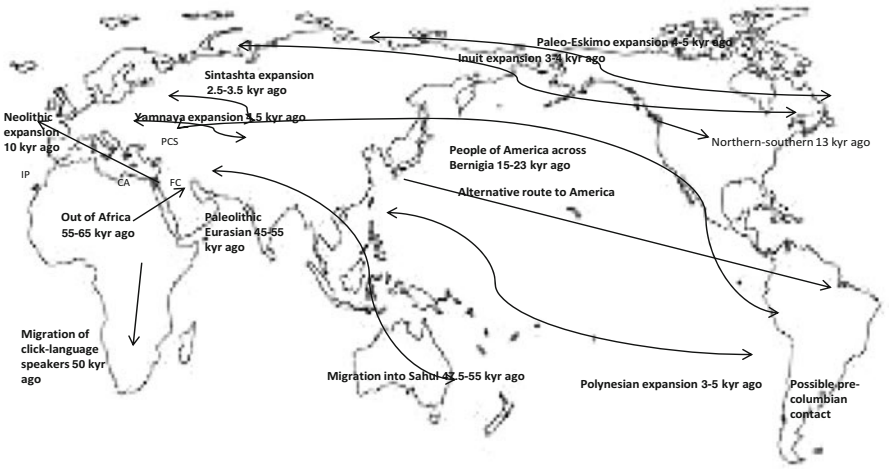


Fig. 2 Major human migrations around the planet as deduced from genomic data. The migration routes utilized to colonize the Americas, for example, are still a source of debate. Genomic data has a limited precision when it comes to determining migratory trajectories since future population movements may obfuscate the spatial patterns that can be discerned from the genomic data. Dashed lines show proposed migratory paths that are still up for debate. CA stands for Central Anatolia; FC is for Fertile Crescent; IP stands for Iberian Peninsula; and PCS stands for Pontic-Caspian Steppe (Nielsen et al., 2017)

populations have remained vast and divided throughout their evolutionary history, with the deepest divisions between human populations occurring in Sub-Saharan Africa. There is also evidence of historical and recent migration events, as well as substantial hybridization, across Sub-Saharan Africa. Other significant migration events include the migration of Pastoralist populations from southern Sudan to eastern and central Africa about 7000 years ago and the migration of Agropastoralists from Ethiopia to Kenya and Tanzania about 5000 years ago. The genetic lineages of click-language-speaking San people of southern Africa capture the deepest divide between human populations, with their separation estimated to have happened roughly 160–110 kyr ago (Veeramah et al. 2012), according to whole-genome sequencing and SNP array data. However, linguistic studies and genetic markers with uniparental inheritance suggest that click-language-speaking hunter-gatherer populations may have originated in eastern Africa and then migrated to southern Africa in the last 50 kyr (Tishkoff et al. 2009), or that they may have originated in eastern Africa and then migrated to southern Africa. Other hunter-gatherer communities that speak click languages, such as the Hadza and Sandawe, live in Tanzania; however, they have low-genomic similarity with the San people of southern Africa (Schlebusch et al. 2012).

The exact origin of anatomically modern humans in Africa is unknown, owing to a lack of fossil and archaeological data in the continent's tropical parts. However, given the possibilities for migration and admixture across the continent, a multi-regional origin of modern humans in Africa (Stringer 2003) is still feasible, in which modern traits arose in a fragmented fashion in different places connected by gene flow.

In Africa, there is evidence of anatomically modern humans admixing with archaic populations (Hammer et al. 2011; Hsieh et al. 2016). Because the environment of the samples from which DNA is extracted, particularly the local climatic conditions, is not conducive to the preservation of genetic material, the characterization of genomes from individuals who lived in Africa more than 10 years ago is difficult. However, statistical analysis of whole-genome sequencing data from geographically diverse hunter-gatherer populations reveals archaic human lineages that underwent introgression (the exchange of genetic material through interbreeding) and diverged from modern human lineages as recently as 35 kyr ago (Hsieh et al. 2016). As a result, the degree of archaic hybridization in Africa is a contentious issue, with numerous continuing initiatives aimed at resolving it.

Out of Africa and the Conjunction with Neanderthals

The exodus of anatomically modern humans from Africa, a significant event in human evolutionary history, left a profound imprint on non-African genetic variation, including lower levels of diversity and higher levels of linkage disequilibrium. The quantity, geographic origin, migratory routes, and timing of significant dispersals, however, remain unknown. For example, evidence suggests that modern humans originated in eastern, central, and southern Africa (Fig. 2). Three studies

(Pagani et al. 2016) use new, high-quality whole-genome sequencing data from individuals from more than 270 places around the world to help answer some of these problems.

They also suggest that a single out-of-Africa dispersal occurred, with all non-African people deriving from the same ancestral population that migrated out of Africa, maybe with modest genetic contributions from an earlier modern human migration wave into Oceania. Furthermore, as modern people left Africa, they may have split into two waves of dispersal. One wave, as proposed by Rasmussen et al. (2011), led to the establishment of Australasia and New Guinea, whereas the other contributed to the lineage of today's mainland Eurasians. However, the precise migration routes in the early diversification of people outside of Africa are still a source of debate and investigation.

It is now known that all non-African people's ancestors came into contact with Neanderthals and were admixed with them (Prüfer et al. 2014). All non-Africans investigated so far have roughly 2% Neanderthal ancestry (Vernot and Akey 2015), implying that admixture took place shortly after the dispersal of anatomically modern humans from Africa, which is compatible with a single-dispersal out-of-Africa hypothesis (Fig. 2). The date of hybridization has been estimated to be around 50–65 kyr ago (Sankararaman et al. 2014), based on linkage disequilibrium characteristics, and knowledge of the time of admixture with Neanderthals helps to constrain estimates of when the ultimately successful out-of-Africa migration occurred.

Estimates of the percentage of Neanderthal ancestry found in modern human genomes (Vernot et al. 2014) suggest a more complicated history of interaction amid Neanderthals and modern humans. East Asians have about 20% more Neanderthal sequences than Europeans, which could be due to natural selection. After the population split from Europeans, there was more admixture in the progenitors of today's East Asians (Vernot et al. 2014; Vernot and Akey 2015; Kim and Lohmueller 2015), or the dilution of Neanderthal ancestry in Europeans due to hybridization with populations from other continents. SNP genotyping of an early modern person from Romania who lived around 40,000 years ago offered more evidence that introgression happened at multiple times and locations throughout Eurasia (Fu et al. 2015), despite the fact that the individual did not contribute detectable ancestry to modern populations. Recent studies have shown a more complicated mixing history than previously anticipated, and that our understanding of admixture models is currently fluid and that additional demographic models are compatible with the observed.

The Peopling of Europe

Three or more genetic components are likely to make up European populations, some of which arrived in Europe at different times (Haak et al. 2015). As early as 43,000 years ago, the earliest anatomically modern humans lived in Europe, because there is evidence of turnover in the genetic composition of Europeans before the Last

Glacial Maximum, possibly in relation to climate oscillations (Fu et al. 2016). After the Last Glacial Maximum, some 11 kyr ago, a new way of life based on animal husbandry, agriculture, and sedentarism that is known as the Neolithic lifestyle began to emerge in numerous Fertile Crescent subregions (Asouti et al. 2013). Ancient DNA analysis revealed that this community of farmers spread from Central Anatolia to Europe, but that other Fertile Crescent locations contributed just a small amount of genetic material to the early European farmers (Günther and Jakobsson 2016). They arrived in the Iberian Peninsula around 7000 years ago, and in Britain and Scandinavia approximately 6000 years ago.

The widespread migration of groups of farmers and the absorption of local hunter-gatherers (Günther et al. 2015), as indicated by genomic data from Neolithic human remains, drove this process, revealing that the Neolithic way of life spread over Europe through people mobility rather than as a concept or a culture. Although archaeological data suggests that the health of Neolithic farmers was sometimes poor, as there were ample signs of malnutrition and caries, the Neolithic lifestyle helped to increase the size of populations, as seen in estimates of effective population sizes generated from genomic data (Skoglund et al. 2014).

During the late Neolithic and early Bronze Age periods, another wave of migration into Europe occurred, introducing the third European genetic component. Herders from the Pontic-Caspian steppe belonging to the Yamnaya civilization migrated to central Europe around 4500 years ago (Haak et al. 2015). The herders were descended from a number of hunter-gatherer tribes from (today) Russia and the Caucasus (Jones et al. 2015). Although some linguistics experts think that Indo-European languages were already spoken by Neolithic farmers (Bouckaert et al. 2012), this movement was likely linked to conquests and technological advancements such as horseback riding and may have spread Indo-European languages to Europe. Clearly, the late Neolithic and Bronze Ages were active periods that resulted in the spread of steppe herders' genetic material across western and northern Europe (Haak et al. 2015; Allentoft et al. 2015).

The contributions of hunter-gatherers to the decolonization of Europe after the Last Glacial Maximum, the migration of Neolithic farmers from Anatolia to Europe in the late-Neolithic period, and the Bronze Age migration to Europe from the east are reflected in the three main genetic components of modern-day European populations. These elements can account for a large portion of the genetic diversity found in today's Europe. For example, in southern European communities such as the Sardinian people (Lazaridis et al. 2014), the Neolithic genetic component appears to be most prominent. Geographical variety in modern-day Europeans is closely connected with genetic variation (Lao et al. 2008), and there is a gradient of decreasing variety as one moves further north.

Despite the fact that the major components of genetic variants were introduced into Europe in different waves, successive processes of gene flow constrained by geography have sculpted the current genetic landscape. In many times of Prehistoric Europe, culture and lifestyle were thus more important factors of genomic divergence and similarity than geography.

The Colonization of Asia and Oceania

The majority of evidence suggests that Asia was colonized in at least two waves. Although additional evidence implies that there was just one dispersal event (Mallick et al. 2016), one wave contained the ancestors of Australians and Papuans, and the other contained other ancestors of East Asians, with admixing between the two (Rasmussen et al. 2011). The specifics of how Asia was colonized are, however, mostly obscure. The genomes of two early modern humans from Asia have been sequenced. The first genome came from a person from the Malta-Buret culture of southern central Siberia who lived about 24,000 years ago (Raghavan et al. 2014), and it shows a strong genetic affinity for both western Eurasians and Native Americans, but a weaker affinity for East Asians and Siberians, implying that there was a very different geographic distribution of genetic signatures during the Upper Paleolithic period in campaigns. When differences in Denisovan admixture are taken into account, the second genome, which comes from a person who lived in the Ust'-Ishim region of western Siberia (Fu et al. 2014) about 45 years ago, displays virtually equal genetic kinship with western Eurasians, East Asians, and Aboriginal Australians.

This, combined with evidence from the Kostenki 14 individual's (36–38-kyr-old) genome from European Russia (Seguin-Orlando et al. 2014), shows a close kinship to contemporary western Eurasians, but not East Asians, and points to a 36–45-kyr-old separation between East Asians and western Eurasians. Two later population expansions into central Asia from Europe and western Asia resulted in mixing with and locally displacing Malta-like hunter-gatherers, according to a study that comprised low-coverage sequencing of the genomes of 101 ancient humans from across Bronze Age Eurasia (Allentoft et al. 2015). The first occurrence was a migration of Yamnaya herders into Asia about 5000 years ago, which coincided with the arrival of the first humans in Asia.

Similar assertions have been made based on linguistic evidence and lithic technology, as well as the arrival of tamed animals like the dingo (Pugach et al. 2013). However, the only comprehensive population genomic study on Aboriginal Australians and Papuans to date (Malaspinas et al. 2016) finds evidence for only one founding event in Sahul, followed by a genetic divergence between the Papuan and Aboriginal Australian ancestral populations and further genetic diversification in the Aboriginal Australian population, which could have coincided with environmental changes such as desertification. As a result, Aboriginal Australians seemed to have been isolated until recently. Polynesians, who are spread across a triangle of islands in the South Pacific bounded to the east by Rapa Nui (also known as Easter Island), represent an expansion into Oceania of individuals with mixed Melanesian and East Asian ancestry, according to a study (Wollstein et al. 2010) of genome-wide SNP data from modern people in Oceania. After the initial Polynesian spread, Melanesian ancestry was added to the existing East Asian ancestry (Skoglund et al. 2016). It is still debatable if Polynesians reached the Americas and mated with Native Americans during their eastward march, which stopped approximately a thousand years ago.

This concept has been supported by a genetic examination of ancient chicken remains from South America (Storey et al. 2007), but it has also been questioned (Gongora et al. 2008). Genome sequencing of human remains discovered in Brazil circa 1650, which predates the reported transportation of Polynesian slaves to South America (Malaspinas et al. 2014), reveals that the individuals are closely linked to modern Polynesians. These findings could corroborate the theory of early interaction between Polynesians and Native Americans, but they might also be the consequence of people being transported by Europeans. The findings of a genome-wide investigation on modern-day inhabitants of Easter Island (Moreno-Mayar et al. 2014), which offered statistical support for Native American claims, are much more convincing. The only way to settle the argument is to find evidence of Polynesian and Native American hybridization in human remains that predate colonization of the Americas.

Populating Americas

The earliest well-characterized archaeological assemblage in the Americas, the Clovis complex (approximately 12.6–13 kyr ago), dates to about 15–14 kyr ago (Jenkins et al. 2012), and widespread colonization of the Americas appeared with the advent of the Clovis complex (approximately 12.6–13 kyr ago) (Fig. 2). However, until about 13,000 years ago, much of North America was covered in ice, making it difficult for humans to migrate from Beringia (today northern Siberia and northwestern North America) to the southern sections of the Americas. After the ice melted, a 1500 km ice-free inner passage emerged. According to metagenomic investigations of lake cores from Canada, this corridor first became biologically viable around 12.6 kyr ago, which makes it an unlikely early route for the southward migration of pre-Clovis and Clovis groups of people, although a study of bison is in disagreement (Heintzman et al. 2016).

It is unclear how and when the first Americans crossed the Pleistocene ice sheets into southern North America, or whether the pre-Clovis and Clovis populations were part of the same migration. However, the most feasible scenario appears to be a movement toward the south along North America's west coast that happened more than 14,000 years ago and was possibly followed by southerly or northerly back-migrations into the interior.

It has been proposed that early Americans were not direct ancestors of contemporary Native Americans but rather were related to Australia-Melanesians, Polynesians, the Ainu people of Japan, or Europeans who were later replaced or assimilated by ancestors of Native Americans from Siberia (Owsley and Jantz 2014), based on cranial morphology and lithic analysis. Several genetic investigations, on the other hand, have generally dismissed these ideas. The earliest and only Clovis-associated human genome from the Americas (discovered in Montana, USA) belonged to an individual who lived around 12.6 million years ago. According to studies, the Clovis group from which the genome was derived was directly ancestral to many modern Native Americans. Similarly, the genome sequence of the 9.5-kyr-old Kennewick

Man skeleton discovered in the state of Washington in the United States (Rasmussen et al. 2015), which had been thought to be closely related to the Ainu and Polynesians based on cranial morphology, determined that he was most closely related to contemporary Native Americans. Furthermore, communities previously thought to represent relics of an early migration into the Americas and closely related to Australo-Melanesians have been discovered to be genetically connected to modern Native Americans (Raghavan et al. 2015).

Based on whole-genome sequencing, estimates of the timing of divergence between Siberians and Native Americans suggest the creation of the Native American gene pool as early as 23 kyr ago (Raghavan et al. 2015), supporting the early arrival of Native American ancestors in the Americas. When the recognized dates for the earliest archaeological sites in the Americas are taken into account, Native American ancestors could have lived in isolation in Siberia or Beringia until roughly 8 kyr ago, after splitting from their Siberian forebears, before traveling eastward into the Americas.

Despite the fact that modern Siberians are Native Americans' closest relatives outside of the Americas, genomic sequencing of a 24-kyr-old Malta skeleton reveals that Native Americans are descended from a combination of Malta-related populations as well as one or more unknown East-Asian lineages. Because the Clovis-associated genome and modern Native Americans have similar levels of the Malta genetic signature (14–38%), the mixing event occurred over 12.6 million years ago. However, it is unknown whether it occurred within or outside of the Americas. Genomic data have been utilized to locate a basal divide in Native Americans that can be dated to around 14,000–13,000 years ago. The southern branch is made up of people who speak Amerindian languages, whereas the northern branch is made up of people who speak Athabascan languages as well as other languages such as Cree and Algonquin. Divergence estimates based on whole-genome sequencing data reveal that both groups diverged from Siberians at the same time, meaning that both Amerindian and Athabascan populations had a single founding event followed by gene flow from Asia (Raghavan et al. 2015).

The exact location of the split between the two Native American branches, whether in Siberia or to the north or south of the American ice sheets, is still a matter of contention, which will require the investigation of more ancient genomes. Similarly, it is unclear whether the Australo-Melanesian signal detected in some current Brazilian Native Americans is due to later gene flow (Raghavan et al. 2015) or a previously undiscovered early founding population. This genetic signature has yet to be discovered in the genomes of ancient humans from the Americas. The Inuit of the American Arctic is said to have come from a different migration than other Native Americans (Reich et al. 2012; Gilbert et al. 2008). However, it has long been debated whether the earliest people to occupy the Arctic, the now-extinct Paleo-Eskimo civilization, who first emerged in the Americas some 5000 years ago, were the progenitors of today's Inuit or a separate founder group from Siberia. DNA analysis of a 4-kyr-old tuft of hair from Greenland revealed that the group the

individual belonged to traveled from Siberia to the North American Arctic independently of Native American and Inuit migrations. The tribe then thrived in the Arctic for almost 4000 years, renewing its subsistence techniques and technologies before being displaced by the Inuit around 700 years ago.

The Denisovans

When the first modern humans began to appear on the continent, at least one other form of archaic human – the enigmatic Denisovans – thrived in Eurasia alongside the Neanderthals (Fig. 2). Denisovans are only known from the genome sequences of a finger bone and three teeth recovered in the Denisova Cave in Siberia (2015; Stringer and Barnes 2015); therefore, nothing is known about their morphology and distribution. They are most closely related to Neanderthals, with a genetic differential similar to the deepest splits between modern humans (Prüfer et al. 2014) but a divergence time estimated to be between 200 and 400 kyr. Denisovans have a number of unique characteristics, for example, they may contain genetic material (obtained through hybridization) from people linked to earlier human species, potentially *Homo erectus*. Denisovans are thought to be the eastern or southern end of a spectrum of archaic humans who lived in Eurasia (and possibly beyond), while Neanderthals are thought to be the western end.

Denisovans interbred with anatomically modern humans in the same way that Neanderthals did. Some groups of people, such as Melanesians in Oceania, can trace 3–6% of their genome back to a Denisovan-like ancestor. Denisovan genetic DNA is found in 0.1–0.3% of continental southeast Asians (Skoglund et al. 2011). On introgression, both Neanderthals and Denisovans' genomes were subjected to genetic selection. Because there is a scarcity of introgressed DNA in functional sections of the genome, most selection in humans appears to be directed against it. In addition, extensive genomic areas devoid of both Neanderthal and Denisovan sequences have been discovered (Vernot et al. 2016), indicating that detrimental sequences were rapidly eliminated. Some introgressed DNA, on the other hand, may have aided human adaption to the local environment, such as Tibetan people's adaptation to high altitudes (Huerta-Sánchez et al. 2014).

Two punctuated and very precise instances for hybridization between anatomically modern humans and archaic humans were hypothesized by studies of the first trustworthy genomic data from archaic humans (Reich et al. 2010). Since then, such hybridization is far more widespread, with several episodes occurring in both directions across distinct groups of contemporary and archaic humans (Vernot et al. 2015; Huerta-Sánchez et al. 2014). As the progenitors of current Australasians traveled across the continent, it is unclear whether the Denisovan introgression into Melanesians and Australians took place in Australasia or Asia. If this had happened in Asia, today's Asians would be primarily descended from other populations who arrived during the subsequent centuries. If that had happened in Asia, the majority of

today's Asians would be descended from other groups who arrived during succeeding waves of migration. Similarly, it is unclear whether East Asians' Denisovan admixture is the consequence of the same mixing event or events that affected Australasians.

The investigation of ancient human genomes has advanced significantly in recent years, with studies of single genomes to population-genomic studies including several ancient individuals. Despite the fact that this has supplied a plethora of information about human evolutionary genomics, the science is still in its early stages. Continued attempts to sequence and analyze the genomes of current and ancient humans, with an emphasis on underrepresented areas of the globe, will aid us in forming a more complete picture of the events that have shaped contemporary humans' cultural and genetic variety.

Future Directions

DNA genetic markers will play an important role in human migration and evolution in the future with rapid advancements in technology and data analysis. Whole-genome sequencing will become more affordable and efficient. The most difficult task for scientists will be to analyze massive data sets generated by massive sequencing technologies. As a result, more emphasis can be placed on clarifying the cultural and environmental influences on gene expression. The future application of genetic markers (DNA fingerprints) is wide open, and study over the next decade will lead to a better understanding of our species' origins and evolution. It is unclear how far back in time ancient DNA studies will go, but these new methodologies will provide anthropologists with a refined story of human history, unraveling the complexities of human migration, admixture, and the successful and unsuccessful ways in which Hominin genomes were selected by their environment. Research into still-unanswered concerns about human evolution and migration would be one of the work's future areas. A better understanding of evolution could lead to novel and creative therapeutic options for many diseases that are currently incurable and untreatable.

References

Allen M, Engström AS, Meyers S, Handt O, Saldeen T, von Haeseler A et al (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J Forensic Sci 43(3):453–464

Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J et al (2015) Population genomics of Bronze Age Eurasia. Nature 522(7555):167–172

Asouti E, Fuller DQ, Barker G, Finlayson B, Matthews R, FazeliNashli H et al (2013) A contextual approach to the emergence of agriculture in Southwest Asia: reconstructing early Neolithic plant-food production. Curr Anthropol 54(3):299–345

Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. Mol Biol Evol 15(10):1298–1311

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB et al (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63(6):1862–1871

Bouckaert R, Lemey P, Dunn M, Greenhill SJ, Alekseyenko AV, Drummond AJ et al (2012) Mapping the origins and expansion of the Indo-European language family. Science 337(6097):957–960

Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 4(1):119–141

Busby GB, Band G, Le QS, Jallow M, Bougama E, Mangano VD et al (2016) Admixture into and within sub-Saharan Africa. elife 5:e15266

Coble MD, Loreille OM, Mark J, Wadhams SME, Maynard K, Meyer CE et al (2009) The identification of the two missing Romanov children using DNA analysis. PLoS One 4(3):e4838

Collins JR, Stephens RM, Gold B, Long B, Dean M, Burt SK (2003) An exhaustive DNA microsatellite map of the human genome using high performance computing. Genomics 82(1):10–19

Cox MP, Hammer MF (2010) A question of scale: human migrations writ large and small. BMC Biol 8(1):1–5

Decorte R (2010) Genetic identification in the 21st century – current status and future developments. Forensic Sci Int 201(1–3):160–164

DiMauro S, Schon EA (2003) Mitochondrial respiratory-chain diseases. N Engl J Med 348(26): 2656–2668

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49(4):746

Forster P (2004) Ice ages and the mitochondrial DNA chronology of human dispersals: a review. Philos Trans R Soc Lond Ser B Biol Sci 359(1442):255–264

Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P et al (2015) An early modern human from Romania with a recent Neanderthal ancestor. Nature 524(7564):216–219

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D et al (2016) The genetic history of ice age Europe. Nature 534(7606):200–205

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Mol Ecol 18(22):4541–4550

Gilbert MTP, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M et al (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320 (5884):1787–1789

Gongora J, Rawlence NJ, Mobegi VA, Jianlin H, Alcalde JA, Matus JT et al (2008) Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. Proc Natl Acad Sci 105(30):10308–10313

Grün R, Stringer C, McDermott F, Nathan R, Porat N, Robertson S et al (2005) U-series and ESR analyses of bones and teeth relating to the human burials from Skhul. J Hum Evol 49(3):316–334

Günther T, Jakobsson M (2016) Genes mirror migrations and cultures in prehistoric Europe – a population genomic perspective. Curr Opin Genet Dev 41:115–123

Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir ÓO et al (2015) Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proc Natl Acad Sci 112(38):11917–11922

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B et al (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522(7555):207–211

Hammer MF, Woerner AE, Mendez FL, Watkins JC, Wall JD (2011) Genetic evidence for archaic admixture in Africa. Proc Natl Acad Sci 108(37):15123–15128

Heintzman PD, Froese D, Ives JW, Soares AE, Zazula GD, Letts B et al (2016) Bison phylogeography constrains dispersal and viability of the Ice Free Corridor in western Canada. Proceedings of the National Academy of Sciences 113(29):8057–8063

Higham T, Douka K, Wood R, Ramsey CB, Brock F, Basell L et al (2014) The timing and spatiotemporal patterning of Neanderthal disappearance. Nature 512(7514):306–309

Holobinko A (2012) Theoretical and methodological approaches to understanding human migration patterns and their utility in forensic human identification cases. Societies 2(2):42–62

Hsieh P, Veeramah KR, Lachance J, Tishkoff SA, Wall JD, Hammer MF, Gutenkunst RN (2016) Whole-genome sequence analyses of Western Central African Pygmy hunter-gatherers reveal a

complex demographic history and identify candidate genes under positive natural selection. Genome Res 26(3):279–290
Huerta-Sánchez E, Jin X, Bianba Z, Peter BM, Vinckenbosch N, Liang Y et al (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512(7513):194–197
Jeffreys AJ, Allen MJ, Hagelberg E, Sonnberg A (1992) Identification of the skeletal remains of Josef Mengele by DNA analysis. Forensic Sci Int 56(1):65–76
Jeffreys AJ, Brookfield JF, Semeonoff R (1985) Positive identification of an immigration test-case using human DNA fingerprints. Nature 317(6040):818–819
Jenkins DL, Davis LG, Stafford TW, Campos PF, Hockett B, Jones GT et al (2012) Clovis age Western Stemmed projectile points and human coprolites at the Paisley Caves. Science 337(6091):223–228
Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4(8):598–612
Johnston E, Crisp A, McKie R, Brawn R (2019) Forensic science & human migration: the role of forensic genetics. Forensic Sci Int Genet Suppl Ser 7(1):450–451
Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R et al (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun 6(1):1–8
Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18(5):830–838
Kayser M (2003) The human Y-chromosome-introduction into genetics and applications. Forensic Sci Rev 15(2):77–89
Kim BY, Lohmueller KE (2015) Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. Am J Hum Genet 96(3):454–461
Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A et al (2008) Correlation between genetic and geographic structure in Europe. Curr Biol 18(16):1241–1248
Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K et al (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513(7518):409–413
Liu W, Martinón-Torres M, Cai YJ, Xing S, Tong HW, Pei SW et al (2015) The earliest unequivocally modern humans in southern China. Nature 526(7575):696–699
Loreille OM, Parr RL, McGregor KA, Fitzpatrick CM, Lyon C, Yang DY et al (2010) Integrated DNA and fingerprint analyses in the identification of 60-year-old mummified human remains discovered in an Alaskan glacier. J Forensic Sci 55(3):813–818
Malaspinas AS, Lao O, Schroeder H, Rasmussen M, Raghavan M, Moltke I et al (2014) Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil. Curr Biol 24(21):R1035–R1037
Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I et al (2016) A genomic history of Aboriginal Australia. Nature 538(7624):207–214
Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F et al (2016) The Simons genome diversity project: 300 genomes from 142 diverse populations. Nature 538(7624):201–206
McDougall I, Brown FH, Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433(7027):733–736
McKiernan HE, Danielson PB (2017) Molecular diagnostic applications in forensic science. In: Molecular diagnostics. Academic Press, New York, pp 371–394
Moreno-Mayar JV, Rasmussen S, Seguin-Orlando A, Rasmussen M, Liang M, Flåm ST et al (2014) Genome-wide ancestry patterns in Rapanui suggest pre-European admixture with Native Americans. Curr Biol 24(21):2518–2525
Nesheva DV (2014) Aspects of ancient mitochondrial DNA analysis in different populations for understanding human evolution. Balk J Med Genet 17(1):5
Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E (2017) Tracing the peopling of the world through genomics. Nature 541(7637):302–310

Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC et al (2021) Ancient DNA analysis. Nature Reviews Methods Primers 1(1):1–26
Owsley DW, Jantz RL (eds) (2014) Kennewick Man: the scientific investigation of an ancient American skeleton. Texas A&M University Press, College Station
Pagani L, Lawson DJ, Metspalu M (2016) Genomic analyses inform on migration events during the peopling of Eurasia. Nature 538:238–242
Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S et al (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505(7481):43–49
Pugach I, Delfin F, Gunnarsdóttir E, Kayser M, Stoneking M (2013) Genome-wide data substantiate Holocene gene flow from India to Australia. Proc Natl Acad Sci 110(5):1803–1808
Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I et al (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505(7481):87–91
Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M et al (2015) Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349(6250):841
Rasmussen M, Guo X, Wang Y, Lohmueller KE, Rasmussen S, Albrechtsen A et al (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 334(6052): 94–98
Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, Poznik GD et al (2015) The ancestry and affiliations of Kennewick Man. Nature 523(7561):455–458
Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY et al (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468(7327):1053–1060
Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N et al (2012) Reconstructing native American population history. Nature 488(7411):370–374
Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136(3):507–513
Roewer L (2013) DNA fingerprinting in forensics: past, present, future. Investig Genet 4(1):1–10
Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298(5602):2381–2385
Saccone C, Gissi C, Lanave C, Larizza A, Pesole G, Reyes A (2000) Evolution of the mitochondrial genetic system: an overview. Gene 261(1):153–159
Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S et al (2014) The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507(7492):354–357
Schlebusch CM, Jakobsson M (2018) Tales of human migration, admixture, and selection in Africa. Annu Rev Genomics Hum Genet 19:405–428
Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F et al (2012) Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science 338(6105):374–379
Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I et al (2014) Genomic structure in Europeans dating back at least 36,200 years. Science 346(6213):1113–1118
Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423(6942):825–837
Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T et al (2014) Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science 344(6185):747–750
Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S et al (2016) Genomic insights into the peopling of the Southwest Pacific. Nature 538(7626):510–513
Storey AA, Ramirez JM, Quiroz D, Burley DV, Addison DJ, Walter R et al (2007) Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. Proc Natl Acad Sci 104(25):10335–10339
Stringer C (2003) Out of Ethiopia. Nature 423(6941):693–695

Stringer CB, Barnes I (2015) Deciphering the Denisovans. Proc Natl Acad Sci 112(51): 15542–15543

Thorne AG, Wolpoff MH (1992) The multiregional evolution of humans. Sci Am 266(4):76–83

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A et al (2009) The genetic structure and history of Africans and African Americans. Science 324(5930):1035–1044

Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, Destro-Bisol G et al (2012) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol Biol Evol 29(2):617–630

Vernot B, Akey JM (2015) Complex history of admixture between modern humans and Neanderthals. Am J Hum Genet 96(3):448–453

Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM et al (2016) Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. Science 352(6282):235–239

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253(5027):1503–1507

Walsh PS, Fildes NJ, Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Res 24(14):2807–2812

Wilson AC, Cann RL (1992) The recent African genesis of humans. Sci Am 266(4):68–75

Witas HW, Zawicki P (2004) Mitochondrial DNA and human evolution: A review. Anthropological Review 67:97–110

Wollstein A, Lao O, Becker C, Brauer S, Trent RJ, Nürnberg P et al (2010) Demographic history of Oceania inferred from genome-wide data. Curr Biol 20(22):1983–1992

Index

H. R. Dash et al. (eds.), *Handbook of DNA Profiling*,
https://doi.org/10.1007/978-981-16-4318-7

D