Hirak Ranjan Dash · Pankaj Shrivastava
Braja Kishore Mohapatra · Surajit Das
Editors

DNA Fingerprinting: Advancements and Future Endeavors

Springer

DNA Fingerprinting: Advancements and Future Endeavors

Hirak Ranjan Dash • Pankaj Shrivastava
Braja Kishore Mohapatra • Surajit Das
Editors

DNA Fingerprinting: Advancements and Future Endeavors

Springer

Editors
Hirak Ranjan Dash
DNA Fingerprinting Unit
State Forensic Science Laboratory
Sagar, Madhya Pradesh, India

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

Braja Kishore Mohapatra
Department of Biology and DNA
Fingerprinting Unit
Central Forensic Science Laboratory
New Delhi, Delhi, India

Surajit Das
Department of Life Science, National
Institute of Technology
Rourkela, Odisha, India

ISBN 978-981-13-1582-4 ISBN 978-981-13-1583-1 (eBook)
https://doi.org/10.1007/978-981-13-1583-1

Library of Congress Control Number: 2018959136

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

Preface

DNA fingerprinting technique is the most trusted gift of science to the mankind, as it is helpful in the field of the criminal justice system, deciphering the genetics of living organisms, diagnosis of the genetic disorder, neonatal diagnosis, cracking of ancestral belongingness, wildlife forensics, and many more. The technique has undergone much advancement with time since its inception. With the advent of time, the technology has spread its tentacles in various other fields which have resulted in the introduction of new academic courses along the globe to fulfill the demand of expertise in diverse aspects of DNA fingerprinting. This demand also resulted in the demand of available literature. However, most of the books available in this field concentrate either on forensic applications or disease diagnosis or legal issues. However, a consensus reference book in this field describing the basics, various applications and use of the technology in real case studies are lacking. Hence the present volume is planned to touch the fields of genetics, tools, and techniques, description of real-time case studies, wildlife forensics, molecular diagnosis of human diseases, legal aspects, and microbial forensics.

The current volume includes four parts: *Part 1*: Basics of DNA Fingerprinting: Tools and Techniques, *Part 2*: Applications of DNA Fingerprinting, *Part 3*: DNA Fingerprinting: Case Studies, and *Part 4*: Future of DNA Fingerprinting. Part 1 consists of four chapters describing the discovery and advancements of DNA technology as well as the involvement of various tools and technology for DNA fingerprinting application. Part 2 covers various applications of DNA fingerprinting including wildlife forensics, identification of mutilated remains, molecular diagnosis of human diseases, human trafficking, and from judicial point of view. Application of various types of currently practiced DNA fingerprinting techniques using Autosomal and Y chromosome STR typing as well as mitochondrial DNA sequencing for criminal justice system has been described in Part 3. Finally, Part 4 harbors some futuristic approach of current day DNA fingerprinting such as whole genome sequencing and microbial forensics.

The current volume has been written in simple English that may require basic biological science background to understand. It will be helpful for the students' from the fields of Zoology, Wildlife, Medicine, Anthropology, Microbiology,

Forensic Science, Genetics, and Law at graduate, postgraduate, and research level. For scientific fraternity, it will be a handy reference to quickly summarize the technological advancements in the field of DNA fingerprinting, to understand the problems faced by this field of science and possible updated solutions to these problems. As nowadays, DNA fingerprinting is used in solving most of the criminal cases, this book will be helpful among the law practicing friends as well. Investigating agencies can also gather a sound knowledge from this book as real case studies have been included here.

We have tried our best to share the available knowledge around the globe in the field of DNA fingerprinting with the aim to provide an important and rationalized resource material in the form of this edited volume. Throughout the editing process of the book, we have faced many problems and hurdles and all have been overcome due to God's grace, self-belief, and nice people surrounding to us. We are highly thankful to each and every one for their support and encouragement during this process. Wishing a good luck to all the readers.

Sagar, Madhya Pradesh, India — Hirak Ranjan Dash
Sagar, Madhya Pradesh, India — Pankaj Shrivastava
New Delhi, Delhi, India — Braja Kishore Mohapatra
Rourkela, Odisha, India — Surajit Das

Contents

About the Editors

Dr. Hirak Ranjan Dash completed his Ph.D. from the Department of Life Science, National Institute of Technology, Rourkela, India, and is currently working as a Scientific Officer (DNA) at the Forensic Science Laboratory, Madhya Pradesh, 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, environmental microbiology, DNA fingerprinting, microbial phylogeny, genetic manipulation of bacterial systems, and microbial diversity. He has developed a number of microbial techniques for the assessment of mercury pollution in marine environments and has successfully constructed a transgenic marine bacterium for enhanced utilization in mercury removal by simultaneous mercury volatilization and sequestration. He has written 3 books and published 28 research papers, 11 book chapters, and 12 conference proceedings.

Dr. Pankaj Shrivastava received his Ph.D. in Microbiology from Rani Durgavati University, Jabalpur. He is presently serving as a Scientific Officer at the DNA Fingerprinting Unit, Forensic Science Laboratory, Madhya Pradesh, India. He has more than 10 years of experience in examining a variety of criminal cases using DNA fingerprinting. The central theme of his research is the DNA analysis of caste and tribal populations of different parts of India, along with the development of new methodologies for improved forensic DNA typing. Till date, he has published 11 books and 61 scientific articles in reputed international journals. He is a visiting faculty of National Police Academy, Hyderabad; National Institute of Criminology and Forensic Science, Government of India, Delhi; and the Central Police Academy, Bhopal, along with many central and state universities of India. He is a recipient of the Pt. Govind Vallabh Pant Samman Award from the Ministry of Home, Government of India; the Anusrijan Samman Award from AISECT University, Bhopal; the Dr. Lalji Singh Memorial Award; and the FICCI Smart Policing Award for the development of a direct protocol in forensic DNA typing.

Dr. Braja K. Mohapatra completed his Ph.D. from Utkal University, Bhubaneswar. He is presently serving as a Senior Scientific Officer and Head of the Department of Biology and DNA Profiling Unit, Central Forensic Science Laboratory (CBI), New Delhi, India. He has more than 10 years of experience in examining various criminal cases using DNA fingerprinting. His research interests include the interpretation of DNA profiles in mixed samples, touch DNA, and population genetics. He has 13 peer-reviewed publications in reputed national and international journals to his credit. He is a recipient of the meritorious service award in forensic science. He has solved various high-profile cases through DNA fingerprinting, both in India and The Republic of Seychelles.

Dr. Surajit Das is an Associate Professor at the Department of Life Science, National Institute of Technology, Rourkela, India. He received his Ph.D. in Marine Biology from the Centre of Advanced Study in Marine Biology, Annamalai University, Tamil Nadu, India, for his research work on marine microbiology. He has been awarded an Endeavour Research Fellowship by the Australian Government to carry out postdoctoral research at the University of Tasmania. As group leader of the Laboratory of Environmental Microbiology and Ecology (LEnME), he is currently conducting research on the biofilm-based bioremediation of PAHs and heavy metals using marine bacteria; nanoparticle-based drug delivery and nanobioremediation; and metagenomic approaches for exploring the diversity of immunoglobulins in the Indian Major Carps, supported by research grants from the Ministry of Science and Technology; Indian Council of Agricultural Research; Ministry of Environment, Forest and Climate Change; and the Government of India. He is an Academic Editor for *PLOS One* and an Associate Editor (Ecological and Evolutionary Microbiology) for *BMC Microbiology*.

Part I
Basics of DNA Fingerprinting: Tools and Techniques

Chapter 1
DNA Fingerprinting: Discovery, Advancements, and Milestones

Jahangir Imam, Romana Reyaz, Ajay Kumar Rana, and Vrijesh Kumar Yadav

Abstract The discovery of DNA fingerprinting is one of the most fascinating scientific discoveries till date. It is not only limited to the laboratory research but also showed a huge potential in forensic science and criminal justice system. It was one of the milestones in resolving crimes by exploring the polymorphism of human DNA in noncoding regions. Since its inception, DNA fingerprinting has taken a great leap in terms of advancements in technology, accuracy, and reliability of the results as well as rapidity of the process for its more efficient application in justice delivery systems. This has become the most valuable armory of the judiciary system to aid in the conviction of guilty as well as exoneration of the innocent. Advancement of DNA fingerprinting technique from RFLP to STR and now NGS has sped up the process of DNA profiling with better discriminating power among individuals with greater efficacy. In this prospect, the current chapter elaborately recapitulates the process of advancement in DNA fingerprinting describing the use of different STR kits, i.e., autosomal STRs, Y-STRs, X-STRs, miniSTRs, etc., for forensic applications. We have also highlighted the importance of SNPs and amalgamation of NGS kits in forensic application. Notably, the importance of wildlife forensic has been discussed for the identification of species as well as its geographic origin. Another important budding aspect of RNA-based identification of forensically relevant biological fluids has also been discussed in much detail.

Keywords DNA fingerprinting · Criminal justice system · STRs · Forensic analysis

J. Imam (✉) · R. Reyaz · A. K. Rana · V. K. Yadav
DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home, Jail and Disaster Management, Government of Jharkhand, Ranchi, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_1

1.1 Introduction

DNA or deoxyribonucleic acid is the hereditary polymer molecule made up of four different nucleotides, i.e., adenine (A), thymine (T), guanine (G), and cytosine (C). These nucleotides arranged in a definite sequence represent the genetic makeup of all known living organisms and some viruses. In humans, the DNA is present within the nucleus of every cell (except mature red blood cells) and is organized into 46 different chromosomes (22 pairs of autosomes and one pair of sex chromosomes). A haploid DNA is approximately three billion base pair in length. Additionally, 99.9% DNA sequences are similar between two individuals [70]. The difference of 0.1% of the DNA sequence between two individuals is usually present along the noncoding regions of the DNA. This difference of sequences is targeted by the forensic scientists to discriminate between individuals by matching DNA isolated from varied samples, i.e., blood, tissue, body fluid, or hair/nails, with high level of accuracy. The difference in DNA sequence basically occurs in terms of around 2–9 base pair repetitions called as microsatellites or of 10–100 bps called as minisatellites [39, 57]. In earlier times, Sir Alec Jeffreys analyzed minisatellites or variable number tandem repeats (VNTR) using Southern blotting to visualize DNA markers [73]. However, currently, microsatellites of around 4 bp repeats are being analyzed along with fluorescent tags of respective markers using multiplex amplification of at least 20 core loci [46, 71] as defined by the Federal Bureau of Investigation (FBI), USA [12, 75]. These loci along with their respective repeat patterns are inherited from parents to their offspring by a random process during meiosis, and every individual can be discriminated from another (as calculated by FBI among Caucasian population) by a probability factor of 1.683×10^{-15} [20], i.e., one individual can have the same DNA pattern with another only in a population of 594 trillion. The exception is monozygotic twins who carry the same loci with the same repeat pattern.

1.2 Discovery

DNA fingerprinting was initially used to find human genetic diseases by linking particular DNA sequences with the help of segregating markers which were present in close proximity within a chromosome [12, 15]. Eventually, it was also used for criminal investigations and forensic science, when an undaunting Ph.D. scholar Alec J. Jeffreys from Leicester University, United Kingdom, took a scientific responsibility to nab the culprit of famous twin girl's rape and murder case from Narborough using classical DNA fingerprinting assay using VNTR method [12, 15, 73]. The geneticist Alec J. Jeffreys discovered that short repetitive DNA sequences are almost unique to every individual, and he called them "minisatellites."

In 1984, at the Leicester University, UK, Dr. Jeffreys was studying hereditary diseases in families and was also focusing on developing methods to resolve

paternity and immigration disputes by linking genetic elements between individuals, and he published his results in *Nature* journal [28, 29]. He discovered that variable number tandem repeats (VNTR) is a part of junk DNA and these repeats vary from one individual to another. He employed various restriction enzymes in his technique to cut these variable DNA sequences and generate length polymorphism. He then run these digested sequences in a very long agarose gel and demarcated the variations from individual to individual. He named his technique "genetic fingerprinting" and demonstrated that this restriction fragment length polymorphism (RFLP) of DNA was unique to each individual and cannot match on earth except for identical twins. However, his technique was launched into the world of forensic science when two murders were committed in a little village of Narborough in the city of Leicestershire, east of Birmingham, UK. In 1983, 15-year-old lady Lynda Mann was found raped and murdered, and in the same village after 3 years in 1986, Dawn Ashworth, also 15, was also raped and murdered in similar manner. Dr. Jeffreys was asked to do DNA analysis of the semen samples recovered from the two victims with the blood sample of a suspect named Richard Buckland, aged 17. Some of the detectives of this case were in suspicion as Buckland was below 14 years of age when the first crime occurred. From the DNA analysis of the samples, Dr. Jeffreys demonstrated and proved that the same killer's semen is present in both the crime scenes and this does not match with Buckland blood sample's DNA. After this, the law enforcement took an exhaustive task to match the blood types and then DNA of total 4582 men from the three towns. Dr Jeffrey did genetic fingerprinting of the 10% men who were tested same for blood group (Group A) and isoenzyme, i.e., phosphoglycerate mutase (PGM). However the DNA could not match with any one of them. After several months, a resident heard a conversation in a pub, where a person named Ian Kelly confessed that he had taken bribe for replacing his photo with a photo of a local baker named Colin Pitchfork in his passport and had submitted his own blood instead of Pitchfork [32]. Twenty-seven-year-old Colin Pitchfork was arrested on September 19, 1987, and when his DNA was compared with the semen samples, it was indistinguishable, i.e., identical. Pitchfork was found guilty in both rapes and murders, and in 1987, Pitchfork became the first person in the world to be identified and convicted as a result of the DNA fingerprinting. He was sentenced to life imprisonment on January 22, 1988 for both murders and is currently in jail. In 1994, Professor Jeffreys was knighted by The Queen of England, and today he is still a professor in the Department of Genetics at the University of Leicester in Great Britain.

1.3 Advancement and Milestones in DNA Fingerprinting

DNA fingerprinting, since its discovery around two and half decades back, has taken a great leap in its advancement and made the justice delivery system more efficient and accurate in the investigation of criminal and civil cases [28–30]. This is much like a valuable armory in the hands of judiciary which aids in the conviction

of the guilty as well as exoneration of the innocent [10]. It has also been proven helpful in linking relationship of reference samples to dead remains of missing person and in mass disasters like plane crash, vehicle collision, earthquakes, etc. [10, 51]. With the discovery and innovation of new techniques for DNA extraction and genotyping, the generation of DNA profiles is becoming more and more accurate and easy, even for challenged and trace DNA samples. The known fact about DNA fingerprinting is its uniqueness among human populations, and this attribute makes it the method of choice [19]. DNA profiling generally involves the five basic steps from sample preparation, DNA extraction, DNA quantitation, DNA amplification to capillary electrophoresis, and profile generation [60]. With the advancement in different fields of science, new technologies are regularly introduced and validated in forensic laboratories to aid the process of DNA fingerprinting with improved sensitivity and informativeness. Day by day, crimes are increasing which poses a paramount pressure on judiciary system. In this regard, the use of automation techniques for sample preparation and data interpretation by the forensic laboratories will be useful to meet this increasing throughput demands on the laboratories [10].

DNA fingerprinting requires good scientific skills and data interpretation ability which is essential in result outcome. Earlier, the methodology used for DNA extraction to profile generation had lots of limitations for the type and quality of biological samples available for forensic investigation. The advancements in genomic and post-genomic era have put the forensic science one step ahead, and now it is much faster, higher, and stronger in crime investigation and judiciary system [10], *faster* in terms of rapid DNA instrumentation, recovering *higher* and good quality of data from biological evidences and *stronger* conclusion on complex evidences [10]. The development in forensic DNA analysis is possible because the pioneer work, progression, and innovation transpire over the past three decades (Table 1.1) [9, 60].

1.3.1 Current Status of DNA Fingerprinting Technique

The discovery of DNA fingerprinting is nothing less than bliss, not only to the scientific world but also to the judiciary system. Alec Jeffreys laboratory was the only one to work on DNA fingerprinting during 1985–1987, and his work and contribution in solving civil and criminal cases with the help of DNA fingerprinting is pioneer in establishment and adoption of this technique in judicial investigations worldwide. The last three decades are the golden periods in the field of forensic science with the advent and implementation of new techniques, the use of advanced commercial DNA typing kits with various genetic marker systems, and NGS in forensic science.

Since the discovery of DNA fingerprinting and its application in forensic judiciary system, the procedure for biological sample collection from the crime scene and DNA extraction methodology from different biological specimens are well established [24, 25, 43, 58, 60, 65, 69, 74]. Here we will discuss the advancement and emerging techniques and methodologies of DNA fingerprinting. Forensic sci-

Table 1.1 Timelines and milestones for forensic DNA analysis

Timeline	Major area	Activities and milestones
1985–1995	Development and exploration	1. Discovery of DNA fingerprinting and first publication by Alec Jeffreys
		2. RFLP using VNTRs
		3. Rapid and sensitive PCR assays
		4. Formation of DNA profiling groups
1995–2005	Stabilization and standardization	1. Improvement in PCR technology
		2. Development of sensitive, fast, and genotypic precision STR-based DNA analysis
		3. National DNA database formation by the United Kingdom with six STR markers
		4. 13 core STR markers designated by the United States for FBI CODIS software
		5. Multiplexing and capillary electrophoresis
		6. Autosomal and Y-STR kit released
		7. Expansion of forensic research in different laboratories
		8. Importance of Human Genome Project (HGP) in forensic DNA technology development
2005–2015	Augmentation	1. Faster and efficient DNA extraction methods adopted
		2. Automation
		3. Accumulation of DNA database
		4. Advancement in different STR kits
		5. Development and validation of new STR kits
		6. Encroachment of DNA forensic in wild life
		7. Forensic DNA phenotyping (FDP)
		8. Implementation of microbial DNA fingerprinting of human fingerprints
		9. Advancement in forensic RNA typing
		10. Introduction and exploitation of NGS in forensic science
2015–2025	Sophistication	1. Development of tools for rapid DNA testing outside of laboratories or at crime scene
		2. Huge population database formation
		3. Development of cost effective forensic DNA analysis
		4. More better multiplexing system and DNA typing
		5. More research in the field of wild life forensic, microbial DNA fingerprinting and forensic RNA typing
		6. Development of cheaper and more robust STR kits
		7. Development of skills and ability to interpret forensic evidence results
		8. Higher sensitive methodologies applied to casework and probabilistic software approached to complex evidence

ence has gone through several stages of development since its discovery and application in the 1980s [19]. The first generation of DNA analysis RFLP-based profiling is obsolete now from the forensic point of view, because it was not suitable for degraded and challenging biological forensic samples as it was not able to analyze the samples with accuracy. The PCR-based second generation of DNA analysis based on dot-blot methods was then developed but could not be fruitful enough as it was not helpful with DNA fragments which were longer in length. The third generation is STR (short tandem repeats) based which is easy, suitable, and most widely accepted for DNA analysis, but, sometimes, for highly degraded DNA samples, getting DNA profile becomes difficult. The fourth generation of DNA analysis which is introduced in forensic science is NGS (next-generation sequencing) which has attracted the forensic community with its high-throughput capacity and low cost [4]. There is a continuous effort to develop more effective, cheap, and fast DNA profiling techniques with more discriminatory power to address the application of forensic science in different fields (Table 1.2). Here, we have highlighted some of the recent progresses made in the analysis of STRs, SNPs, low-template DNA, mitochondrial DNA, DNA methylation, microbial forensic, and NGS in forensic and illustrate how different technologies can be integrated for new-generation forensic science.

1.3.1.1 Evolution of Capillary Electrophoresis as a Tool for Forensic DNA Analysis

Capillary electrophoresis (CE) is one of most important instrumental advancements to be implemented for forensic DNA typing. After PCR invention, scientists consider it as the second most needed development. Application of capillary electrophoresis in forensic DNA analysis not only makes the work easier but also makes it more accurate and authentic, which is of paramount importance in forensic DNA analysis [63]. The application of capillary electrophoresis is not only limited for biological samples but also has a huge importance in the analysis of gunshot residues, explosive residues, and drugs. For forensic DNA analysis, STR profiling (highly polymorphic markers) which is based on fragment analysis is of great value for human identification (HID) due to the single-base resolution capability of CE [63]. Introduction of capillary electrophoresis in STR typing circumvents the tedious and expensive approach of DNA sequencing for STR typing. The approach of CE like precise sizing, its sensitivity for the detection of fluorescence emitted by different dyes, automatic electrophoresis, and data collection software are key factors in the worldwide adoption of CE as the preferred platform for forensic DNA analysis. The most common CE systems used in forensic DNA analysis include the ABI PRISM® 310, 3100, 3100 Avant, 3130, 3130xL, 3500, and 3500xL Genetic Analyzers (GAs). The advanced CE automated machines are developed with advanced features which are useful for forensic scientists. It has many advantages like normalization of peak height, accurate sizing of fragments, sample injection, single-base resolution, high run-to-run precision, good temperature control and

Table 1.2 Different commercial STR kits for forensic application produced after the year 2000

S.No.	STR kit name	No. of markers	Application and advantages	Make
1	Identifiler™, IdentifilerPlus™ and IdentifilerDirect™	Amelogenin Plus 15 marker (13 CODIS + D2S1338 and D19S433)	Greater sensitivity and improved performance on inhibited forensic samples. More robust master mix and modified thermal cycling parameters	Life Technologies
2	PowerPlex® 16 and PowerPlex® 16HS	Amelogenin Plus 13 CODIS marker	Enhanced buffer system with improved robust performance	Promega Corporation
3	PowerPlex® 18D	Amelogenin Plus 15 marker (13 CODIS + D2S1338 and D19S433 + Penta E and Penta D)	Improved allelic ladder featuring many additional alleles not found in original PowerPlex® 16 PowerPlex® 16HS kit. It is also compatible with direct amplification from FTA card punches as well as non-treated paper	Promega Corporation
4	NGM™	Amelogenin Plus 15 marker (ENFSI loci -mini- and midi-STRs)	It enables simultaneous, multicolor fluorescence detection in a single PCR by using five-dye chemistry	Life Technologies
5	NGM SElect™ and NGM SElect™ express	Amelogenin Plus 16 marker (ENFSI loci- mini- and midi-STRs)	It enables simultaneous, multicolor fluorescence detection in a single PCR by using five-dye chemistry	Life Technologies
6	PowerPlex® ESX-16 and PowerPlex® ESI-16	Amelogenin Plus 15 marker (ENFSI loci – mini- and midi-STRs)	It enables simultaneous, multicolor fluorescence detection in a single PCR by using five-dye chemistry	Promega Corporation
7	PowerPlex® ESX-17 and PowerPlex® ESI-17	Amelogenin Plus 16 marker (ENFSI loci – mini- and midi-STRs)	It enables simultaneous, multicolor fluorescence detection in a single PCR by using five-dye chemistry	Promega Corporation
8	ESSplex and ESSplex SE	Amelogenin Plus 15 marker (ENFSI loci – mini- and midi-STRs)	It enables simultaneous, multicolor fluorescence detection in a single PCR by using five-dye chemistry	Qiagen, Hilden, Germany
9	PowerPlex® ES	Amelogenin Plus 8 marker (SE33 locus)	Addition of highly polymorphic SE33 locus. It has limited use in relationship testing due to highest mutation rate	Promega Corporation

(continued)

Table 1.2 (continued)

S.No.	STR kit name	No. of markers	Application and advantages	Make
10	SEfiler™ and SEfilerPlus™	Amelogenin Plus 10 marker (SE33 locus)	Addition of highly polymorphic SE33 locus. It has limited use in relationship testing due to highest mutation rate. Improved synthesis and purification processes, enhanced sensitivity for inhibited samples	Qiagen, Hilden, Germany
11	PowerPlex® 21	Amelogenin Plus 20 marker (13 CODIS + D2S1338 and D19S433 + Penta E and Penta D + 3 marker)	It can work with a variety of sample types, including casework samples. It is also compatible with direct amplification from FTA card punches as well as non-treated paper	Promega Corporation
12	GlobalFiler™ Express	Amelogenin Plus 23 marker (13 CODIS + D2S1338 and D19S433 + 7 marker + Y Indel gender loci)	Additional loci included (MiniFiler loci) which are designed to meet the expanded US core loci requirements. It is optimized for efficient amplification of low level DNA and to overcome common inhibitors of the PCR. Optimized for the amplification of single-source samples	Life Technologies
13	PowerPlex® Fusion	Amelogenin Plus 23 marker (13 CODIS + D2S1338 and D19S433 + Penta E and Penta D + 6 marker)	Additional loci included (MiniFiler loci) which are designed to meet the expanded US core loci requirements. It is a dual-purpose kit in that it can be used for common casework samples as well as direct amplification of reference samples stored on paper with only minor changes to the PCR amplification conditions	Promega Corporation
14	PowerPlex® CS7 System (nonstandard STR marker system)	Seven STR loci	It is used as a confirmatory kit in paternity applications	Promega Corporation
15	Investigator HDplex kit (nonstandard STR marker system)	Amelogenin Plus 13 marker (highly polymorphic markers)	It is developed specifically to discriminate closely related individuals. It is designed for difficult forensic and paternity cases	Qiagen, Hilden, Germany
16	MiniFiler ™	Amelogenin Plus 8 CODIS marker	For genotyping degraded DNA samples. First commercial kit designed to amplify miniSTRs	Life Technologies
17	PowerPlex Y	12 Y-STR loci	First sex-chromosome STR kit developed to identify male lineages	Promega Corporation

18	Yfiler ™	17 Y-STR loci	Most commonly used sex-chromosome STR kit to identify male lineages as it works well in most outbred populations	Applied Biosystems
19	Argus Y-12 QS	12 Y-STR loci + internal control	Sex-chromosome STR kit developed to identify male lineages. The internal control system provides helpful information about PCR efficiency and about the presence of inhibitors in tested samples	Qiagen, Hilden, Germany
20	PowerPlex® Y23	12 Y-STR loci of Yfiler Kit plus six additional new informative loci for male-lineage differentiation	It allows Y-STR analysis of both human forensic samples and database samples. It features fast amplification time and better tolerance of inhibitors of the PCR when compared to previous generations of Y-STR multiplexes	Promega Corporation
21	Argus X-12 kit	12 X-STR loci	Simultaneous amplification of 12 X-chromosomal markers for kinship and paternity testing, as well as population genetics and anthropological studies. Also suited for forensic stains, such as female traces in male background	Qiagen, Hilden, Germany

automation, better sensitivity, high throughput, user-friendly, and easy software features to analyze the raw data to the level of precise accuracy [63]. Definitely the incorporation of CE in forensic application must be considered as a milestone for the service of the mankind.

1.3.1.2 STR and Next-Generation STR Genotyping Kits for Forensic Application

Forensic DNA typing has been constantly evolving driven by innovations from academic laboratories as well as kit manufacturers [47]. Much technological advancement took place during the last 30 years, but the PCR-based STR genotyping is central to all. STRs are now the markers of choice for various human identification (HID) applications as the STR loci are considered polymorphic as they are unique to each individual [8]. The basis of individual identification by STRs is the measurement of length of different alleles which exhibit the highest variability among individuals [21]. Mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats of STRs are available, but tetranucleotides are commonly used and preferred in STR analysis because the chance of stutter production is minimal and it can analyze the amplicons that are one repeat less than the true allele [60]. Now, PCR-based multiplexing of STRs and capillary electrophoresis enables the analysis of several different loci at the same time with better accuracy and ready to available data for interpretation [44]. In 1997, the United States established a core set of 13 STR loci known as the Combined DNA Index System (CODIS) loci. The 13 STR loci set had strong distinguishing power with average random match probability of one in a quadrillion (1×10^{-15}). After the establishment of CODIS loci, Promega Corporation (Madison, Wisconsin) and Life Technologies developed the commercial kits to meet the demand of the forensic community [47].

1.3.1.3 Multiplex STR Kits

After 2000, Promega designed the PowerPlex 16 System, and Life Technologies came up with AmpFLSTR Identifiler PCR Reaction kits which were capable of amplifying all 13 CODIS markers in a single PCR reaction along with the amelogenin sex identification marker [14, 38]. The AmpFLSTR Identifiler PCR Reaction Kit and the PowerPlex 16 System have been widely used for forensic database generation and casework analysis, both. This helps in the generation and addition of millions of STR profiles to the DNA databases helping the criminal judiciary system. With huge criminal cases happening frequently, forensic laboratories are continuously seeking and adopting enhanced technologies which help them to process database and casework samples more efficiently and effectively. The AmpFLSTR

Identifiler Plus PCR Amplification Kit and PowerPlex HS System were developed for various challenging forensic samples for greater sensitivity and improved performance. Various autosomal STR kits have been produced after the year 2000 for challenging and inhibited biological samples which are listed in Table 1.2. Later on in 2011, CODIS announced an increase in the number of CODIS loci to reduce the likelihood of adventitious matches, to increase international compatibility, and to increase the discrimination power (Mulero and Hennessy). Life Technologies developed GlobalFiler™ and GlobalFiler™ Express PCR Amplification kits, and Promega Corporation developed PowerPlex Fusion System as a next-generation STR kit to meet this challenge by incorporating extra loci in the kits, hence making it more robust (Table 1.2). For genotyping degraded DNA samples, the first commercial STR kit, "MiniFiler Kit," was developed to amplify "miniSTRs." The improvement was done to amplify degraded samples by repositioning the primers as close to STR repeat region, as possible (Table 1.2) [11, 13, 66]. New STR kits were developed for human identification by incorporating highly polymorphic STR locus, SE33 (a tetranucleotide STR). Inclusion of SE33 in multiplex has shown a huge advantage due to its structural variations and polymorphism with differentiation of around 1 bp [55, 59]. This marker also possesses the highest mutation rate, thus limiting its application in forensic use such as relationship testing [16, 49]. PowerPlex® CS7 System (Promega Corporation) and Investigator HDplex Kit (Qiagen), a nonstandard STR marker system, were also developed for very special cases which involved in kinship analysis and testing with samples deficient of close relatives. It has been developed specifically to discriminate closely related individuals for difficult forensic and paternity cases (Table 1.2). Many sex-chromosome STR kits were also developed to identify male and female lineages (Table 1.2).

The next-generation commercial kits listed in the Table 1.2 have been developed by improving the performance with inhibited samples, increased sensitivity, high throughput with direct amplification, increased CODIS loci, and paternity- and lineage-specific STRs which provide a "driving force" for the progress of forensic science.

1.3.1.4 ABO Typing with Multiplexes STRs

Few years back, Jiang et al. [31] developed a successful technique, in which both ABO genotyping and STR analysis were combined in a single reaction, where a forensic scientist could get both the information from a biological sample in a single reaction. The developed technique has a combination of all the 15 autosomal STR loci, gender-determining locus amelogenin, and markers for six ABO genotypes. This was an important development as it is more accurate and has better confirmatory power than normal ABO blood group typing.

1.3.1.5 Autosomal SNPs and Indels in Forensic Analysis

Many STRs have been developed (autosomal, MiniFiler, and sex chromosome STRs) in lieu to overcome the challenges of generating DNA profiles from forensic samples. This is only possible because of the greater variability of DNA polymorphisms. But forensic science always faces tough challenges in terms of low-level DNA or degraded DNA, to obtain complete STR profiles. Degraded DNA or low template (LT) causes problems in STR typing either with allele dropouts or allele drop-ins which in many times are unable to get even with increased sensitivity and miniSTR typing [1, 2]. The possible alternatives to increased PCR sensitivity of STRs for degraded DNA or low template (LT) are to use SNPs and insertion/deletions (indels). SNPs used in LT DNA can result in fewer allele drop-ins [5]. SNPs (single-nucleotide polymorphisms) and indels (insertion/deletion polymorphisms) are the most common short binary markers of the human genomic variations. SNPs allow allele detection of comparatively small amplicon size ranges (about 41 bp) and therefore can be a better option with highly degraded samples [6]. But still, SNPs have not been whole heartedly adopted in forensic science as the marker of choice, for highly degraded samples, as new and advanced STR technologies have come up which can enhance the profiling performance even for highly degraded DNA [7]. Therefore, SNPs are not very much suitable for normal forensic casework and database entries when working with DNA mixtures. It is most suitable for identification of missing persons and relationship establishment. STRs present better information than single SNPs, but as we increase the number of SNP markers, they could provide better discriminating power. One of the valuable characteristics of SNPs is variation at heterozygosity level of the genome. Other important advantages of SNPs is that it does not require separations on the basis of size which makes multiplexing and automation easier compared to the STR analysis and low mutability rate making it more stable genetic marker [18].

Two SNP multiplexes have been developed for forensic identifications: a 52-SNP assay developed by the SNPforID consortium, comprising a 52-plex PCR followed by tandem 21- and 29-plex primer extension reactions [50, 61] and a 44-plex PCR followed by tandem 18- and 26-plex extensions [41] based on the Kidd Lab forensic identification marker panels, consisting of a list with almost twice as many ID-SNPs than the 44 collated in this assay [53]. The 52-plex multiplex is best suited for highly degraded DNA samples, i.e., for very old skeletal remains or body recovered from the river or sea. This shows its importance in identification of missing person. Here, it is also important to note that the number of SNPs required to match the informativeness of STRs is higher in relationship testing than in identification applications. Table 1.3 listed a few techniques and variations of SNP analysis.

Indels (insertion/deletions) comprise about 5% of known polymorphisms in human genome and are thus considered as potential markers in forensic identification as they have combined the application of both SNPs and STRs [48]. The advantages of indels over SNPs are as follows: first, ease of analyzing indels from very short amplicon size as compared to SNPs, and, second, the ease of doing indel analysis by combining the advantage of direct PCR-to-capillary electrophoresis

Table 1.3 List of few techniques and variants of SNP analysis

Sl. No.	SNP techniques	Details	Application	References
1	SNaPshot assay	1. Capable of multiplex SNP analysis in small amount of template DNA samples	SNPs analysis for degraded DNA samples	
		2. Highly sensitive, can analyze samples shorter than 70 bp amplicons		
2	TaqMan assay	1. Capable of multiplex SNP analysis in small amount of template DNA samples	SNPs analysis for degraded DNA samples	
		2. Highly sensitive, can analyze samples shorter than 70 bp amplicons		
3	Y-SNPs multiplex system	1. It is sensitive, human specific	Applicable for Chinese Han population	
		2. Informative in cases of degraded DNA and male-male mixture		
		3. Resistant to high background of female DNA in male-female mixture		
4	Mitochondrial DNA SNPs (mtSNPs)	1. Mt DNA SNPs are lineage markers	Lineage-specific SNP analysis for degraded teeth and bone DNA samples	
		2. Not capable of genetic individualization		
		3. Can be employed for degraded bone and teeth samples		
5	Nucleosome SNP assay (18-plex single-base extension assay)	1. Based on the principle that histone-DNA complexes found in nucleosomes offer protection from DNA degradation process	Best suited for highly degraded DNA samples. It provides a new marker set that can be used to supplement existing SNP assays	Freire-Aradas et al. [18]
		2. It offers better results than other existing forensic SNP assays		
6	21 SNP multiplex system	1. Target amplicon length for 21 SNPs is from 63 bp to 192 bp	This system provides a reliable technique for individual identification of nondegraded and degraded DNA samples	
		2. More efficient in the analysis of degraded DNA compared with standard STR typing		

typing, as this is not possible with SNP typing using SNaPshot assay. Different indel typing kits were developed like Investigator DIPplex Kit (Qiagen, Hilden, Germany) and indel-plex. Both have shown successful application for typing highly degraded DNA samples in forensic science.

1.4 Next-Generation Sequencing and Its Application in Forensic Sciences

Next-generation sequencing (NGS), the technology which has overcome the limitations of conventional Sanger sequencing, has grown rapidly in recent years in the field of genomics research because of its high-throughput capacity and low cost and ancient DNA analysis [54, 62]. Over the last 10 years, NGS methods and platforms have evolved, and sequencing quality has now reached a level where NGS can be launched in forensic science, and in the last 2 years, there has been an explosion in scientific articles, with forensic applications of NGS. Since the number of casework samples which require DNA processing is increasing day by day, the CE-based methods which have fixed capabilities are sometimes unable to stand. Therefore, NGS technology and platforms show promising results in DNA testing and identification of missing persons, kinship testing, ancestry investigation, and other human identification applications. In NGS technology, simultaneous amplification of multiple STR marker types and SNPs can be achieved in a single run for large number of samples.

1.4.1 Next-Generation Sequencing Kits or Systems in Forensic

1.4.1.1 HID-Ion AmpliSeq Ancestry Panel

HID-Ion AmpliSeq ancestry panel (Life Technologies) enables simple and fast target selection of hundreds of SNPs using multiplex PCR. Thousands of primer pairs can be used in a single tube for target amplification followed by next-generation sequencing (NGS) on the Ion PGM™ System. This ready-to-use panel consists of 165 autosomal markers that provide biogeographic ancestry information. Fifty-five of these markers were selected based on a poster by Dr. Kenneth Kidd [35, 36], and 123 markers were selected based on a publication by Dr. Michael Seldin [37]. Ion AmpliSeq technology makes it possible to multiplex 165 PCR reactions in one tube with only 1 ng of input DNA. With small amplicon sizes, the panel is optimized for degraded DNA samples that provide the biogeographic ancestry information and guide the investigation process.

1.4.1.2 Precision ID NGS System for Human Identification

Precision ID NGS System for human identification (Applied Biosystem) for human identification can help in solving tough cases by getting more information from the challenging samples. It is the combination of Ion Chef System and Ion S5 or Ion S5 XL Systems with forensically relevant Precision ID panels that utilize Ion AmpliSeq technology. It includes the same 21 autosomal STRs, along with Y indel and amelogenin sex marker found in the GlobalFiler DNA amplification kit. Instead of using SE33, this panel includes nine additional multiallelic STR markers (for a total of 33 targets) to aid in mixture interpretation of complex casework samples.

1.4.1.3 ForenSeq™ DNA Signature Prep Kit

ForenSeq™ DNA Signature Prep Kit (Illumina) is developed by incorporating multiple STRs kits which include over 200 forensically relevant genetic markers in a single, streamlined workflow, eliminating the need for multiple STR kits [64]. This kit includes global autosomal STRs, Y-STRs, X-STRs, identity-informative SNPs (iiSNPs), phenotypic-informative SNPs (piSNPs) (eye and hair color), and biogeographical ancestry SNPs (aiSNPs) in a single platform, which is not available with traditional CE-based methods. This kit overcomes the limitations of other CE-based kits for degraded, mixed, or PCR-inhibited samples.

1.4.1.4 PowerSeq™ Systems

PowerSeq™ systems (Promega) for forensic identification include three systems: (a) PowerSeq™ Auto includes 23 STR loci and amelogenin, (b) the PowerSeq™ Mito system generates ten small amplicons (adapted from Eichmann and Parson) covering the control region of the mitochondrial genome, and (c) PowerSeq™ Auto/Mito/Y combines both sets of amplicons in one multiplex plus 23 Y-STR loci. PowerSeq™ Auto system is a 24-plex kit for analyzing autosomal STRs, amelogenin, and DYS391. PowerSeq™ Mito system is based on sequencing of the mtDNA control region (HV1 and HV2). PowerSeq™ Auto/Mito/Y system has been configured for the simultaneous analysis of 22 autosomal STRs, amelogenin, 23 Y STRs, and the control region of the mitochondrial genome [3, 17, 72].

1.4.2 Forensic Application Prospects of NGS Technology

Forensic science technology has embraced DNA technology as the main weapon to address various crimes and help the judiciary. Today, PCR- and CE-based DNA typing is the backbone of forensic science and criminalistics. Various STRs

Table 1.4 Forensic application prospects of NGS technology

Sl. No.	Application	Advantages
1	STR typing	High throughput, low cost, simultaneous detection of large numbers of STR loci (autosomal and sex chromosome STRs), and the ability to distinguish alleles with similar length facilitate the identification of mixed DNA samples and analysis of complex paternity cases
2	Mitochondrial genome analysis	Important in maternal lineage identification; whole mitochondrial sequence for identification with high discrimination power
3	Y-chromosome analysis	Establishes paternal relationship between male individuals
4	Forensic microbiological analysis	NGS is suitable for whole-genome typing of microbial pathogens during forensic and epidemiological investigations which can detect even the rare polymorphisms and thus give forensic data higher resolution and greater accuracy for accurate identification of criminals and biological terrorists
5	Animal and plant DNA analysis	NGS technology has allowed the DNA typing in plant and animal species identification
6	Ancestry studies and phenotypic inferences	NGS technology for whole-genome sequencing will be helpful in determination of ancestry and personal characteristics like ethnicity, physical and physiological characteristics, and age
7	Epigenetic analysis	Epigenetic changes like methylation pattern can be easily studied and employed in forensic genetics with the aid of NGS technology
8	MicroRNA analysis	NGS has proved to be very critical in analysis of millions of miRNA sequences for rapid identification of organ and developmental stage-specific expression and expression in different diseases which is a powerful tool in forensic analysis

(autosomal, miniSTRs, sex chromosome, phenotypic STRs) are CE based, and CE still is the method of choice for forensic analysis because of its accuracy, specificity, discriminatory power, and easy handling. The use of NGS in forensic science as in human identification (HID) and determination of phenotypic traits leads to its larger application in forensic analysis. Definitely NGS technology has a lot of advantages over tradition CE-based typing, and there is little doubt that NGS will be implemented and used in forensic laboratories in the future. Table 1.4 presents the overview of NGS application prospects in forensic application.

1.5 RNA Profiling and Its Application in Forensic Science

The presence of biological evidences at the crime scene and its correct screening for the possible source of DNA have always been the challenges for the forensic expert. Many conventional biochemical and immunological assays are there for the screening of biological fluids, but they are time-consuming and laborious and even consume the important evidentiary material. Because of this problem, many forensic scientists bypass these preliminary screening processes and directly proceed for

DNA analysis which many times lead to failure to provide the information regarding the nature of crime [23]. In recent times, molecular approaches for the identification of body fluids have been developed which have significantly improved the sensitivity. The use of RNA profiling strategies is considered the better option for the identification of forensically relevant biological fluids and tissues such as saliva, vaginal secretions, menstrual blood, and skin [23].

The forensic identification of human body fluids and tissues by means of messenger RNA (mRNA) profiling is a long-studied methodology that is being increasingly applied to casework samples [68]. From a singleplex PCR technique to multiplex RT-PCR platform, mRNA profiling has evolved in a big way in providing expression of data on multiple genes simultaneously [40]. A single mRNA-based system, 19-plex system, has been developed for the discrimination of common forensic body fluids as well as skin cells [40]. This 19-plex system is able to establish both the donor and the cell type of the samples. This 19-plex mRNA assay showed good results with body fluids, with high sensitivity and specificity. The 19-plex mRNA assay targets six different cellular origins to provide better assessment, which is important in forensic casework.

1.6 Wildlife Forensics

Soon after the discovery, the forensic DNA profiling has revolutionized criminal investigation process in humans. Today, forensic DNA analysis has become an indispensable tool for different criminal cases and helps in the arrest of many perpetrators as well as exonerations of many innocent individuals who were wrongly convicted. As the forensic DNA analysis is growing day by day and its applications are also being introduced in other fields, its need is also felt in the investigation of wildlife-related crimes and wildlife conservation. Wildlife and their products constitute the third most illegally traded commodity worldwide, after arms and drugs [34, 42, 45]. An important difference between crimes against humans and wildlife is that an animal becomes a “silent witness” to a human crime scene and it’s like no “victim” to provide information regarding the investigation. Another important issue with wildlife forensic science is that various species of animals or plants have to be analyzed as against single species in human identification cases [33]. Various sample types can be there in wildlife forensic science like whole animals (dead or alive), skins or skeletal of animals, exoskeleton and shells, and animal body parts such as the leg, wings, head, fur, scales, teeth, beak, claws, skin, carcass, horns, organ parts, blood samples, and many more [33]. Therefore, good and improved preservation techniques are required to support the prosecution in smuggling, poaching, and laundering of wildlife and their products [45]. In wildlife forensics, species identification is more important than individual identification. In addition, geographical region of the samples can also be analyzed.

Table 1.5 List of currently available techniques in wildlife forensic analysis

Sl. No.	Techniques used	Application	References
1	STR loci	1. Individual identification, i.e., investigation to determine if two samples are from the same individual	Johnson et al. [33] and van de Goor et al. [67]
		2. Pedigree analysis of a particular individual	
		3. Dinucleotide repeat STRs are common in identification of domesticated species	
2	Mitochondrial DNA (mtDNA) typing	1. Most commonly used in forensic DNA typing because:	Johnson et al. [33], Ivanova et al. [27], Guha and Kashyap [22], Imaizumi et al. [26], Pun et al. [56] and Osborne et al. [52]
		A. Mitochondrial loci are used for molecular taxonomy and phylogenetic	
		B. The presence of universal primers can be applied to unknown samples	
		C. It works well with highly degraded samples	
		2. Commonly used mitochondrial loci are cytochrome *b* (cyt*b*) and cytochrome oxidase 1 (COI)	
		3. Other loci are ribosomal RNA genes, D-loop region, and subunits of mitochondrial encoded NADH dehydrogenase	
		4. MtDNA sequencing for wildlife forensic identification	
3	Pyrosequencing techniques	1. Direct sequencing of thousands of small DNA fragments (degraded DNA) in a single run	Karlsson and Holmlund [34]
		2. Individual DNA typing of previously unknown species	
		3. It is rapid, accurate, and flexible and allows parallel processing	
4	NGS or high-throughput sequencing	1. Mass parallel sequencing for the identification of repetitive DNA sequences, for the identification of new STRs and SNPs	Johnson et al. [33] and Kidd et al. [35], [36]

DNA-based analysis is now frequently used and is most commonly applied in species identification cases. Now it's also being introduced for the analysis of geographic location of the species captured/claimed. The DNA markers applicable for the wildlife forensic science are different from human identification (HID) markers. As stated earlier, this is because of the complexities of species identification. STRs and SNPs are used for individual identification, pedigree analysis, and assignment of an unknown sample to a population [33]. Table 1.5 lists some of the currently available techniques of wildlife forensic analysis.

1.7 Conclusion

In this review, a brief overview of discovery, advancements, and milestones in the field of forensic DNA analysis has been outlined. Many new discoveries took place after the discovery of DNA fingerprinting, and these new approaches continue to be explored for more effectiveness. DNA fingerprinting relies most basically on the quality of DNA, DNA isolation from the biological samples has improved a lot, and good-quality DNA can be extracted even from highly degraded samples. The discovery of PCR and then advancements in real-time PCR have put the forensic science on the road of successful DNA fingerprinting. The advancement in STR technology and kits further improved the ability to decipher and interpret DNA results from challenging samples which provided the opportunity for future advances in forensic DNA analysis. The various forms of STR kits have revolutionized the field of forensic DNA profiling. The evolution in capillary electrophoresis technology and tough competition among the firms lead to cheaper and better kits, and this is the reason why forensic DNA fingerprinting has advanced so much, and now, it is helping the judiciary to solve the ever-increasing criminal and civilian cases. NGS has also knocked the door of forensics, and the day is not very far when it will be validated and incorporated as a conclusive technique to serve the judiciary system. RNA profiling is also proving to be a helping hand of forensic science in many relevant cases. The development in various molecular tools to investigate wildlife-related crimes has advanced the wildlife forensic analysis, and very soon, a well-placed technique will be available for individual and species identifications. Although various new techniques and scientific improvements are coming up, the current methods of STR typing are reliable, valid, and widely applicable.

Acknowledgment The authors are thankful to the Director of State Forensic Science Laboratory, Ranchi, Jharkhand, for the support.

References

1. Benschop CC, Haned H, de Blaeij TJ, Meulenbroek AJ, Sijen T (2012) Assessment of mock cases involving complex low template DNA mixtures: a descriptive study. Forensic Sci Int Genet 6:697–707
2. Benschop C, Haned H, Sijen T (2013) Consensus and pool profiles to assist in the analysis and interpretation of complex low template DNA mixtures. Int J Legal Med 127:11–23
3. Bornman DM, Hester ME, Schuetter JM, Kasoji MD, Minard-Smith A, Barden CA et al (2012) Short-read, high throughput sequencing technology for STR genotyping. BioTechniques:1–6
4. Borsting C, Mauling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18:78. https://doi.org/10.1016/j.fsigen.2015.02.002
5. Børsting C, Mogensen HS, Morling N (2013) Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. Forensic Sci Int Genet 7:345–352
6. Budowle B, van Daal A (2008) Forensically relevant SNP classes. BioTechniques 44:603–608
7. Budowle B, van Daal A (2009) Extracting evidence from forensic DNA analyses: future molecular biology directions. Biotechniques 46:339–340

8. Butler JM (2006) Genetics and genomics of core STR loci used in human identity testing. J Forensic Sci 51(2):253–265
9. Butler JM (2010) Chapter 3: historical methods. In: Fundamentals of forensic DNA typing. Elsevier Academic Press, San Diego, pp 43–78
10. Butler JM (2015) The future of forensic DNA analysis. Phil Trans R Sac B 370:20140252
11. Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 48:1054
12. Chambers GK, Curtis C, Millar CD, Huynen L, Lambert DM (2014) DNA fingerprinting in zoology: past, present, future. Investig Genet 5:3
13. Coble MD, Butler JM (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. J Forensic Sci 50:43–53
14. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR amplification kit. J Forensic Sci 49:1265–1277
15. Crawford MH, Beaty KG (2013) DNA fingerprinting in anthropological genetics: past, present, future. Investig Genet 4:23
16. Dauber EM, Kratzer A, Neuhuber F et al (2012) Germline mutations of STR-alleles include multistep mutations as denied by sequencing of repeat and flanking regions. Forensic Sci Int Genet 6:381–386
17. Eichmann C, Parson W (2008) "Mitominis": multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. Int J Legal Med 122:385–388
18. Freire-Aradas A, Fondevila M, Kriegel AK, Phillips C, Gill P et al (2012) A new SNP assay for identification of highly degraded human DNA. Forensic Sci Int Genet 6:341–349
19. Gaensslen RE, Harris HA, Lee HC (2007) Introduction to forensics & criminalistics. McGraw-Hill Companies, Inc. USA
20. 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
21. Giardina E, Spinella A, Novelli G (2011) Past, present and future of forensic DNA typing. Nanomedicine (Lond) 6:257–270
22. Guha S, Kashyap VK (2006) Molecular identification of lizard by RAPD & FINS of mitochondrial 16s rRNA gene. Legal Med (Tokyo, Japan) 8(1):5–10
23. Hanson EK, Ballantyne J (2013) Rapid and inexpensive body fluid identification by RNA profiling-based multiplex high resolution melt (HRM) analysis. F1000Res 2:281. https://doi.org/10.12688/f1000
24. Hopwood AJ, Elliott K (2012) Forensic DNA research: keeping it real. Int J Legal Med 126(2):343–344
25. Hudlow WR, Buoncristiani MR (2012) Development of a rapid, 96-well alkaline based differential DNA extraction method for sexual assault evidence. Forensic Sci Int Genet 6(1):1–16
26. Imaizumi K et al (2007) Development of species identification tests targeting the 16S ribosomal RNA coding region in mitochondrial DNA. Int J Legal Med 121(3):184–191
27. Ivanova NV, Clare EL, Borisenko AV (2012) DNA barcoding in mammals, methods. Mol biol (Clifton, NJ) 858:153–182
28. Jeffreys AJ, Wilson V, Thein SL (1985a) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73
29. Jeffreys AJ, Wilson V, Thein SL (1985b) Individual-specific 'fingerprints' of human DNA. Nature 316:76–79
30. Jeffreys AJ, Wilson V, Thein SL (1985c) Individual-specific 'fingerprints' of human DNA. Nature 314:67–74
31. Jiang X, He J, Jia F, Shen H, Zhao J et al (2012) An integrated system of ABO typing and multiplex STR testing for forensic DNA analysis. Forensci Sci Int Genet 6:785–797
32. Jobling MA (2013) Curiosity in the genes: the DNA fingerprinting story. Investig Genet 4:20

33. Johnson RN, Wilson-Wilde L, Linacre A (2014) Current and future directions of DNA in wildlife forensic science. Forensic Sci Int Genet 10:1–14
34. Karlsson AO, Holmlund G (2007) Identification of mammal species using species-specific DNA pyrosequencing. Forensic Sci Int 173:16–20
35. Kidd K et al. (2012a) Better SNPs for better forensics: ancestry, phenotype, and family identification. Poster presented at the National Institute of Justice (NIJ) annual meeting, Arlington VA, June 2012
36. Kidd KK et al (2012b) Expanding data and resources for forensic use of SNPs in individual identification. Forensic Sci Int Genet 6(5):646–652
37. Kosoy R, Nassir R, Tian C et al (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 30(1):69–78
38. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amiott EA, Sprecher CJ (2002) Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 47:773–785
39. Legendre M, Pochet N, Pak T, Verstrepen KJ (2007) Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Res 17:1787–1796
40. Lindenbergh A, de Pagter M, Ramdayal G et al (2012) A multiplex (m) RNA-profiling system for the forensic identification of body fluids and contact traces. Forensci Sci Int Genet 6:565–577
41. Lou C, Cong B, Li S et al (2011) A SNaPshot assay for genotyping 44 individual identification single nucleotide polymorphisms. Electrophoresis 32:368–378
42. Manel S, Berthier P, Luikart G (2002) Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. Conserv Biol 16:650–659
43. Matte M, Williams L, Frappier R, Newman J (2012) Prevalence and persistence of foreign DNA beneath fingernails. Forensic Sci Int Genet 6:236–243
44. McDonald J, Lehman DC (2012) Forensic DNA analysis. Clin Lab Sci 25:109–113
45. McGraw SN, Keeler SP, Huffman JE (2013) Forensic DNA analysis of wildlife evidence. In: Jaiprakash S, Ray L (eds) Forensic DNA analysis (current practices and emerging technologies). ISBN 9781466571266
46. Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R 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 25:175–181
47. Mulero JJ, Hennessy LK (2013) Next-generation STR genotyping kits for forensic applications. In: Jaiprakash S, Ray L (eds) Forensic DNA analysis (Current practices and emerging technologies). ISBN 9781466571266
48. Mullaney JM, Mills RE, Pittard WS, Devine SE (2010) Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet 19:R131
49. Müller M, Sibbing U, Hohof C, Brinkmann B (2010) Haplotype-assisted characterization of germline mutations at short tandem repeat loci. Int J Legal Med 124:177–182
50. 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
51. Nandinene MR, Prasad SPR, Goud CV, Negi DS, Nagaraju J, Gowrishankar J (2010) DNA-based identification of the victims of the Mangalore air crash of may 2010. Curr Sci 99:3
52. Osborne MJ, Christidis L, Norman JA (2002) Molecular phylogenetics of the Diprotodontia (kangaroos, wombats, koala, possums, and allies). Mol Phylogenet Evol 25(2):219–228
53. Pakstis AJ, Speed WC, Fang R et al (2010) SNPs for a universal individual identification panel. Hum Genet 127:315–324
54. Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B et al (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 311:392–394

55. Polymeropoulos MH, Rath DS, Xiao H, Merrill CR (1992) Tetra nucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-ac-psi-2 (ACTBP2). Nucleic Acids Res 20:1432
56. Pun KM et al (2009) Species identification in mammals from mixed biological samples based on mitochondrial DNA control region length polymorphism. Electrophoresis 30(6):1008–1014
57. Ramel C (1997) Mini- and microsatellites. Environ Health Perspect 105(Suppl 4):781–789
58. Richert NJ (2011) Swabbing firearms for handler's DNA. J Forensic Sci 56(4):972–975
59. Rolf B, Schurenkamp M, Junge A, Brinkmann B (1997) Sequence polymorphism at the tetranucleotide repeat of the human beta-actin related pseudogene H-beta-Acpsi-2 (ACTBP2) locus. Int J Legal Med 110:69–72
60. Romeika JM, Yan F (2013) Recent advances in forensic DNA analysis. J Forensic Res S12:001. https://doi.org/10.4172/2157-7145.S12-001
61. Sanchez JJ, Phillips C, Børsting C et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–1724
62. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. ProcNatl Acad Sci U S A 74:5463–5467
63. Shewale JG, Qi L, Calandro LM (2012) Principles, practice, and evolution of capillary electrophoresis as a tool for forensic DNA analysis. Forensic Sci Rev 24(2):79–100
64. Shrivastava P, Jain T, Trivedi VB (2016) DNA fingerprinting: a substantial and imperative aid to forensic investigation. Eur J Forensic Sci 3:23. https://doi.org/10.5455/ejfs.204929
65. Thomasma SM, Foran DR (2012) The influence of swabbing solutions on DNA recovery from touch samples. J Forensic Sci 58(2):465–469
66. Tsukada KK, Takayanagi H, Asamura M, Ota FH (2002) Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01, TPOX, CSF1PO, and vWA loci. Legal Med 4:239–245
67. Van de Goor LHP, Panneman H, van Haeringen WA (2009) A proposal for standardization in forensic bovine DNA typing: allele nomenclature of 16 cattle-specific short tandem repeat loci. Anim Genet 40:630–636
68. van den Berge M, Bhoelai B, Harteveld J, Matai A, Sijen T (2016) Advancing forensic RNA typing: on nontarget secretions, a nasal mucosa marker, a differential co-extraction protocol and the sensitivity of DNA and RNA profiling. Forensic Sci Int Genet 20:119–129
69. Vandewoestyne M, Van Hoofstat D, Franssen A, Van Nieuwerburgh F, Deforce D (2013) Presence and potential of cell free DNA in different types of forensic samples. Forensic Sci Int Genet 7:316–320
70. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291: 1304–1351
71. Wang DY, Gopinath S, Lagace RE, Norona W, Hennessy LK et al (2015) Developmental validation of the GlobalFiler((R)) express PCR amplification kit: a 6-dye multiplex assay for the direct amplification of reference samples. Forensic Sci Int Genet 19:148–155
72. Warshauer DH, Lin D, Hari K, Jain R, Davis C, Larue B et al (2013) STRait Razor: a length-based forensic STR allele-calling tool for use with second generation sequencing data. Forensic Sci Int Genet 7:409–417
73. Zagorski N (2006) Profile of Alec. J Jeffreys Proc Natl Acad Sci U S A 103:8918–8920
74. Zech WD, Malik N, Thali M (2012) Applicability of DNA analysis on adhesive tape in forensic casework. J Forensic Sci 57:1036–1041
75. Zietkiewicz E, Witt M, Daca P, Zebracka-Gala J, Goniewicz M, Jarzab B (2012) Current genetic methodologies in the identification of disaster victims and in forensic analysis. J Appl Genet 53:41–60

Chapter 2
DNA Fingerprinting Techniques for Forensic Application: Past, Present, and Future

Nisha Bara, Ramkishan Kumawat, and Jahangir Imam

Abstract The discovery of DNA fingerprinting has revolutionized the world of forensic science and started the wave that it can be helpful in solving crime cases and assisting the criminal justice system. It is one of the milestones in solving crimes with the help of DNA-associated polymorphisms in human beings. The DNA fingerprinting started with Restriction Fragment Length Polymorphisms (RFLP), which was tedious and time consuming but opened the doors for new developments in the arena. Later, with the development of CODIS, STR markers and now NGS have sped up the process of DNA profiling with better discriminating power and enhanced accuracy. The markers being used are short tandem repeats (STRs), species-specific primers, SNPs (single nucleotide polymorphism), NGS (next-generation sequencing), Y-STR, X-STR, and mitochondrial DNA (mtDNA). The identification of human from the DNA profile, generated through the Genetic Analyzer (by electrophoresis of amplified DNA), is the most favored method which is often used in sexual assault cases, paternity disputes, burning, and murder cases as it is believed that DNA is unique to each individual. In recent years, completely automated DNA-profiling system and diverse genetic markers have been introduced. The rapid DNA instruments integrate different steps such as DNA extraction, PCR amplification, separation, detection, sizing, and genotyping of the products on one single platform. This chapter is an insight on the development of DNA fingerprinting over the years and its application in forensic sciences.

Keywords DNA fingerprinting · DNA marker · NGS · Forensic investigation

N. Bara · J. Imam (✉)
DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home, Jail and Disaster Management, Government of Jharkhand, Ranchi, India

R. Kumawat
State Forensic Science Laboratory, Jaipur, Rajasthan, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_2

2.1 Introduction

DNA fingerprinting, also referred to as DNA typing or DNA profiling, is the most common forensic method which is used to identify an individual in criminal cases. In the year 1984, Sir Alec Jeffrey discovered the DNA fingerprinting while working in the Department of Genetics at the University of Leicester [9]. He found that there are patterns of genetic material that are unique to almost every individual and called them minisatellites. In the criminal justice system, DNA fingerprinting plays a vital role in saving the innocent and convicting the guilty. DNA fingerprinting is very useful in solving paternity and immigration disputes. Nowadays, the DNA recovering capabilities of laboratories have enhanced due to the introduction of various new technologies with enhanced sensitivity. Although different genetic markers are used for different applications, short tandem repeat (STR) typing continues to be at the top in forensic DNA analysis.

2.2 Methods Used in DNA Fingerprinting

2.2.1 RFLP Analysis: Restriction Fragment Length Polymorphism

At the beginning of DNA fingerprinting analysis, RFLP method was used. In this, a single restriction enzyme is used to cut the same stretch of genomic DNA into fragments of different lengths which are then separated according to their lengths by gel electrophoresis. Separated DNA fragments are transferred and fixed in the nitrocellulose or nylon membrane. These separated strands are then denatured and hybridized with radiolabeled DNA probes which are complementary to the repeated DNA sequences, thus favoring DNA/DNA hybridization (Southern blotting). These repeat sequences are called variable number tandem repeat sequence (VNTR), and they vary in length among different individuals except monozygotic twins. After hybridizing with probes, the blot is exposed to X-ray film. Distinct fluorescent bands are produced on the film by the DNA fragments with positive hybridization to the probe molecules. The band pattern on the autoradiograph is compared to find out whether the sample contains DNA from one or more person; hence, the contributors are distinguished. RFLP application provides the DNA fingerprint of a person, which can be used to compare the samples at a crime scene or to solve paternity disputes. RFLP analysis was used widely earlier, but it is time-consuming and difficult to automate tedious process requiring a greater amount of good-quality DNA compared to the polymerase chain reaction (PCR)-based DNA analysis technique. In forensics, RFLP analysis is not possible sometimes because samples obtained are degraded, challenged, and of poor quality.

2.2.2 Polymerase Chain Reaction (PCR) Analysis

There is a shift in the use of RFLP to PCR-based analysis techniques for many applications. DNA fingerprinting technique focuses on microsatellite regions, which are short segments of two- to six-nucleotide repeat units. The number of times a particular sequence has repeated within a given microsatellite region varies among individuals. The length of the sequence repeat can be determined by using PCR-based technique where primers are designed in such a way that they flank the microsatellite region. High-resolution gel electrophoresis technique is usually used to determine the size of the resulting PCR products. One of the RFLP variants, "terminal restriction fragment length polymorphism (TRFLP)," includes both PCR and RFLP analysis and has applications in characterization of bacteria and related communities. In this technique, fluorescent-labeled primers are used for DNA amplification, followed by restriction enzyme-mediated digestion of amplified DNA and RFLP analysis.

2.2.3 Reverse Dot Blot

In reverse dot-blot analysis technique, allele-specific oligonucleotide probes are immobilized on strips of nylon membrane. Human leukocyte antigen (HLA)-DQα1 gene found on chromosome-6 was the first commercially available PCR-based forensic DNA fingerprinting kit. This technique is helpful in detecting the sequence variation (HLA)-DQα1 gene. This kit utilizes sequence-specific oligonucleotide (SSO) probes present at specific locations on a test strip. Denatured PCR products are annealed to the SSO probes, followed by calorimetric analysis. The first DQα kit was able to distinguish six different alleles 1.1, 1.2, 1.3, 2, 3, and 4. The original DQα strip contained a total of nine SSO probes. Later, Perkin-Elmer/Roche released a slightly more sophisticated kit DQα1, which contained 11 SSO probes. HLA-DQ-α reverse dot-blot strips were admired by the forensic scientists. They were very handy and useful in rapid detection. One of the major limitations of these strips is lower discriminating power compared to RFLP technique mostly in cases of mixed samples.

2.2.4 AmpFLP: Amplified Fragment Length Polymorphism

During the early 1990, AmpFLP technique came into practice for DNA fingerprinting. This technique was faster, less labor-intensive than RFLP analysis, and highly reproducible. It uses restriction enzyme-digested genomic DNA as template for PCR amplification. It works on the principle of variable number tandem repeat

(VNTR) polymorphisms to distinguish various alleles. The amplified products are separated on a polyacrylamide gel using an allelic ladder (in contrast to a molecular weight ladder). The bands on gel are visualized using silver staining. The popular locus D1S80 is mostly used for fingerprinting. The problem associated with almost all PCR-based methods is allelic dropout (leading to misinterpretation between a heterozygote and homozygote) or other stochastic effects when highly degraded or low quantity of template DNA is used. Additionally the presence of very high number of repeats may bunch together at the top of the gel, thus complicating the analysis process. AmpFLP analysis can be highly automated which allows trouble-free construction of phylogenetic trees, based on comparing DNA of individual samples.

2.2.5 STR Analysis: Short Tandem Repeats

A short tandem repeat, also referred to as microsatellite or simple sequence repeats (SSRs), is a short DNA fragment of usually less than 100 bps and is made up of tandem repeats of 2–13 bps. In DNA, repeated regions are mostly found near the chromosomal centromere and are known as satellite DNA. STR analysis is used to evaluate specific loci on DNA from two or more samples by measuring the exact number of repeating units. During early 1990s STR markers were used as valuable tools for human identity testing. STR analysis does not use restriction enzymes to cut the DNA but rather uses probes to specific regions on the DNA followed by multiplex PCR to detect the lengths of the short tandem repeats in vitro. Sequence-specific primers are used to amplify STR loci, employing the PCR. PCR products are analyzed by high-resolution gel electrophoresis which can detect small differences in the lengths of DNA fragments. STR utilizes highly polymorphic regions of DNA with short repeated sequences. However, these STR markers are shared by around 5–20% of individuals. Hence, it is required to take the highest polymorphic STR markers and multiple STR loci, simultaneously in account. In October 1998, the USA launched the National DNA Index System (NDIS) which contained 13 core STR loci which were TH01, vWA, FGA, D8S1179, D18S51, D21S11, CSF1PO, TPOX, D3S1358, D5S818, D7S820, D13S317, and D16S539. Autosomal STRs and Y-chromosome STR markers came in use after commercialization of the core loci. Different STR-based DNA-profiling systems are used by different countries, but tetranucleotide repeat STR system has become more popular than di- or trinucleotide repeats. In human genome, penta- and hexanucleotide repeats are less common, however currently being used by some of the laboratories. The beauty of this STR technique is that small product size of STR markers is usually compatible with degraded DNA, and hence a small amount of DNA template is required for the test. In a single test, high discriminatory power is possible when multiplex amplification of STR markers with fluorescence detection is done. Capability for national

and international sharing of criminal DNA profiles has enhanced due to the uniform set of core STR loci. These characteristics of STR markers make them the most popular in the field of forensics.

2.2.6 CODIS STR Loci

Since 1996, the Combined DNA Index System (CODIS) is the national DNA database of the USA, created and maintained by the Federal Bureau of Investigation (FBI). It combines forensic science and computer technology into a tool which enables federal, state, and local forensic laboratories to exchange and compare DNA profiles electronically. In that way the CODIS STR loci help in linking serial violent crimes to each other and to identify the offenders. The DNA Identification Act of 1994 dignified the FBI's authority to establish a National DNA Index System (NDIS) for law enforcement purposes. Across the USA, over 190 public law enforcement laboratories participate in NDIS. More than 90 forensic laboratories in more than 50 countries use the CODIS software for their own database initiatives. Thirteen core STR loci were selected to be the basis of the future CODIS national database on November13–14, 1997, in STR Project meeting. Later, additional seven loci (D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, and D22S1045) were added to the CODIS core in 2015, which came to effect from January 1, 2017. Different countries employ different STR-based DNA-profiling systems, e.g., 20 core CODIS universal in North America, 17 loci system in the UK (DNA-17), and 18 loci system in Australia. CODIS STR loci DNA-profiling system is based on multiplex reactions which result in testing many STR regions at the same time.

2.3 Lineage Markers in Forensic Analysis

2.3.1 Y-Chromosome Analysis

Lineage markers have important applications in mass screening approaches of forensic science. The autosomal STR markers are commonly chosen due to high discrimination potential; however, they are not ideal for rape cases where more than one male accused is involved. Differential organic DNA extraction method used for isolating DNA from rape case evidences has the ability to separate male DNA from the female DNA followed by autosomal STR analysis for individual identification. In rape cases, where multiple males are involved, Y-STR markers have power of discrimination between male lineages. A Y-STR marker utilizes STRs solely present on paternally inherited Y chromosome. Y-STR examination is

very useful in the identification of paternally associated male individual. One of the major disadvantages associated with Y-STR is that it fails to discriminate between male relatives as the rate of genetic recombination in this region is negligible. So, unless any mutation occurs, it is difficult to differentiate male members within a family, using Y-STR markers.

2.3.2 Mitochondrial Analysis: Mitochondrial DNA

Mitochondria have their own genome. The human mitochondrial genome consists of a double-stranded circular DNA of 16.6 kb. Mitochondrial DNA (mtDNA) is also used as the lineage marker and for forensic DNA analysis [4]. During fertilization, mtDNA is contributed by the egg. Therefore, both male and female inherit mtDNA from their mother. In forensic DNA analysis, mtDNA can play a significant role and can be used as matri-lineage marker. Forensic scientists amplify, sequence, and compare the hypervariable segment 1 and 2 of the mtDNA [10], and the single nucleotide alteration is taken into account. MtDNA is useful in solving kinship cases, such as those of missing people when a maternally linked relative can be found. Hence, females have only one type of mtDNA (F-type), whereas males have both F-type and M-type mtDNA in their somatic and germinal cells, respectively. Within a family, all the females share the same mtDNA unless rare mutations occur. Moreover, they have many advantages over other DNA samples. MtDNA is present in high copy number per human cell, so it is required in less quantity [1]. In cases of old remains like bones, teeth, and hairs, mtDNA is useful for collecting information, as it survives during postmortem degradation process [8, 14]. Like Y-STR marker, mtDNA markers also do not have individual specificity, resulting in a lack of discrimination capacity.

2.3.3 X-Chromosome Analysis

A large number of short tandem repeat markers are present in human X chromosome, which are used in forensic genetic testing [11]. X-chromosomal short tandem repeat (X-STR) loci are mostly employed in forensic practice in complex kinship cases. Investigator Argus X-12, the commercially available kit for X-STR, is used to examine 12 STR markers belonging to four linkage groups of X chromosome in a single PCR. X-STR markers have a particular inheritance pattern: women contain two X chromosomes transferred by both parents, and men contain single X chromosome received from the mother. In paternity cases with limited reference samples, where half-sisters and grandmothers are examined, X-chromosomal loci analysis may be beneficial. The daughters carry the whole paternal X chromosome, whereas the son carries the mother's [7]. X-STR analysis provides easy and informative haplotyping kinship analysis. X-STRs are considered more complex genetic

markers in forensic analysis because they are tightly linked. Thus, X-STR profiling is a powerful tool for solving the cases with limitation of samples from the close relatives.

2.4 Past, Present, and Future of DNA Fingerprinting

DNA profiling has brought the scientific revolution in the field of forensics, but it was not immediate. It started in 1984 when Alex Jeffrey first used RFLPs for human identification [9]. In 1984, DNA typing was first used to solve an immigration case. The immigration authorities were not satisfied that the woman claiming a Ghanaian boy as her son was in fact his mother. Serum proteins and erythrocyte antigens and enzyme analysis showed that the alleged mother and son were related but could not determine whether the woman was the boy's mother or aunt. In this complicated matter, father was not available for analysis. In the beginning, DNA fingerprinting was highly contentious and not always accepted by the courts. The arguments and uncertainties related to the techniques and data interpretation anomalies related to DNA-profiling analysis led to a study of National Research Council (NRC 1992) organized by the National Academy of Sciences which has been published subsequently. The FBI began using DNA fingerprinting in 1998. John M. Butler, in 2015, divided development of DNA fingerprinting in forensics in the last three decades into different phases which is summarized in Table 2.1 [3]. First is "exploration phase," where different methodologies like RFLP and some of the basic PCR assays were attempted. Although the convenience of DQα and simultaneous analysis of multiple markers led to their extensive use during 1985–1995, VNTRs dominated the prospect. The first drawback of VNTRs was that they required somewhat larger amount of DNA. In the later stage, PCR technique solved the issue. Sometimes, there was problem in amplification of large size of VNTRs by PCR, so, from 1995 onward, the shorter STRs began dominating. Third is "growth phase," where DNA databases were rapidly increasing. Fourth is "sophistication phase" where escalating instrument sets provided new capabilities for DNA analysis and use. The next-generation sequencing (NGS) also called as massively parallel sequencing results in simultaneous collection of information from numerous STR markers [12]. To solve forensic cases, the UK Forensic Science Service (FSS) started using STR markers with quadruplex consisting of TH01, vWA, FES/FPS, and F13A1 and later second-generation multiplex (SGM) consisting of six STR loci (TH01, vWA, FGA, D8S1179, D18S51, and D21S11) and the sex-typing marker amelogenin [13]. The first national DNA database was launched by the UK in 1995 [15], whereas National DNA Index System (NDIS) was developed by the USA in 1998 with 13 core STR loci [2]. Then, the autosomal STRs and lineage-specific STRs came into force. Subsequently, the replacement of slab-gel electrophoresis followed by silver staining fluorescence detection technique with capillary electrophoresis and laser-induced fluorescence led to the possibility of the highly sensitive, automated, and high-throughput analysis.

Table 2.1 Different STR markers used in forensic DNA fingerprinting: utility, limitations, and present and future

Genetic markers	Utility	Limitations	Present	Future
Autosomal STR	High discrimination capacity	Not suitable in gang rape cases.	Core loci used to solve forensic cases and generate DNA profile databases with CE system	More core autosomal STR loci can be discovered by using NGS
Y-chromosome STRs	Useful in rape cases where multiple males are involved	No genetic recombination so unable to discriminate between male relatives Does not have individual specificity	12–27 Y-STR loci with haplotype frequencies searched in population databases (e.g., YHRD.org); for case study	More core Y-STR loci can be discovered by using NGS. Rapidly mutating Y-STRs markers can be discovered which can discriminate close male relatives
X-chromosome STRs	Useful in paternity cases with sample limitation from close relatives and kinship cases	Highly linked	Population data collected for 12 bp loci but only used occasionally in kinship cases	X-STRs markers can regularly use to solve challenging kinship cases along parallel with autosomal STRs
Mitochondrial DNA	Beneficial in cases of old remains and also in kinship cases	Does not have individual specificity	Sequencing of hypervariable control regions followed by estimation of haplotype frequencies through population database searches (e.g., EMPOP.org)	Larger population databases can be generated

In recent years, more data was recovered from biological samples due to improved sensitivity and technological advancements such as automated DNA-profiling system and the use of diverse genetic markers. Different genetic markers used in DNA fingerprinting for solving criminal cases, their utility, limitations, and present and future prospect have been summarized in Table 2.1. The rapid DNA analysis instruments integrate different steps in one single platform such as DNA extraction, rapid PCR amplification of STR loci, their separation, detection, sizing, and genotyping. In future, DNA testing protocols and related instruments are expected to improve manifold. In the present scenario, DNA recovery or DNA profiling from challenging, poor-quality samples is a huge problem. Different PCR inhibitors present in the respective samples also create an immense problem to produce an informative DNA profile. In future dealing with PCR inhibitors and poor-quality DNA, samples can improve. For data recovery from degraded DNA templates, miniSTRs can be amplified

from smaller or degraded DNA templates [5, 6]. In future, additional small size miniSTR markers can be discovered to be used by NGS for enhanced detection limit of degraded samples.

Acknowledgment The authors are thankful to The Director, State Forensic Science Laboratory, Ranchi, Jharkhand, for the support.

References

1. Bender K, Schneider PM, Rittner C (2000) Application of mtDNA sequence analysis in forensic casework for the identification of human remains. Forensic Sci Int 113(1–3):103–107
2. Budowle B, Moretti TR, Niezgoda SJ, Brown BL (1998) CODIS and PCR-based short tandem repeat loci: law enforcement tools. In Proceedings from the second European Symposium on Human Identification. http://www.promega.com/~/media/files/resources/con
3. Butler JM (2015) The future of forensic DNA analysis. Phil Trans R Soc B 370:20140252
4. Butler JM, McCord BR, Jung JM, Wilson MR, Budowle B, Allen RO (1994) Quantitation of PCR products by capillary electrophoresis using laser fluorescence. J Chromatogr B 658:271–280
5. Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 48:1054–1064
6. Coble MD, Butler JM (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. J Forensic Sci 50:43–53
7. Edelmann J, Hering S, Kuhlisch E, Szibor R (2003) Validation of the STR DXS7424 and the linkage situation on the X-chromosome. Forensic Sci Int 125:217–222
8. Ginther C, Issel-Tarver L, King MC (1992) Identifying individuals by sequencing mitochondrial DNA from teeth. Nat Genet 2:135–138
9. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73
10. Kline MC, Vallone PM, Redman JW, Duewer DL, Calloway CD, Butler JM (2005) Mitochondrial DNA typing screens with control region and coding region SNPs. J Forensic Sci 50:377–385
11. Nothnagel, M., Szibor, R., Vollrath, O., Augustin, C., Edelmann, J., Geppert, M. et al. Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome. Forensic Sci Int Genet 2012; doi: https://doi.org/10.1016/j.fsigen.2012.02.015
12. Seo SB, King JL, Warshauer DH, Davis CP, Ge J, Budowle B (2013) Single nucleotide polymorphism typing with massively parallel sequencing for human identification. Int J Legal Med 127:1079–1086
13. Sparkes R et al (1996) The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies. Int J Legal Med 109:186–194
14. Sullivan KM, Hopgood R, Gill P (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. Int J Legal Med 105:83–86
15. Werrett DJ (1997) The national DNA database. Forensic Sci Int 88:33–42

Chapter 3
Techniques Involved in DNA Fingerprinting: Isolation, Quantification, PCR, Genotyping, and Analysis

Braja Kishore Mohapatra

Abstract The primary objective of forensic DNA fingerprinting is to establish the relatedness between the questioned samples and reference samples with absolute certainty beyond any reasonable doubt. This sophisticated and robust technique however requires a series of methodologies and procedures which are to be followed by the forensic scientists to obtain an accurate result. The procedures of DNA fingerprinting have accelerated through a series of major stages of technological advancement. It began with an initial methodology called "restriction fragment length polymorphism (RFLP)" crossing the stages of "minisatellites" or VNTR analysis and currently advancing through the stages of "microsatellite" or STRs and SNPs. Similarly the instrumentation processes involved also changed from horizontal gel to vertical gel system to automated analyzers consisting of sensitive microcapillary arrays with laser-based detection system. Quality softwares are now available for data analysis and generation of user-friendly reports. Even though the techniques involved in DNA fingerprinting vary greatly from laboratory to laboratory throughout the world, the basic concept remains the same and is represented here.

Keywords DNA fingerprinting · STRs · DNA isolation · Multiplex PCR · Genotyping · Capillary electrophoresis

3.1 Introduction

During the course of time, the technology of DNA fingerprinting has evolved as one of the most reliable techniques in the field of human identification. The high degree of discriminating power of the technique has made it the most dependable tool for crime investigation purposes. Its accuracy and acceptability in the court of law have changed the scenario of conviction rates throughout the world, and truly speaking

B. K. Mohapatra (✉)
Department of Biology and DNA Fingerprinting Unit, Central Forensic Science Laboratory (CBI), New Delhi, Delhi, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_3

crime investigation has become more sophisticated and easier with the advancement of the technique of DNA fingerprinting and its implementation in a wide range of crime cases [26]. The stability of the DNA molecule and reproducibility from trace evidence samples have marked a high degree of success rate in crime investigations. However, for a forensic scientist, the challenge lies in the employment of accurate procedures in a defined manner during the process of DNA fingerprinting to get a final output, i.e., a complete DNA profile from the evidence samples for comparison with the reference samples [3]. Biological materials contained in the forensic evidence samples are normally degraded, impure, and contaminated and contain various inhibitors. In such situations isolation of sufficient quantity of high molecular weight DNA, its purification and quantification for multiplex PCR amplification, and subsequent genotyping have become important tasks.

3.2 Isolation of DNA from Biological Samples

The first and most important step of the technology of DNA profiling is effective extraction of purified DNA from the available biological samples which can be used for all downstream reactions. The sensitive enzyme-based polymerase chain reaction (PCR) involved for selective amplification of the desired DNA markers requires input of optimized quantity of high molecular weight purified DNA samples devoid of impurities and inhibitors [49]. A forensic scientist often needs to analyze the samples which are degraded, contaminated, and traced in nature. Extracting good quality DNA from such challenged samples can be treated as success, if a complete DNA profile is generated from such samples for analysis and comparison. Currently various methods are available for isolation of DNA from biological samples. However, the choice of the method depends upon the nature and quantity of available biological samples. The commonly used DNA isolation methods include organic isolation of DNA, nonorganic isolation of DNA, column-based isolation of DNA, and magnetic bead-based isolation of DNA. However, in all the instances, the principle involved can be categorized into four basic steps, i.e., cell lysis via disruption of cellular membranes, inhibition of nucleases, removal of denatured proteins and other cell debris, and purification of DNA [19].

3.2.1 Organic Isolation of DNA from Biological Cells

Organic isolation of DNA from any type of biological samples is a common method that uses organic chemicals and buffers to extract genomic DNA as well as mitochondrial DNA. This is the most widely used procedure followed in many forensic science laboratories. The method is highly cost-effective and gives maximum yield of DNA for downstream reactions. The procedure involves the lysis of biological cells using optimized organic buffers followed by enzymatic hydrolysis of cellular proteins and subsequent purification of DNA via a biphasic liquid separation system [28].

3.2.1.1 Lysis of Membrane, Denaturation, and Hydrolysis of Proteins

The organic buffer used in almost all forensic science laboratories for lysis of biological cell membrane is commonly known as "extraction buffer" or forensic DNA extraction buffer. The basic components of the stain extraction buffer include Tris, NaCl, SDS, EDTA, and water. These buffers are slightly alkaline and normally contain Tris base or Tris-HCl (pH 7.5–8.0). Further dithiothreitol (DTT) and proteinase K are added to the buffer for further enhancement of extraction process. The DNA remains protected from unwanted degradation and denaturation in this process by a magnesium chelator, i.e., EDTA. EDTA inhibits nucleases from denaturing the DNA. Tris interferes with lipopolysaccharides present on the cell membrane and helps it to make it permeable. This effect is improved with the addition of EDTA, and NaCl adds to the positive ion concentration of the aqueous medium [39].

The cell, nucleus, and majority of cell organelles of living organisms are covered with a protective cell membrane which is mainly made of a lipid bilayer structure. The lipid bilayer structure acts as a selective barrier which keeps ions, proteins, and other molecules where they are needed and prevents them from diffusing into areas where they are not required (Fig. 3.1). Such bilayers are usually made of up phospholipids which have a hydrophilic head and two hydrophobic tails. When phospholipids are exposed to water, they arrange themselves into a two-layered structure (a bilayer) with all of their tails pointing toward the center of the structure. The center of this bilayer structure contains almost no water and also excludes molecules like sugars or salts that dissolve in water.

Detergents like SDS or sarkosyl, which are a part of the extraction buffer, lyse the cell membranes, separate histone proteins from DNA, denature histone proteins, and disrupt secondary and tertiary structures of proteins, which reduce their solubility in aqueous solution. Detergent sarkosyl is usually used if a lysis process is to be conducted at low-temperature conditions, because sodium dodecyl sulfate (SDS) precipitates out of the solution at a condition less than room temperatures. SDS is the detergent which is normally used when lysis procedures are conducted at room temperatures. Detergents have hydrophobic-associating properties by virtue of their nonpolar tail groups (Fig. 3.2). These detergents are water-soluble and allow the dispersion (miscibility) of water-insoluble, hydrophobic compounds into aqueous media which includes removal and solubilization of membrane proteins [5].

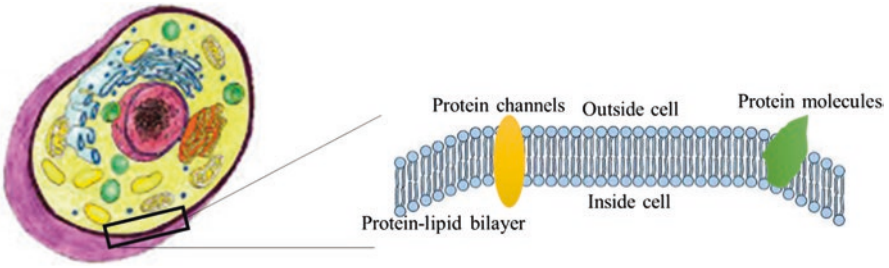

Fig. 3.1 Structure of typical eukaryotic cell and cell membrane

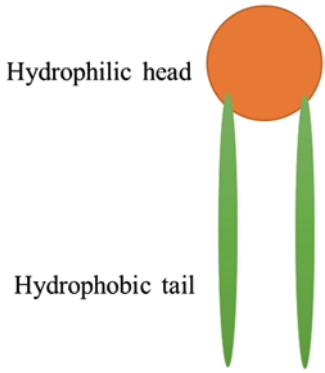

Fig. 3.2 Structure of a typical phospholipid molecule with hydrophilic head and hydrophobic tail

Enzyme proteinase K is added to the medium to hydrolyze histone proteins and is suitable for the process as it remains active over a wide range of pH (4–12.5), it also remains active in the presence of detergents, and it is not inhibited by EDTA. Enzyme proteinase K is produced by the fungus *Tritirachium album* Limber. It is an endolytic enzyme that breaks peptide bonds of protein structure [40]. Enzyme proteinase K in the extraction buffer also reduces nuclease activities and aids to lysis of cell membrane to free nuclear DNA. Additionally, dithiothreitol (DTT) reduces disulfides to dithiols, allowing the release of the DNA from its protective proteins and subsequent degradation of the proteins by proteinase K. DTT is used as an essential component for sperm cell lysis as this cell membrane contains a high concentration of disulfides.

3.2.1.2 Removal of Denatured Proteins and Other Cell Ingredients

Normally a biphasic (organic and aqueous phase) separation system is employed for removal of denatured proteins and other cell particles from the extraction medium. The organic phase (nonpolar) contains phenol, chloroform, and isoamyl alcohol in the ratio of 25:24:1, and the inorganic or aqueous phase contains water.

Phenol is a member of a group of organic compounds which contains a hydroxyl group fixed to an unsaturated carbon in a benzene ring (Fig. 3.3). Phenols are not true alcohols; they are more acidic than alcohols but less so than carboxylic acids. Polysaccharides and proteins are soluble in phenol, allowing for their separation from DNA [50]. *Chloroform* (CHCl3), or trichloromethane, is a colorless liquid that is slightly water-soluble and miscible with organic solvents such as phenol. It is more dense than water or buffer (in which DNA is soluble) and less dense than phenol. It improves the clarity of interface between the organic and aqueous layers. Further chloroform also solubilizes lipids [41]. During the extraction procedure, cellular debris can be observed at the interface. *Isoamyl alcohol*, or 3-methyl-1-butanol, is often included in genomic extraction procedures to prevent foaming and making it easier to visualize the interface between the organic and aqueous phases [21].

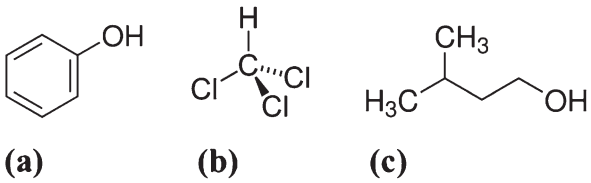

Fig. 3.3 Chemical structure of the common reagents used for DNA isolation: (**a**) phenol, (**b**) chloroform, and (**c**) isoamyl alcohol

The phenol-chloroform-isoamyl alcohol in the ratio 25:24:1 is mixed with the extract and is subjected to high-speed centrifugation to separate the DNA molecules from other cell ingredients. DNA being negatively charged (polar) due to its highly charged phosphate backbone gets solubilize in water which is also highly polar, "like dissolves like" (dielectric constant of water is 80 at 25 °C) [25]. The relatively weak electric force of the positive ions present in the solution prevents them from forming stable ionic bonds with phosphates and precipitating DNA out of the solution. High-speed centrifugation makes three distinct separating levels in the tube containing the extract. The lower dense phase is the phenol-containing denatured proteins that are subsequently hydrolyzed with proteinase K. The intermediate separating layer contains the lipid dissolved in the chloroform, and the DNA remains in the top aqueous phase.

3.2.1.3 Precipitation of DNA

Absolute ethanol is very much less polar in comparison with water; its dielectric constant is 24.3 (at 25 °C) which means that adding ethanol to aqueous phase disrupts screening of charges by water. If sufficient ethanol is added, electrical attraction between phosphate groups and any positive ions present in the solution becomes strong enough to form stable ionic bond results in precipitation of DNA. This happens when ethanol makes around above 64% (absolute alcohol) of the solution. As the mechanism suggests, the solution has to contain positive ions for precipitation to occur; usually Na^+, NH_4^+, or Li^+ plays this role [18]. Low temperature allows maximum recovery of DNA as the polarity further reduces at ultralow temperature and also prevents DNA from further degradation. Isopropanol can also be used instead of ethanol. However even though isopropanol has better efficiency for precipitation of DNA, because of its less volatility in comparison with ethanol, it needs more time to dry and hence not preferably used. Further in order to remove the leftover salts, 70% ethanol is added to the pellet. After complete mixing it is again centrifuged, and the pellet containing the extracted DNA is allowed to dry. Finally, the pellet is resuspended in water or other desired buffer for further downstream reactions.

A typical organic DNA isolation procedure from fabrics containing biological cells can be represented in the following steps:

1. Put the stain containing the biological cells in a fresh centrifuge tube.
2. To it add
   400 µl extraction buffer + DTT (0.006 gm/1 ml of SEB) + 10 µl proteinase K.
   Vortex for 10 s. Short spin.
   Incubate at 56 °C in water bath (minimum for 6 h)

3. Take the extract portion in a fresh Eppendorf tube. Add 400 µl of phenol-chloroform-isoamyl alcohol (ratio 25:24:1); vortex for 30 s; and centrifuge at 15,000 rpm for 6 min.
4. Transfer the aqueous layer (topmost layer) and place in a fresh 1.5 ml Eppendorf tube.
5. To the aqueous layer, add 1.0 ml of cold, absolute EtOH. Mix the tube by inversion. Keep the tube at −20 °C for 15 min. Centrifuge at 10,000 rpm for 6 min.
6. Discard the supernatant by decantation.
7. Wash the pellet with 70% EtOH. Centrifuge at 10,000 rpm for 6 min.
8. Discard the supernatant by decantation.
9. Air-dry.
10. Dissolve the pellet in TE buffer or MQ water.
11. DNA ready for quantification.

3.2.2 Nonorganic Isolation of DNA

Keeping in view of the toxic nature of phenol, several laboratories have adopted nonorganic extraction methods like Chelex extraction method and FTA card-based DNA extraction methods. In Chelex extraction method, 5% suspension of Chelex resin is prepared and added to the biological cells or the materials containing biological cells. Chelex has a very high affinity for polyvalent metal ions such as Mg^{2+} [53]. It chelates the polyvalent metal ions and effectively removes them from the solution.

A typical Chelex DNA isolation procedure from fabrics containing biological cells can be represented in the following steps:

1. The biological cells or materials containing biological cells are added to 500 microliter of Tris-EDTA buffer and incubated for 15 min.
2. The extract is subjected to high-speed centrifugation to get the cellular materials as pellet.
3. The supernatant is removed.
4. The pellet is resuspended in 5% Chelex.
5. Incubate the solution for 2 h at 56 °C.
6. Further centrifuge the extract and the supernatant that contains the DNA for downstream reactions.

FTA card is a specially designed card with a complex matrix system, which retains the DNA molecules of biological samples. It can be kept at normal temperature condition for indefinite period and transported to the laboratory [1]. Biological samples are put on the encircled areas over the card. When absorbed, the matrix system ruptures the cell walls, nuclear walls, proteins, lipids, etc. and retains the DNA molecules on its matrix network (Fig. 3.4). The DNA molecules can be preserved for much longer durations even for several years as the specially designed

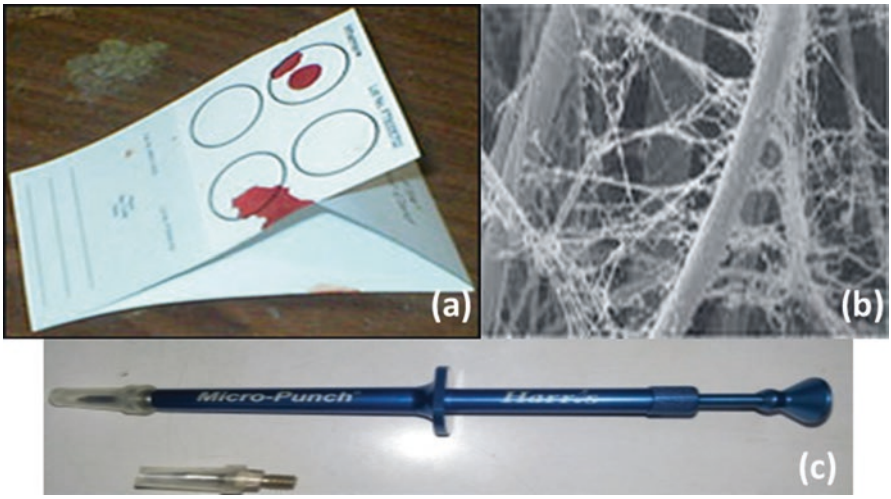

Fig. 3.4 FTA card and its accessories, (**a**) FTA card with blood stains, (**b**) typical matrix of FTA card, and (**c**) micro-puncher to cut blood stains from FTA card

matrix also prevents the infectious microbes tending to destroy the DNA on it. The FTA wash buffer is used to wash or remove the cellular debris and other inhibitors from the micro-punches of approximately 0.3 mm diameter. The cleaned micro-punches of the FTA card are washed with TE buffer, dried, and taken for further analysis.

3.2.3 Column-Based Isolation of DNA

Specially designed spin columns are available which contain a positively charged material commonly silica and other salt ions at the bottom [31]. Lysis buffer along with the cells is poured into these spin columns. After lysis of the cells and nucleus, the DNA binds to the positively charged materials. During initial centrifugation, the cell debris and other ingredients/inhibitors are passed/drained out because of their non-affinity with the positively charged material. Further the column is washed with wash buffer followed by elution of DNA with buffer or molecular biology grade water.

3.2.4 Magnetic Bead-Based Isolation

This is the most advanced and quicker method of DNA isolation from biological cells. Many of the forensic science laboratories use automated DNA extraction machines which are mainly based on this technology [20]. Prefilled DNA extraction

cartridges are available which contain all the requirements like the lysis buffer, wash buffer, suspended magnetic beads, and elution buffers. After cell lysis magnetic beads are added to the lysed cells and are brought to electric field. Magnetic beads get positively charged under the influence of electricity. Additionally, because of their affinity, nucleic acid binds to the magnetic beads and can be eluted with buffer or water when electric field is removed from the magnetic beads making them neutral.

3.3 Purification/Concentration of Isolated DNA

The DNA recovered during the process of isolation can be further purified and concentrated using available ready-to-use purification kits. It is recommended that the kits used are required to be validated in-house in own setups. Purification kits generally contain spin columns having materials with DNA binding affinity at the bottom [16]. The binding affinity of DNA is enhanced with the addition of binding buffer to it, and the inhibitors are drained out by wash buffers when subjected to high-speed centrifugation. This is a common procedure for column-based purification and concentration process which involves the following steps:

1. Dissolve the air-dried DNA with 30 µl of ultrapure distilled water followed by vortex and short spin.
2. Add DNA binding buffer to the sample (double the volume of DNA sample); mix by vortexing.
3. Transfer the mixture to the spin column placed in a collection tube.
4. Centrifuge at 15,000 rpm for 6 min.
5. Discard the flow-through.
6. Add 200 µl wash buffer to the column.
7. Centrifuge at 15,000 rpm for 6 min; repeat wash step.
8. Discard the collection tube and place the column in a 1.5 ml micro-centrifuge tube, and add 25 µl of purified water; allow to stand for 15 min.
9. Centrifuge at 15,000 rpm for 6 min to elute the DNA.

3.4 Quantification of DNA

Quantification of DNA is much essential as an optimized quantity of DNA is required for the process of amplification based on polymerase chain reaction (PCR) technology [35]. Adding the accurate quantity of DNA for amplification process will yield best quality results as a higher quantity of input of DNA results in pre-utilization of available primers and less input of DNA results in nonavailability of DNA templates for primer binding. Generally three methods are employed for estimation of quantity of DNA.

3.4.1 Yield Gel Method

1% agarose gel is prepared by dissolving the agarose in running buffer (preferably 1X Tris-acetic acid-EDTA buffer). Intercalating dye such as ethidium bromide (EtBr) is added to the gel. Further the extracted DNA along with an indicator dye like bromophenol blue is poured into the wells of the agarose gel, and the gel is subjected to electrophoresis in submerged condition. The high concentration of sugar (mostly sucrose) in the loading dye allows the DNA to sink to the bottom of the well. Under the influence of electricity, the negatively charged DNA molecules migrate toward the positive side via the intracellular spaces of the agarose gel. During the course of migration, the DNA molecule forms a complex with the intercalating dye. The degree of migration can be viewed by the movement of the indicator dye. After migration of a certain distance, the gel is viewed under ultraviolet transilluminator or gel documentation system. The DNA-intercalating dye complex emits fluorescence when exposed to ultraviolet light. The fluorescence is viewed in the shape of fluorescence bands. The thickness of the band is co-related with the quantity of DNA present in the medium. The concentration of DNA can be calculated on comparison basis with known quantity of DNA sample running in the same gel in another well. Further the pattern of migration also indicates about the quality of the DNA [52].

3.4.2 Absorption Spectrophotometry

Absorption of light by DNA molecule is maximum at 260 nm. This character of DNA is used to quantitate DNA present in a solution. The extracted DNA is scanned within a range of 220–300 nm either by nano-drop techniques or by using cuvette and conventional spectrophotometers. The optical density (OD) of a medium with a 1 cm path length, containing 50 μg/ml of double-stranded DNA or 40 μg/ml of single-stranded DNA, is 1.00 at a wavelength of 260 nm [12]. The quality or purity of the sample can be determined by comparing the measurements at 260 and at 280 nm.

3.4.3 Quantitation via Real-Time PCR Method

The sophisticated and most advanced method of DNA quantitation is based on the technique of real-time PCR [11]. This is the most accurate, precise, and efficient method currently used for human DNA quantitation. The real-time PCR instrument consists of a thermal cycler integrated with a digital fluorometer. It is a process that detects and measures the accumulation of fluorescent dyes as the reaction progresses. The accumulation/emission of fluorescence is achieved by making probes containing fluorescent dyes bind to the template DNA strands [22]. The initial quantity of DNA in the sample is detected by monitoring the exponential growth phase of

the reaction and measuring the cycle number at which the fluorescent intensity of the sample overcomes the background noise or threshold. This cycle number is directly proportional to the quantity of DNA in the reaction. Analysis of the quantity of DNA in the sample is performed using software that compares the unknowns with the best-fit regression line constructed from the standards, i.e., phase one, exponential/geometric; phase two, linear phase; and phase three, plateau region [47].

During the exponential phase of the PCR process, the reaction results in a theoretical doubling of amplicons with each cycle. The doubling is close to 100%, yielding a consistent relationship of input DNA to product. At the beginning of the exponential phase, the baseline is determined by measuring the background fluorescence signal (noise). This baseline establishes the threshold. During the amplification process of a sample, the point (in terms of amplification cycles) in which the level of fluorescence exceeds the threshold is referred to as the cycle threshold (C_T). The C_T value is lower for a sample with a higher initial concentration and is higher for a lower concentration sample [47]. Other parameters calculated during the process of real-time PCR have been described in Fig. 3.5. The standard curve is a regression line that is derived from the C_T values of the standards plotted against the log of the concentration of the standards. The concentration of each sample is determined by comparing its C_T value against the standard curve. During the linear phase, the consumption of one or more PCR reagents during the amplification process will impede the efficiency causing the onset of the linear phase. Product amplification slows down yielding an inconsistent ratio of input DNA to product. Therefore, the linear phase is not commonly used in the data analysis process [47]. During plateau region the final amplification phase signifies the depletion of critical

Fig. 3.5 Parameters calculated during the process of quantitative real-time PCR, i.e., (**a**) amplification plot, (**b**) standard curve, (**c**) multicomponent plot, and (**d**) raw data plot

reagents and is known as the plateau region. During this phase, the amplification process ceases [47].

Two principles are generally used in a real-time PCR system for quantification of DNA, i.e., SYBR Green and TaqMan chemistry. The principle behind SYBR Green chemistry is the gradual increase in fluorescence from SYBR Green dye with the progress in PCR amplification process. As the SYBR Green dye binds to double-stranded amplicons, it undergoes a conformational change and emits fluorescence at a greater intensity [43]. Additionally, the TaqMan probe principle relies on the 5′–3′ nuclease activity of Taq polymerase to cleave a dual-labeled probe during hybridization to the complementary target sequence and fluorophore-based detection [6].

3.5 Amplification of Desired Loci of DNA Templates

Kary Mullis and his co-workers developed the technique of polymerase chain reaction (PCR) in the year of 1985 which revolutionized the field of forensic DNA profiling and human genetics research. The technique involves the enzymatic procedure, whereby specific desired regions of the DNA templates are replicated repeatedly corresponding to each cycle under the influence of heating and cooling of the sample supplemented with specifically designed primers and free nucleotides [51]. The process leads to the evolution of the PCR machine, commonly known as thermocycler. In a typical thermocycling, three different temperature levels are rotated in each cycle. The key steps of a PCR process involve the denaturation, annealing and extension. In denaturation step the double stranded DNA denature to form the single stranded form. Whereas, during annealing, the primers bind to the specific DNA sequence at specific temperature. Finally, during extension, DNA polymerase extends the primers by coping the complimentary target region using the deoxynucleotide triphosphate building blocks (dNTPs). Amplification of the template DNA is achieved by repeating the three-step process through 25–30 cycles in a thermal cycler generating a PCR product called as an amplicon. The result is an exponential accumulation of the specific target fragment, approximately 2^n, where "n" is the number of cycles of amplification performed. However, the efficiency of the process reduces at higher cycle numbers as after 30 cycles, approximately a billion copies of the target DNA template are generated [36].

During initial stages of development, only single locus of the DNA template was amplified. However, with the advancement of technology, more than one region (loci) of the template DNA are copied simultaneously by adding more than one primer set to the reaction known as multiplexing [15]. The targeted regions of the DNA molecule mostly contain tandem repeats. For human identification purpose, two types of tandem repeats are used, i.e., minisatellites commonly known as variable number tandem repeats (VNTR) and microsatellites commonly known as short tandem repeats (STR). The basic difference between minisatellites and microsatellite markers lies in the base-pair length within each repeat unit (Fig. 3.6). The number of core repeat sequence of VNTRs ranges in size from 6 to 100 bp. However, keeping in view of the resolution and amplification efficiency, most of the laboratories currently use the PCR-based amplification method which is commonly known as STR technique [46].

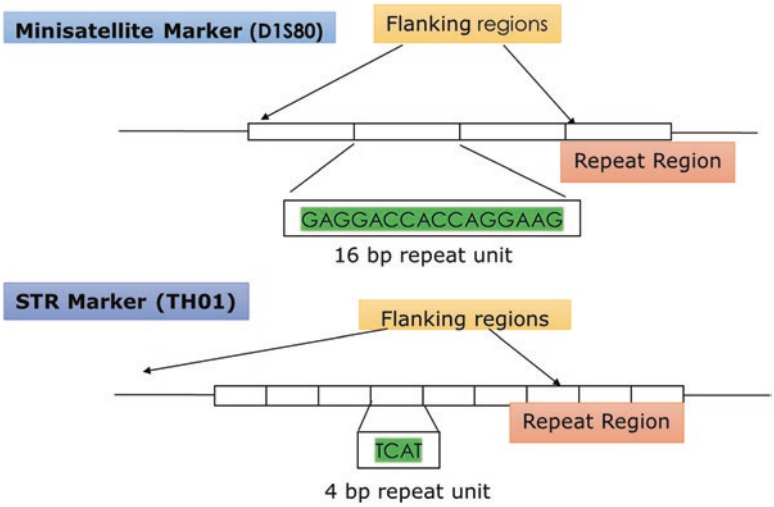

Fig. 3.6 Fundamental differences between minisatellites and microsatellites

3.5.1 PCR Components

A PCR reaction mix is prepared by adding several individual components as well as extracted DNA templates to achieve the desired volume and concentration of each of the components. Commercial kits with pre-mixed components may also be used for PCR which have simplified the utility of PCR technology in DNA fingerprinting laboratories [55]. The major components of PCR have been discussed below.

3.5.1.1 Thermostable DNA Polymerase

A wide choice of enzymes is now available that vary in the fidelity, efficiency, and ability to synthesize large DNA products. For routine PCR *Taq* polymerase, the thermostable DNA polymerase extracted originally from thermophilic bacterium *Thermus aquaticus* is widely used [4]. DNA polymerase catalyzes template-directed synthesis of DNA using dNTPs catalyzed by Mg^{2+} ion and initiated by a primer having a free 3′ hydroxyl end. Gradually suitable modifications have been carried out, and modified Taq DNA polymerase with the requirement of a thermal activation has been developed. Such enzymes the AmpliTaq Gold have greatly benefited the specificity of PCR amplifications [30].

3.5.1.2 A Pair of Synthetic Oligonucleotides

A pair of synthesized oligonucleotide sequences known as primers are of utmost importance in a PCR process as their specificity determines the amplification of the desired products. In this regard, yield of PCR is highly dependent on the annealing

Table 3.1 General guidelines for suitable PCR primers

Parameters	Optimal values
Length (in bases)	18–30
Melting temperature (in °C)	55–72
GC content (in %)	40–60
No. of hairpin structure	≤ 3 contiguous bases
Primer dimer	≤ 3 contiguous bases (at 3′ ends)
Distance between two primers	<2000 bases

characteristics of the primers [48]. The primers should possess the following characteristic features for proper amplification of the target DNA sequences in PCR (Table 3.1).

3.5.1.3 Deoxynucleotide Triphosphates (dNTPS)

An equimolar concentration of all the dNTPs, i.e., dATP, dTTP, dCTP, and dGTP mostly at the concentration of 200–250 μM, is required for the smooth functioning of DNA polymerase during a PCR process [35]. Additionally, 1.5 mM $MgCl_2$ is used in each PCR reaction with a volume of 50 μl for the amplification of ~6–6.5 μg of template DNA. These much of the components are sufficient to carry out multiplex reaction using eight or more primer pairs simultaneously. High dNTP concentration (>4 mM) may also act reversely which might be due to sequestering effect of Mg^{2+}. However, studies have shown a satisfactory result of multiplex PCR with the use of dNTP concentrations as low as 20 μM–0.5–1.0 pM for a target of ~1 kb in length [35].

3.5.1.4 Template DNA

For an ideal PCR functioning, a single copy of template DNA is required; however, in practical cases, several thousand copies of the template DNA are seeded to the initial reaction. From human DNA fingerprinting aspect, a template of 1.0 to 2.0 μg per μl of template DNA can be utilized for a successful PCR product generation [9].

3.5.2 Phases of Polymerase Chain Reaction

3.5.2.1 Hot Start

Initiating PCR at an elevated temperature can avoid low-temperature mispriming. In a recently developed technique called as "hot start" PCR [32], the most crucial component of DNA polymerase is introduced once the sample temperature rises above the desired annealing temperature. This technique minimizes the possibilities of mispriming and misextension events due to nonavailability of DNA polymerase during reaction setup [45].

3.5.2.2 Denaturation

Denaturation of a double-stranded DNA template is dependent on its G + C content. In this context, the denaturation temperature is directly proportional to the total G + C content of the DNA. Additionally, the incubation time for the denaturation of double-stranded DNA is dependent on the length/size of the DNA molecule [54]. In this case, if suitable temperature and incubation time are not chosen properly, only AT-rich regions may get denatured, thus generating incomplete single-stranded DNA molecules. In most commonly used PCR techniques, denaturation is carried out at 94–95 °C for about 45 sec, which is the highest temperature Taq DNA polymerase can sustain for 30 or more PCR cycles without excessive cleavage [30].

3.5.2.3 Annealing

Annealing, the binding of primers to the template DNA, is one of the most crucial steps of PCR. A very low amplified DNA can be detected with a higher annealing temperature where primers bind poorly to the template DNA. Similarly, low annealing temperature leads to non-specific annealing of primers resulting in the amplification of undesired segments of DNA. In most of the cases, the annealing temperature is used at the 3–5 °C lower temperature than the calculated melting temperature at which the oligonucleotide primers tend to dissociate from the template DNA [29].

3.5.2.4 Extension

Extension of oligonucleotide primers annealed at the specific DNA sequence is carried out mostly at the optimal temperature of thermostable DNA polymerase, i.e., 72–78 °C. In first two cycles, extension of one primer precedes beyond the sequence complementary to the binding site of the other primer. In further cycles, first the molecules are produced whose length is equal to the segment of the DNA delimited by the binding sites of the primers. Finally, from the third cycle onward, this segment of the DNA is amplified geometrically, whereas longer amplification products accumulate arithmetically. The efficiency of Taq DNA polymerase in terms of its polymerization rate is ~2000 nucleotides per minute at its optimal temperature of 72–78° [34].

3.5.2.5 Number of Cycles

The number of cycles required for amplification of a target DNA sequence by polymerase chain reaction depends on the initial number of copies of template DNA and the efficiency of primer extension and amplification [2]. Once the amplification reaches the geometric phase, the reaction proceeds until one of the components becomes limiting. However, at least 25 cycles are required to achieve acceptable levels of amplification of single-copy target sequence in mammalian DNA templates [27].

3.5.3 *Multiplex PCR*

By adding multiple primer sets to the reaction mixture, more than one region on DNA can be amplified by PCR [15]. Such simultaneous amplification of more than one region of DNA is commonly known as multiplexing or multiplex PCR. However, compatible primer sets are required for the proper functioning of a multiplex PCR. More technically, the primers should possess similar annealing temperatures and lesser regions of complementarities, i.e., formation of primer dimer and the chance of a primer to bind to different regions on template DNA can be avoided [7]. A multiplex PCR should be optimized properly for simultaneous multiple primer annealing events without interfering with each other. Mostly in a multiplex PCR, the extension time is increased so that the enzyme, DNA polymerase, gets enough time to generate a complete copy of the target DNA. A schematic multiplex PCR using three sets of primers for amplification of three regions on DNA has been presented in Fig. 3.7.

In forensic DNA fingerprinting, PCR and multiplex PCR are widely used for amplification of biological evidences due to the usability of trace amount of DNA template available from the source of crime exhibits. Additionally, in the case of forensic samples, many times degraded samples are found, hence the degraded fragments of DNA. However, in the case of PCR, only a few hundred base-pair lengths of a DNA sample can be used for effective amplification. Another advantage is the simultaneous amplification of large number of specific DNA sequences by multiplex reactions. There is a less/no chance of amplification of the contaminant DNA, mostly the microbial DNA, as the human-specific DNA primers are used in PCR. Many kits for simultaneous amplification of multiple DNA fragments are available commercially nowadays [42]. In this context, the FBI had issued guidelines for the inclusion of 13 STR markers (CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11) collectively called as the Combined DNA Index System (CODIS). However, seven additional STR loci (D1S1656, D2S441, D2S1338, D10S1248,

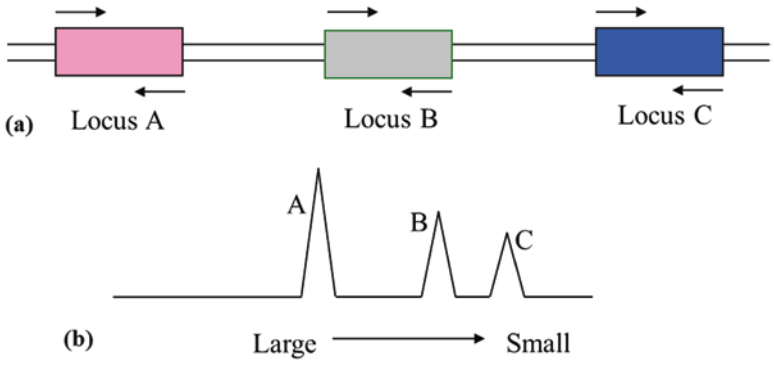

Fig. 3.7 Schematic representation of a multiplex PCR for the amplification of three regions of a DNA strand using three primers sets simultaneously

Table 3.2 List of CODIS STR loci for recommended use in forensic DNA typing

Sl. no.	STR locus	Chromosome position	Nucleotide repeats	Sl. no.	STR locus	Chromosome position	Nucleotide repeats
1	D1S1656	1q42	Tetra	11	D10S1248	10q26.3	Tetra
2	TPOX	2p25.3	Tetra	12	TH01	11p15.5	Tetra
3	D2S441	2p14	Tetra	13	vWA	12p13.31	Tetra
4	D2S1338	2q35	Tetra	14	D12S391	12p13.2	Tetra
5	D3S1358	3p21.31	Tetra	15	D13S317	13q31.1	Tetra
6	FGA	4q31.3	Tetra	16	D16S539	16q24.1	Tetra
7	D5S818	5q23.2	Tetra	17	D18S51	18q21.33	Tetra
8	CSF1PO	5q33.1	Tetra	18	D19S433	19q12	Tetra
9	D7S820	7q21.11	Tetra	19	D21S11	21q21.1	Tetra
10	D8S1179	8q24.13	Tetra	20	D22S1045	22q12.3	Tri

Table 3.3 Commonly used fluorescent dyes for multiplex PCR and their characteristic features

Sl. no.	Dye	Color	Ab_{max}	Em_{max}
1	6-FAM	Blue	495 nm	520 nm
2	HEX	Green	535	556
3	TET	Green	521	536
4	VIC	Green	538	554
5	NED	Yellow	494	517
6	JOE	Green	520	548

D12S391, D19S433, and D22S1045) have been selected by the CODIS Core Loci Working Group with the aim to implement the new 20 CODIS core loci by January 1, 2017 [24] (Table 3.2).

The basic and most important problem associated with multiplexing system is the resolution of different amplified regions which is reflected in the form of peaks generated by capillary-based genetic analyzer [38]. Labeling of all the primer sets with a single fluorescent dye may lead to the overlapping of peaks, and a distinctive resolution of different markers may not be achieved. In this regard, multiple dye chemistry is employed for simultaneous amplification of multiple genes with high degree of resolution [13]. Use of multiple dye chemistry leads to the segregation of alleles with base-pair length of close vicinity by different colored fluorescent-labeled primers. Few commonly used fluorescent dyes used to label the primers for amplification of multiple DNA fragments by multiplex PCR have been described in Table 3.3.

3.6 Genotyping and Analysis

Post-PCR steps of DNA fingerprinting are highly crucial which include the genotyping and analysis. Mostly, the generated PCR products are separated according to their size and dye color using electrophoresis followed by laser-induced

fluorescence with multiwavelength detection [23]. The DNA fragments of known size called as internal size standard along with the amplified DNA fragments with differently labeled primer sets are subjected to electrophoresis simultaneously. Electrophoresis results in the generation of multicolored electropherograms which are analyzed by software, and the STR allele sizes are automatically detected based on the standard curve produced from the internal size standard. In this regard, the STR genotyping is generated by comparing the allele sizes of each sample to the sizes of alleles present in an allelic ladder. Allelic ladder harbors the common alleles that have been sequenced previously [44]. Finally, the allelic ladder along with the internal size standard is subjected to capillary electrophoresis (CE) in one injection, and the sample alleles with same internal size standards are run in other subsequent injections on capillary in a sequential fashion.

In most of the genetic typing laboratories, resolution of the amplified products is carried out by a sophisticated capillary electrophoresis unit commonly called as automated genetic analyzer. In this process the fluorescently labeled amplicons are migrated along the narrow capillaries prefilled with polymer of suitable quality and viscosity under the influence of electric current. The rate of migration of the amplified PCR products is always inversely proportional to their sizes in terms of the base-pair length [33]. Fluorescence emitted by the PCR products of different sizes is detected by the laser detection cell of the instrument at particular distance and is displayed in the form of colored peaks after analysis.

3.6.1 Capillary Electrophoresis

Capillary electrophoresis (CE) is one of the new discoveries in the field of electrophoresis. In this regard, separation of DNA by CE was first performed in the late 1980s [8].

3.6.1.1 Advantages of Capillary Electrophoresis over Slab Gels

There are many advantages of analyzing DNA through capillary electrophoresis. Some of the advantages include the following points:

- During CE, the injection, separation, and detection steps can be of fully automated in nature, thus permitting multiple samples to be run simultaneously and unattended.
- Very minute quantities of amplified DNA samples are required during the injection process which plays an important role in usability of precious forensic specimens for their further storage and usages.
- As higher voltage is applied during capillary electrophoresis, the process of separation of amplified products is conducted in lesser time with improved heat dissipation from the capillaries.

- Following the completion of a run, a quantitative information is readily available in the electronic format.
- No chance of cross-contamination as leaking over of the samples from the adjacent cells is not permitted in CE which is a major problem in slab gel electrophoresis.

3.6.1.2 Components of CE

The primary components of a typical CE instrument include a narrow capillary, two buffer vials, two electrodes connected to a high-voltage power supply, laser excitation source, fluorescence detector, autosampler to hold the sample tubes, and computer to control the sample injection and detection (Fig. 3.8).

The capillary used during CE is mostly a narrow glass tube of approximately 36–50 cm in length and 50 μm in diameter. The capillary is filled with a viscous polymer solution which acts as a gel generating a sievelike structure for the amplified DNA products to pass through it. In this process, the DNA samples are placed in a tray followed by injection onto the capillary by applying high voltage to the samples sequentially [8]. In this process a high voltage of around 15,000 volts is applied across the capillary after injection in order to separate the DNA fragments in a lesser time. Finally, the fluorescent dye-labeled products are analyzed when they pass through the detection window and are excited by a laser beam. This is followed by the computerized data acquisition which enables rapid analysis of data and digital storage of the results. The electronic fields generated during CE are 10–100 times stronger in comparison with the slab gels, thus generating the results faster. The time span from sample injection to sample detection is detected with a laser placed near the end of the capillary [37]. The DNA fragments migrated through gel under influence of electric current reach the detection point as per their molecular weight, as the smaller molecules move faster. The acquired data in the form of function of relative fluorescence intensity is used to quantify the DNA molecules passing through the detector.

3.6.2 Sample Preparation and Injection

A small amount of PCR products are diluted with de-ionized high-quality formamide with low conductivity prior to subjecting in capillary electrophoresis. Formamide is a strong denaturant which is used for the generation of single-stranded DNA fragments to be detected by the genetic analyzer. A small fraction of formamide is sufficient enough to denature a high amount of DNA molecules [54]. Additionally, the samples are subjected to temperature shock from 95 °C to snap-cooling on ice to ensure a complete denaturation of the double-stranded DNA molecules.

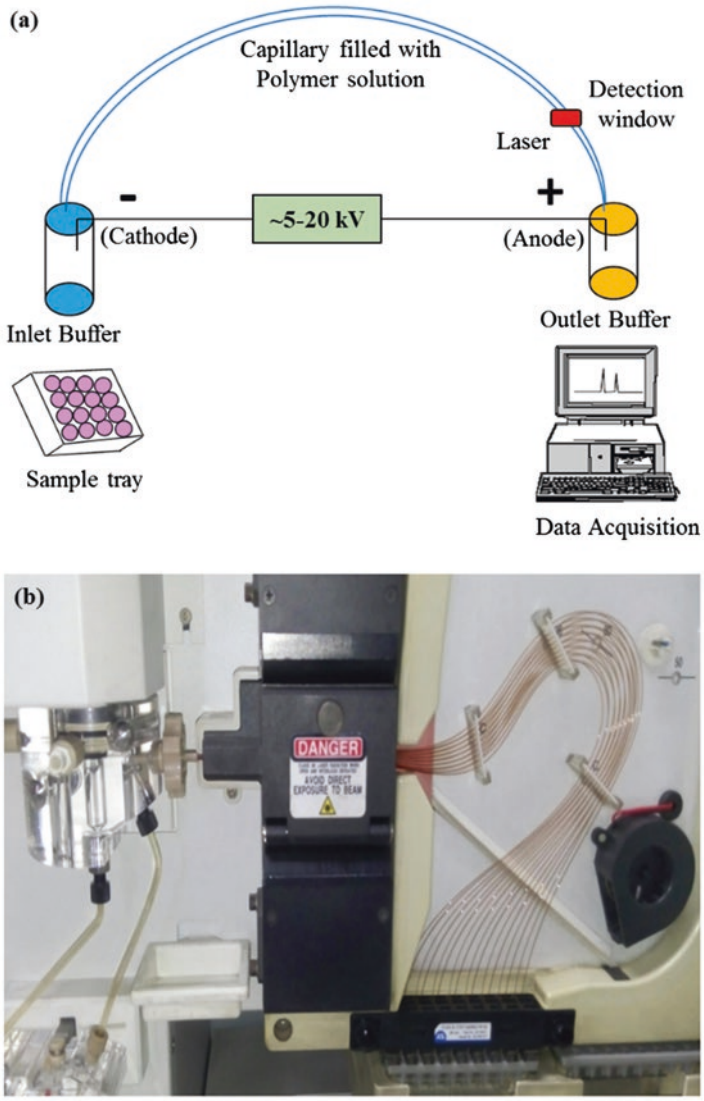

Fig. 3.8 (**a**) Schematic presentation of capillary electrophoresis technology used for DNA analysis and (**b**) a typical genetic analyzer used in most of the DNA typing laboratories

3.6.2.1 Electrophoretic Separation

During electrophoresis, the position of a DNA band on gel is measured in comparison with the positions of the standard containing known DNA fragment sizes. Thus, electrophoresis can be called as a relative measurement technique, and the molecular weight markers are run along with the samples. In this way, the fragment size of

an unknown sample can be estimated by comparing with the fragment size of the closest sample post identical electrophoretic conditions [33].

3.6.2.2 Sample Separation

In addition to the samples, many other factors also affect the separation of samples in a significant magnitude, i.e., the polymer used, the capillary and buffer used during electrophoresis, and the strength of the electric field. The resolution of the degree of sample separation can be determined from the values of the STR allelic ladders used along with the samples during electrophoresis [56]**.**

Many types of sieving media are utilized for electrophoretic separations depending on the physical characteristics of the separating media. In this context, chemical gels such as the common polyacrylamide gels are used in denaturing slab gel electrophoresis. These polyacrylamide gels are rigid cross-linked materials linked together by strong covalent bonds. Additionally, another material called as agarose is also used to prepare physical gels. The shape of the agarose is the result of weaker intermolecular forces produced due to the entanglement of different strands of agarose molecules. Another type of sieving materials used is called the entangled polymers. These materials are also characterized by their intermolecular interactions. Due to their inability of holding the shape, they are not called as the true gels [10]. Rapid increase in viscosity with the polymer concentration reaching the threshold limit is a typical characteristic feature of the entangled polymers. There are evidences of using entangled polymers for separation of biopolymers. The combination of capillary electrophoresis technology and entangled polymers is highly accredited due to their efficient anti-convective and heat dissipation properties which allow the separation of materials with higher resolution. A dilute and low-viscosity polymer solution can be used as a separation medium for high-resolution separation of DNA mixtures. However, an ideal polymer should possess the characteristic feature of the classical gel coupled with the characteristic of low viscosity which will allow its easy replacement which has been achieved by using the performance optimized polymers [14].

3.6.3 Detection of STRs

The amplified products with the fluorescent dye-labeled primers are detected by fluorescent detection. The effective simultaneous use of different colored dyes results in the development of organized peaks. The dyes used for labeling the primers in a multiplex PCR are fluorescent in nature which emit colors at different wavelengths. The usability of this technique is however dependent on the multiwavelength detection permitting multiplexing of STR markers. In this technique the dye of interesting unique properties is attached to the 5′ (nonreactive) end of each primer set. The dye sets are excited by a single argon-ion laser of 488 nm but fluoresce in

different wavelengths. Finally, a charged coupled device (CCD) camera with properties of multiwavelength analyzer is used to determine the dye. This is based on the emission of each fragment with the passage through the detector window. Thus, this technique allows the analysis of DNA fragment of similar size but labeled with different dyes of different wavelengths [17].

3.6.4 Interpretation of Results

The major principle used for interpretation of results includes that, though the dyes are of different colors, overlapping of these spectra occurs as the wavelengths that determine the color do not occur discretely. Due to these spectral overlaps, a standard matrix is run using the instrument. The matrix is created by running each dye separately, and finally the instrument compares the color of the questioned sample with the standard one to generate the result (Fig. 3.9). The matrix file is represented in tabular format (Table 3.4). The instrument software uses the matrix file to normalize the fluorescence between the dyes to create the virtual filter, thus separating the colors for sample analyses.

Finally, a standard curve is drawn by the software using the migration of the known size standards vs. the time taken for migration of the fragments of different sizes. In most of the cases, the known standards are labeled with a distinguished dye, and the size standards are run simultaneously with the unknown test samples, and the size of the test sample is derived by comparing its rate of migration with the standard samples. Another important factor associated with the interpretation of results is the allelic ladder. Allelic ladder contains all the possible alleles of all the markers used in multiplex PCR which is run simultaneously with the each set of test samples. Hence, the length of the unknown samples is measured in terms of base pairs by plotting the base pairs of unknown samples in the standard curve derived from the repeat numbers of the corresponding markers and the allelic ladder.

3.7 Conclusion

DNA fingerprinting technique, the state-of-the-art method, is regarded as one of the most important breakthroughs of forensic science. The most commonly used DNA fingerprinting technique follows the common set of experiments, i.e., DNA isolation, quantification of DNA, multiplex PCR, genotyping by capillary electrophoresis, and interpretation of results. A generalized DNA fingerprinting experiment may yield three possible results, i.e., the inclusion, the exclusion, and the null result. Thus, current focus of research should target the null result possibilities to reduce its occurrence. The rapidity and cost-effectiveness aspect of the DNA typing technology should also be explored.

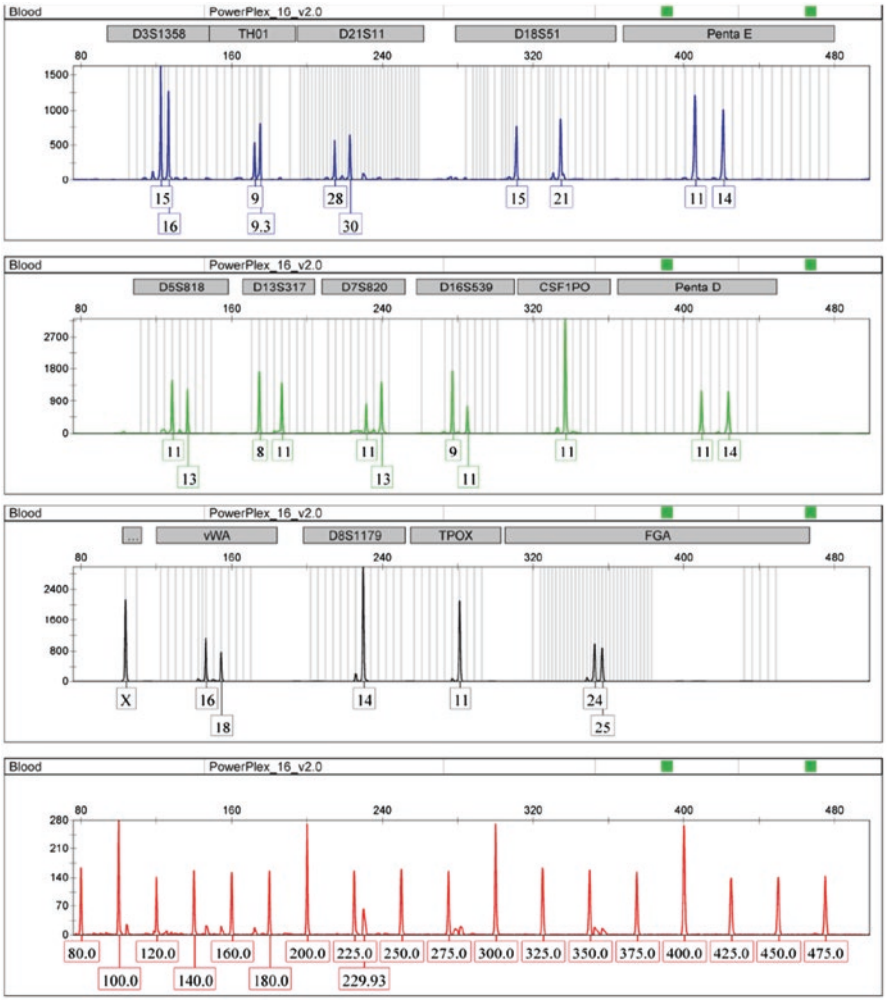

Fig. 3.9 A generalized human autosomal STR DNA profile

Table 3.4 A generalized matrix file with corresponding maximum emission of 1000 at a particular wavelength

	B	G	Y	R
B	1.0000	0.5423	0.0428	0.0010
G	0.6601	1.0000	0.5460	0.0059
Y	0.3459	0.5735	1.0000	0.0893
R	0.1693	0.3236	0.5922	1.0000

References

1. Ahmed HA, MacLeod ET, Hide G, Welburn SC, Picozzi K (2011) The best practice for preparation of samples from FTA® cards for diagnosis of blood borne infections using African trypanosomes as a model system. Parasit Vectors 4:68
2. Akbari M, Hansen MD, Halgunset J, Skorpen F, Krokan HE (2005) Low copy number DNA template can render polymerase chain reaction error prone in a sequence-dependent manner. J Mol Diagn 7:36–39
3. Alonso A, Martin P, Albarrán C, Garcia P, Fernandez de Simon L, JesúsIturralde M, Fernández-Rodriguez A, Atienza I, Capilla J, García-Hirschfeld J, Martinez P, Vallejo G, García O, García E, Real P, Alvarez D, León A, Sancho M (2005) Challenges of DNA profiling in mass disaster investigations. Croat Med J 46:540–548
4. Al-Soud WA, Rådström P (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. Appl Environ Microbiol 64:3748–3753
5. Arnold T, Linke D (2008) The use of detergents to purify membrane proteins. Curr Protoc Protein Sci 4:4.8.1–4.8.30
6. Billard A, Laval V, Fillinger S, Leroux P, Lachaise H, Beffa R, Debieu D (2012) The allele-specific probe and primer amplification assay, a new real-time PCR method for fine quantification of single-nucleotide polymorphisms in pooled DNA. Appl Environ Microbiol 78:1063–1068
7. Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, Little S (1997) The elimination of primer-dimer accumulation in PCR. Nucleic Acids Res 25:3235–3241
8. Butler JM, Buel E, Crivellente F, McCord BR (2004) Forensic DNA typing by capillary electrophoresis using the ABI prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 25:1397–1412
9. Cha RS, Thilly WG (1993) Specificity, efficiency, and Fidelity of PCR. Cold Spring Harbor Laboratory ISSN 1054-9805/93
10. Christoph Heller C, Pakleza C, Viovy JL (1995) DNA separation with field inversion capillary electrophoresis. Electrophoresis 16:1423–1428
11. Deepak SA, Kottapalli KR, Rakwal R, Oros G, Rangappa KS, Iwahashi H, Masuo Y, Agrawal GK (2007) Real-time PCR: revolutionizing detection and expression analysis of genes. Curr Genomics 8:234–251
12. Dell'Anno A, Fabiano M, Duineveld GCA, Kok A, Danovaro R (1998) Nucleic acid (DNA, RNA) quantification and RNA/DNA ratio determination in marine sediments: comparison of spectrophotometric, fluorometric, and high performance liquid chromatography methods and estimation of detrital DNA. Appl Environ Microbiol 64:3238–3245
13. Dobnik D, Štebih D, Blejec A, Morisset D, Žel J (2016) Multiplex quantification of four DNA targets in one reaction with bio-rad droplet digital PCR system for GMO detection. Sci Rep 6:35451
14. Durney BC, Crihfield CL, Holland LA (2015) Capillary electrophoresis applied to DNA: determining and harnessing sequence and structure to advance bioanalyses (2009–2014). Anal Bioanal Chem 407:6923–6938
15. Edwards MC, Gibbs RA (1994) Multiplex PCR: advantages, development, and applications. Genome Res 3:S65–S75
16. Fan A, Byrnes S, Klapperich C (2013) Purification of DNA/RNA in a microfluidic device. Methods Mol Biol 949:403–411
17. Ferri F (1997) Use of a charge coupled device camera for low-angle elastic light scattering. Rev Sci Instrum 68:2265–2274
18. Flock S, Labarbe R, Houssier C (1996) Dielectric constant and ionic strength effects on DNA precipitation. Biophys J 70:1456–1465
19. Ghatak S, Muthukumaran RB, Nachimuthu SK (2013) A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. J Biomol Tech 24:224–231

20. Gong R, Li S (2014) Extraction of human genomic DNA from whole blood using a magnetic microsphere method. Int J Nanomedicine 9:3781–3789
21. Greco M, Sáez CA, Brown MT, Bitonti MB (2014) A simple and effective method for high quality co-extraction of genomic DNA and total RNA from low biomass *Ectocarpussiliculosus*, the model Brown alga. PLoS One 9:e96470
22. Gudnason H, Dufva M, Bang DD, Wolff A (2007) Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. Nucleic Acids Res 35:e127
23. Gupta S, Charakana C, Sreelakshmi Y, Sharma R (2011) Fluorescent dye labeled DNA size standards for molecular mass detection in visible/infrared range. BMC Res Notes 4:12
24. Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. Forensic Sci Int Genet 17:33–34
25. Hobbie RK, Roth BJ (2007) Intermediate physics for medicine and biology. Springer Science & Business Media, New York, p 616
26. Houck MM (2005) DNA and the criminal justice system the technology of justice. J Clin Invest 1115:1398–1398
27. Housley DJE, Zalewski ZA, Beckett SE, Venta PJ (2006) Design factors that influence PCR amplification success of cross-species primers among 1147 mammalian primer pairs. BMC Genomics 7:253
28. Iqbal M, Tao Y, Xie S, Zhu Y, Chen D, Wang X, Huang L, Peng D, Sattar A, Shabbir MAB, Hussain HI, Ahmed S, Yuan Z (2016) Aqueous two-phase system (ATPS): an overview and advances in its applications. Biol Proced Online 18:18
29. Ishii K, Fukui M (2001) Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Appl Environ Microbiol 67:3753–3755
30. Kebelmann-Betzing C, Seeger K, Dragon S, Schmitt G, Möricke A, Schild TA, Henze G, Beyermann B (1998) Advantages of a new Taq DNA polymerase in multiplex PCR and time-release PCR. BioTechniques 24:154–158
31. Koshy L, Anju AL, Harikrishnan S, Kutty VR, Jissa VT, Kurikesu I, Jayachandran P, Jayakumaran Nair A, Gangaprasad A, Nair GM, Sudhakaran PR (2017) Evaluating genomic DNA extraction methods from human whole blood using endpoint and real-time PCR assays. Mol Biol Rep 44:97–108
32. Lebedev AV, Paul N, Yee J, Timoshchuk VA, Shum J, Miyagi K, Kellum J, Hogrefe RI, Zon G (2008) Hot start PCR with heat-activatable primers: a novel approach for improved PCR performance. Nucleic Acids Res 36:e131
33. Lee PY, Costumbrado J, Hsu CY, Kim YH (2012) Agarose gel electrophoresis for the separation of DNA fragments. J Vis Exp (62):3923
34. Liu D, Daubendiek SL, Zillman MA, Ryan K, Kool ET (1996) Rolling circle DNA synthesis: small circular oligonucleotides as efficient templates for DNA polymerases. J Am Chem Soc 118:1587–1594
35. Lorenz TC (2012) Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. J Vis Exp (63):3998
36. Ma TS (1995) Applications and limitations of polymerase chain reaction amplification. Chest J 108:1393–1404
37. Marino MA, Devaney JM, Davis PA, Girard JE (1999) Optimization of intercalation dye concentration for short tandem repeat allele genotyping using capillary electrophoresis with laser-induced fluorescence detection. J Chromatogr B Biomed Sci Appl 732:365–374
38. Markoulatos P, Siafakas N, Moncany M (2012) Multiplex polymerase chain reaction: a practical approach. J Clin Lab Anal 16:47–51
39. Mohammadi Z, Shalavi S, Jafarzadeh H (2013) Ethylenediaminetetraacetic acid in endodontics. Eur J Dent 7:S135–S142
40. Nigam PS (2013) Microbial enzymes with special characteristics for biotechnological applications. Biomol Ther 3:597–611
41. Pamidimarri DVNS, Meenakshi SR, Boricha G, Reddy MP (2009) A simplified method for extraction of high quality genomic DNA from *Jatrophacurcas* for genetic diversity and molecular marker studies. Indian J Biotechnol 8:187–192

42. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis 30:3682–3690
43. Ponche F, Toomes C, Bransfield K, Leong FT, Douglas SH, Field SL, Bell SM, Combaret V, Puisieux A, Mighell AJ, Robinson PA, Inglehearn CF, Isaacs JD, Markham AF (2003) Real-time PCR based on SYBR-green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. BMC Biotechnol 3:18
44. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01 [AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–958
45. Puskás LG, Bottka S (1995) Reduction of mispriming in amplification reactions with restricted PCR. Genome Res 5:309–311
46. Romsos EL, Vallone PM (2015) Rapid PCR of STR markers: applications to human identification. Forensic Sci Int Genet 18:90–99
47. Rutledge RG, Stewart D (2008) Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. BMC Mol Biol 9:96
48. Rychlik W, Spencer WJ, Rhoads RE (1990) Optimization of the annealing temperature for DNA amplification in vitro. Nucleic Acids Res 18:6409–6412
49. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York
50. Tan SC, Yiap BC (2009) DNA, RNA, and protein extraction: the past and the present. J Biomed Biotechnol 2009:574398
51. Templeton NS (1992) The polymerase chain reaction. History, methods, and applications. Diagn Mol Pathol 1:58–72
52. Tweedie JW, Stowell KM (2005) Quantification of DNA by agarose gel electrophoresis and analysis of the topoisomers of plasmid and M13 DNA following treatment with a restriction endonuclease or DNA topoisomerase I. Biochem Mol Biol Educ 33:28–33
53. 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
54. Wang X, Lim HJ, Son A (2014) Characterization of denaturation and renaturation of DNA for DNA hybridization. Environ Health Toxicol 29:e2014007
55. Zhang S, Tian H, Wu J, Zhao S, Li C (2013) A new multiplex assay of 17 autosomal STRs and amelogenin for forensic application. PLoS One 8:e57471
56. Zhu M, Hansen DL, Burd S, Gannon F (1989) Factors affecting free zone electrophoresis and isoelectric focusing in capillary electrophoresis. J Chromatogr A 480:311–319

Chapter 4
STR Typing and Available Kits

Pankaj Shrivastava, Hirak Ranjan Dash, R. K. Kumawat, Ankit Srivastava, and Jahangir Imam

Abstract This chapter describes a time-scale development of short tandem repeats (STRs) and STR-based DNA technology used in forensic DNA analysis. The text describes the subsequent advancements in the development of the STR multiplex systems. This sequential development in STR-based multiplex systems has increased efficiency, sensitivity, and inhibitor tolerance with improved buffers. The development of different variants of STR-based multiplex systems utilizing different variants including autosomal, mini-, Y-, and X-STRs has increased the flexibility and ease of the forensic DNA analyst. With the availability of genetic analyzers, utilizing sixth dye has created new possibilities toward availability of expanded STR multiplex without affecting the requirement of input DNA for PCR. Rapid PCR and application of next-generation sequencing are also discussed along with the already validated capillary electrophoresis technology.

Keywords STRs · DNA analysis · Y and X STRs · Multiplex PCR · Autosomal · NGS

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

R. K. Kumawat
State Forensic Science Laboratory, Jaipur, Rajasthan, India

A. Srivastava
Dr. A.P.J. Abdul Kalam Institute of Forensic Science, Bundelkhand University, Jhansi, Uttar Pradesh, India

J. Imam
DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home, Jail and Disaster Management, Government of Jharkhand, Ranchi, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_4

4.1 Introduction

Eukaryotic genomes contain a large number of repeated sequences (sometimes, in thousands) of DNA [1]. These sequences are present in varied sizes and are mostly present near the chromosomal centromere and called satellite DNA. In frequent observations of minor satellite bands in equilibrium density gradient centrifugation experiments, the term satellite DNA was coined and assigned to these sequences [2, 3]. Short tandem repeats (STRs) are tandemly repeated segments of a core repeated DNA sequence with varied length which are dispersed all over the genome. The probable reason for this variation may be insertion, deletion, or mutation in nucleotide sequence. The core repeat unit of around 10–100 bases in length is also called as minisatellites or VNTR (variable number of tandem repeats) [4, 5]. Another category of 2–6 bp repeat regions of DNA is known as microsatellite or simple sequence repeats (SSRs) or most popularly the short tandem repeats (STRs). STRs with the characteristics of definite amplification without the issue of differential amplification have recognized them as the markers of choice nowadays. Another reason for the popularity of these STR markers might be due to the fact that both the alleles in any heterozygous position are analogous and small in size. The polymorphic nature of these STR markers makes them a good candidate for individual identification. The present DNA profiling technique is based on PCR and uses these simple sequences or STRs. Unrelated people almost certainly have different numbers of repeat units, and because of this fact, STRs are being used to discriminate between unrelated individuals. The present forensic DNA typing technique is PCR based and uses STRs. It is easy to discriminate between unrelated individuals using STRs because at any position, unrelated individuals have different numbers of repeat units. The use of tetranucleotide repeats, containing 4-bp repeat structure, is very common in forensic laboratories (Fig. 4.1). In 1997, the Federal Bureau of Investigation (FBI) Laboratory identified 13 STR loci as the core STR loci to support sharing of worldwide DNA data, and this initiated the formation of the US national DNA database.

Flanking regions around the STR repeats are never changing in nature and are used invariably for STR analysis. These STR loci which are actually specific locations on a chromosome are targeted with primers (sequence specific) and PCR amplified. The DNA fragments that are produced after this PCR amplification are separated and finally detected using capillary electrophoresis by using genetic analyzer. After the first report of trimeric and tetrameric tandem repeats in 1991 [6] and their utilization in a forensic case in the same year (1991), the use of STRs/microsatellites has now become the backbone of the forensic DNA testing globally [6–8]. The best suitable STR markers for human identification purposes are those which are polymorphic, i.e., which show the maximum interindividual variability.

The present forensic DNA typing is done using a set of STR markers (already proven to be polymorphic) which are structurally similar to minisatellites markers

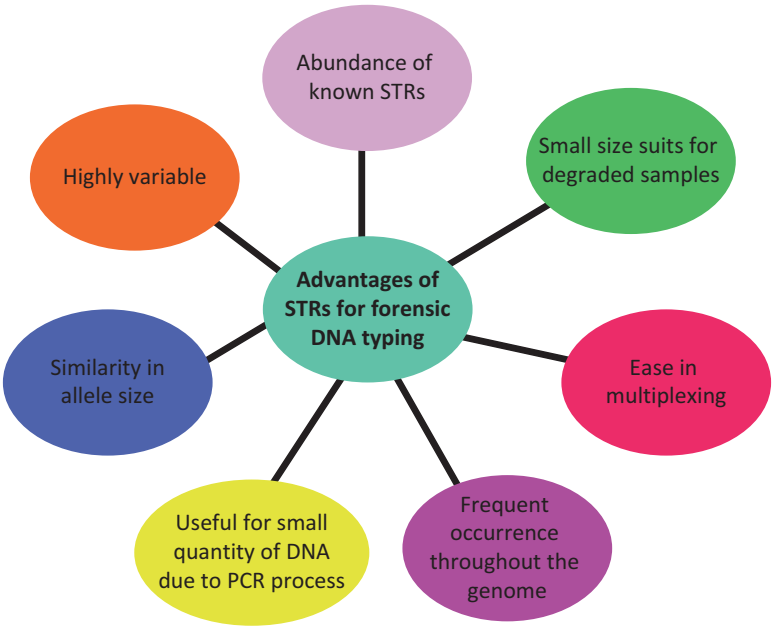

Fig. 4.1 Advantages of STRs for forensic DNA typing

but have much shorter repeat sequence, and thus their PCR-based amplification is easier in multiplex [9]. Nowadays, generating DNA profile of an individual more than 20 STRs can be analyzed in a single injection/run of capillary electrophoresis (CE). Multiplex kits from all the leading manufacturers are available now for this purpose. As per the guidelines requested by criminal databases around the globe, two sets of STR markers exist, i.e., 12 markers according to European standard [10] and 13 markers further expanded to 20 markers as per US CODIS standard [11].

4.2 Sequential Developments of STRs in Forensics

The sequential initial developments in the use of STRs in forensic DNA typing are listed in Table 4.1 and Fig. 4.2. For separation and detection of fragments, capillary electrophoresis (CE) and gel electrophoresis are two common methods. Of these two methods, present-day forensic DNA technology mostly utilizes genetic analyzers which are based on CE.

Table 4.1 Sequential initial developments of STRs in forensics (From 1991 to 2000)

Year	Development
1991	First STR was discovered
1993	Gender-specific marker amelogenin was discovered
1994	Forensic Science Service (FSS) developed quadruplex (containing four loci – THO1, FES/FPS, vWA, and F13A1)
	First commercial STR kit for silver stain analysis was launched by Promega which comprised the three loci – a triplex "**CTT**" constituting **C**SF1PO, **T**POX, and **T**HO1
1995	In April 1995, the United Kingdom initiated their first ever national DNA database with six STRs, i.e., TH01, vWA, FGA, D8S1179, D18S51, and D21S11
	One additional marker D3S1358, in addition to the UK standard, was adopted as European standard set
1996	FBI started efforts for common core loci by launching STR project which was run between April 1996 and April 1997
	Seventeen candidate STR loci were evaluated
1997	First CODIS loci (CODIS 13) was chosen
1999	European core loci (ESS 12) were declared
2000	Both ABI and Promega started offering multiplex kits containing CODIS loci

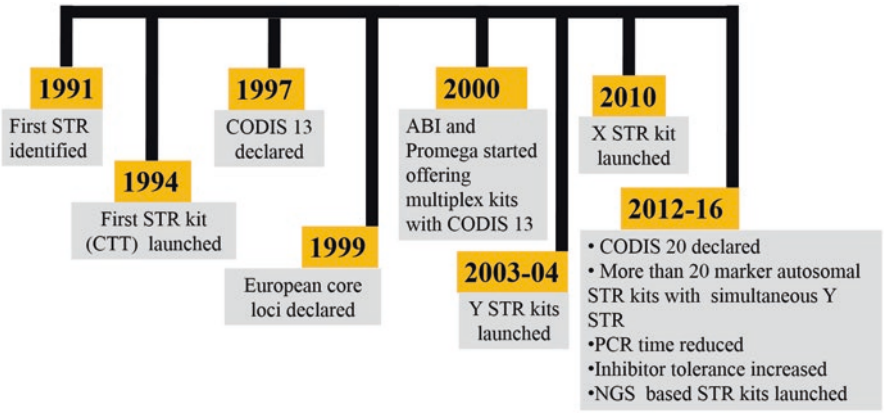

Fig. 4.2 Chronological advancements of STR technique

4.3 Characteristics of STR Markers

Among various STRs being applied for forensic DNA analysis, tetranucleotide repeats are in an upper hand in comparison with di- or trinucleotide repeat STRs. Most of the multiplex systems being used in forensic DNA typing have used tetranucleotide repeats. Though penta- and hexanucleotide repeats are present in the human genome, they are less commonly used in currently available multiplex kits

used for forensic DNA typing. The following characteristics of STR markers have been utilized for their use in forensic DNA application [12, 13]:

- High discriminating power with around 70% observed heterozygosity.
- Loci from separate chromosomal locations are chosen to avoid linked loci.
- These should be robust and should ensure reproducibility of results during multiplexing with other STRs.
- They must exhibit low stutters.
- Mutation rate of these markers should be low.
- Smaller size for their better use with degraded forensic samples (Predicted length 90 to 500 bp).

4.4 Nomenclature of STRs and Development of STR Multiplex

The STR loci which are in use for the purpose of forensic DNA typing are noncoding but mostly are present in proximity to the introns or within them. In this regard, the STR loci present within introns are mostly named afterward the associated gene, e.g., vWA, a STR locus is named after its occurrence in intron 40 of a gene von Willebrand factor [8] on chromosome 12. Similarly, another STR marker present on intron 10 of chromosome 2 coding for thyroid peroxidase gene has been named as TPOX.

The naming of STR markers is simple mostly and can be easily deciphered by seeing the name of the STR marker. For example, for STR locus D21S11, D stands for DNA, 21 is the number of chromosome, S specifies the presence of a lone copy, and the final integer designates the position of the repeat to be the 11th of 21st chromosome. In 1994 only four loci (Table 4.1) [14, 15] were available and known for forensic DNA typing which extended in the next 1 year to 7 [16–19]. The initial multiplex reactions with these STR markers were formulated by the FSS, UK [20]. After the initial inception and use of STR markers, two leading brands in the field, Applied Biosystems (now Thermo Fisher Scientific, USA) and Promega Corporation, USA, took the lead in further development and started their marketing globally. By the end of year 1997, the available multiplex kits started amplifying nine STR loci in addition to a sex-determining marker amelogenin. In 2000 the list of available markers in STR multiplex increased to 16 [21], and from 2000 to almost more than a decade, the forensic DNA examination was based on these 16 markers (Fig. 4.3). In 2010, the third brand Qiagen also launched STR multiplex systems, and after 2010 some other brands also showed their presence in the global STR multiplex market with 16 markers. After 2011, the number of STR markers further increased, and today more than 20 markers of STR multiplex systems are being used in forensic DNA typing.

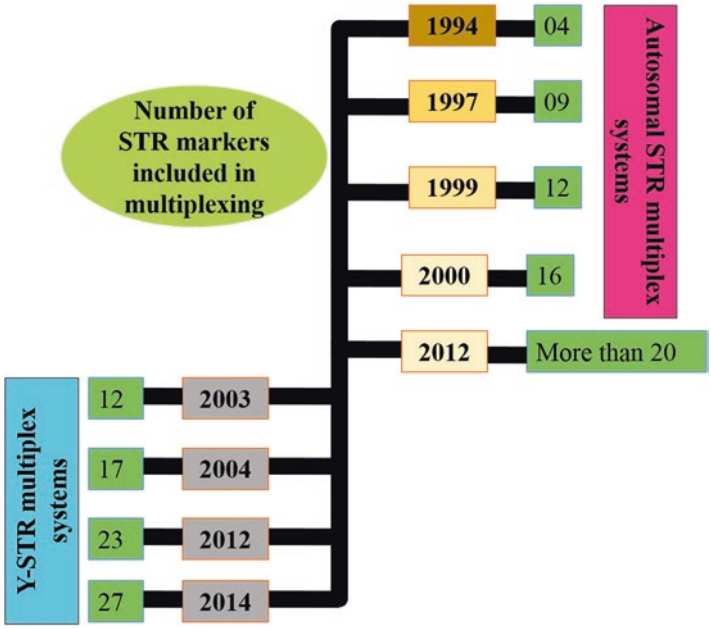

Fig. 4.3 Gradual increase in induction of STR markers to the multiplex system

4.5 Commercially Available STR Kits

The whole process for STR typing is a multistep process as shown in Fig. 4.3. After this the process ends with interpreting the results by comparing the results with the provided referral samples as match or non-match. The obtained STR profile may be from a single source, or it may be a mixed profile which is obtained mostly in evidences of sexual assault cases. Interpretation of mixed DNA profiles is a challenging job as the peak height of the STR alleles is obtained in different ratios with respect to the amount of body fluids of the victim and perpetrator's bodily fluids contributing to the complex and challenging results.

Commercial kits are choice of most forensic laboratories due to their prevalidated nature with standardized workflow providing the ease to the end users irrespective of their high cost. The user has to be particular in accessing the quantity of DNA from their sample after DNA extraction as per the requirement of the multiplex kit being used to get a clean STR profile (Fig. 4.4). These pre-formulated systems not only simplified the protocols but also removed the difficulties of PCR. Commercially available STR kits are the pre-formulated mixtures of primers, standard buffer containing the polymerase enzyme and dNTPs to get the best possible STR profiles. These STR kits are provided with size standard and allelic ladders for the purpose of genotyping after amplification with all predefined procedures and protocols.

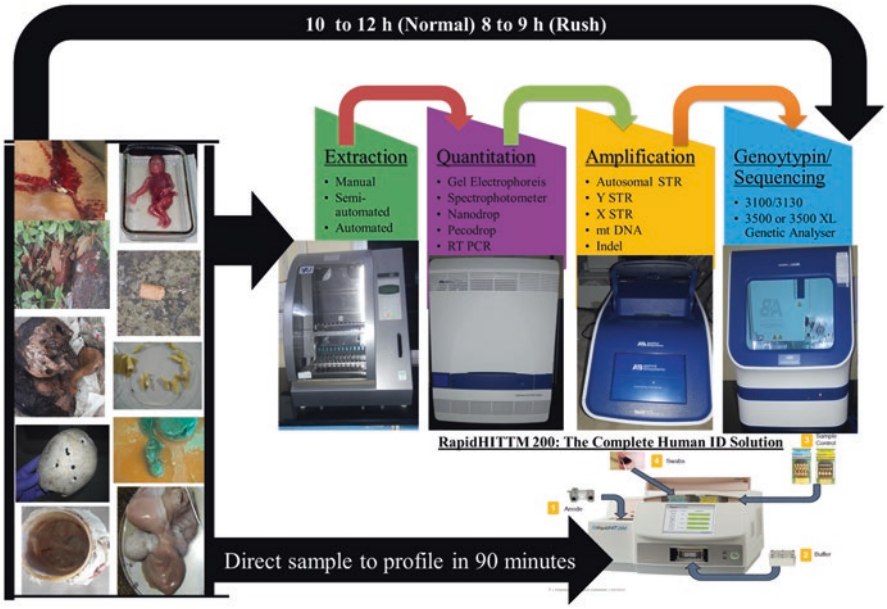

Fig. 4.4 Steps in present CE-based STR typing

Table 4.2 Currently available STR kits, the number of loci analyzed, and the respective probability of identity value (PI)

STR kit	Number of loci	PI ($N = 1036$)
CODIS 13	13	5.02×10^{-16}
ESS 12	12	3.04×10^{-16}
CODIS 20	20	9.35×10^{-24}
GlobalFiler™ STR kit	23 (excluding amelogenin)	7.73×10^{-28}
Investigator® 24plex kit	23 (excluding amelogenin)	7.73×10^{-28}
PowerPlex® Fusion System	24 (excluding amelogenin)	6.58×10^{-29}
PowerPlex® Fusion 6C System	27 (excluding amelogenin)	2.30×10^{-32}

"Variability of new STR loci and kits in US population groups," profiles in DNA https://www.promega.com/resources/profiles-in-dna/2012/variability-of-new-str-loci-and-kits-in-us-population-groups/Published 2012

4.6 New-Generation STR Kits

There is regular advancement in the forensic DNA typing. The increase in the number of STR markers is primarily meant for significant gains in discriminatory power, further reducing the chances of adventitious matches (Table 4.2). Now with the new-generation multiplex systems using varied dye sets (Fig. 4.5), PCR time is also reduced, and the multiplex system is more sensitive toward low quantity of DNA with more inhibitor tolerance. To foresee the CODIS expansion guidelines,

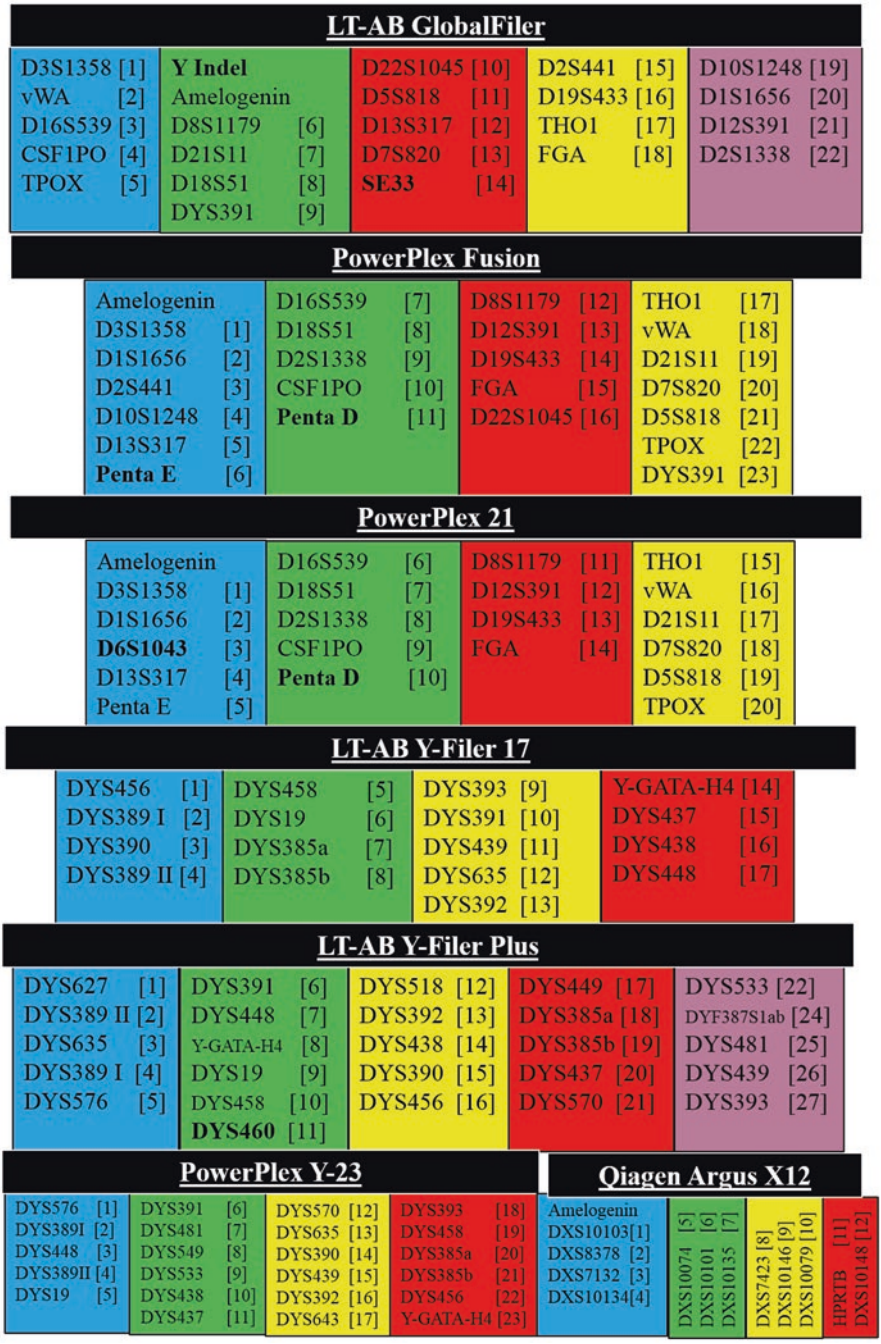

Fig. 4.5 Most recent available multiplex kits for autosomal, Y-, and X-STRs and their constituent dye label colors

both the leading kit manufacturers ABI and Promega have expanded the number of STRs (from 15 to 16 STRs to 20 plus STRs) in their largest set of multiplex kits maintaining amplification efficiency with the increased loci along with increased sensitivity [22].

4.7 Rapid PCR

Before 2012, most of the commercially available multiplex STR systems were not optimized for rapid PCR. ABI IdentiFiler, ABI IdentiFiler plus, PowerPlex 16, PowerPlex 16HS, and IDPlex which were developed to cope up and/or to meet the fulfillment of the CODIS 13 used to take approximately 3 h for completing the amplification. In 2012, the first phase toward fast/rapid PCR started with the launch of six-dye-based GlobalFiler from Thermo Fisher (previously known as Life Technologies and/or Applied Biosystems) and including the launch of PowerPlex Fusion and Powerplex Y23 from Promega Corporation for the pre-available five-dye-based capillary array (CE) platforms and few other six-dye-based kits from Promega and Qiagen (Table 4.3). Six-dye multiplex PCR systems can only be run on 3500 series and on 3130 (with advanced software) of capillary array (CE)-based

Table 4.3 Rapid multiplex STR systems available in and after 2012

Sl. No.	Name of the kit	Manufacturer	Total no. of markers	Type of markers in the multiplex			
				Autosomal	Sex determination	SNP	Others
1	GlobalFiler	Thermo	24	21 nos.	Amelogenin	1 Y-Indel	1 Y STR
2	GlobalFiler Express	Thermo	24	21 nos.	Amelogenin	1 Y-Indel	1 Y-STR
3	YfilerPlus	Thermo	27	–	–	–	27 Y-STRs
4	PowerPlex 21 system	Promega	21	20 nos.	Amelogenin	–	–
5	PowerPlex Fusion	Promega	24	22 nos.	Amelogenin	–	1 Y-STR
6	PowerPlex Y23	Promega	23	–	–	–	23 Y-STRs
7	PowerPlex Fusion 6C	Promega	27	23 nos.	Amelogenin	–	3 Y-STRs
8	Investigator 24plex	Qiagen	24	21 nos.	Amelogenin	–	1 Y-STR, 2 quality sensors
9	Investigator 24plex GO	Qiagen	24	21 nos.	Amelogenin	–	1 Y-STR, 2 quality sensors

ABI genetic analyzers; however five-dye-based multiplex systems can run on 3100, 3130, and 3500 series of ABI genetic analyzers. These multiplex systems are capable of amplifying in less than 2 h and are also direct PCR amplification compatible for a few forensic samples with some pretreatment, thus eliminating the extraction and quantitation steps by some pretreatment which also eliminates within the genotyping process.

4.8 Rapid and Fully Integrated Devices for STR Typing

DNA fingerprinting is a multistep process, and integration of all of the processes in a single machine has been a challenge for long. In the last couple of years, this challenge has also been successfully taken care of and as result of several continual few devices was not only developed but validated as well for forensic use. With these devices, DNA extraction to final DNA profile is possible now in a single machine. These rapid DNA systems are fully integrated robotic machines which generate STR profiles from samples within 90 min (Table 4.4).

4.9 Genotyping STRs with NGS

With the advent of NGS (next-generation sequencing) in the last few years, it is now possible to generate genotyping data of STRs by using sequencing technology instead of using CE. The added advantages of using NGS over CE systems are elaborated in Fig. 4.6. More numbers of markers can be typed using NGS along with simultaneous typing of a variety of markers on different chromosomes (autosomal, Y, and X) including mitochondrial DNA, phenotypic features, and even ancestry information which can be deciphered in a single run. But despite of multifaceted use, NGS is yet to be used in routine forensic DNA investigation.

Table 4.4 Summary of integrated platforms (sample in and STR profile out) performing STR typing

System	Chemistry	Time required	Reference number
ANDE	PowerPlex 16	84 min	[23]
RapidHIT 200	PowerPlex 16	90 min	[24]
RapidHIT 200	GlobalFiler Express	120 min	[25]
RapidHIT ID system	GlobalFiler Express	Less than 90 min	[26]
ANDE	PowerPlex 16	84 min	[27]

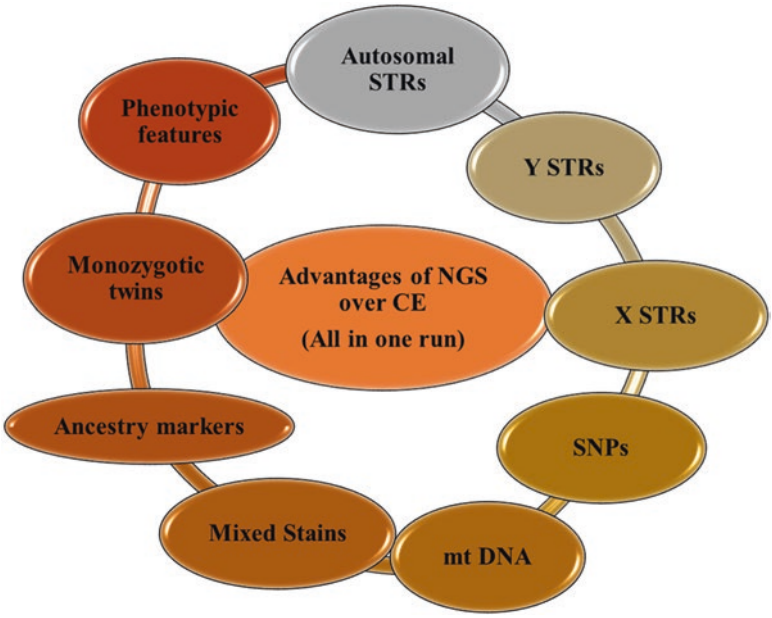

Fig. 4.6 NGS, the new technology for STR typing with much potential

4.10 Web Resources on STRs

Fast and validated processing of forensic samples to get results in a shorter time is need of hour today. The most important aspect after generation of a DNA profile from a forensic sample is to match it with the suspect(s) to aid in forensic investigation. With this factual state comes the need of population database. Though the growth of DNA database has faced the issue of privacy concerns in the last few years, still the importance of DNA database cannot be ignored. On-web databases are available for autosomal, Y-, and X-STR data. A list of web resources on STRs and STR-based database is presented in Table 4.5.

4.11 Conclusion

The virtue of analyzing the challenged samples and producing substantial results for human identification and identification-based research has made the application of STR-based technologies a success in forensic DNA typing. The relentless improvements in almost all the facets of STR-based typing have enhanced the potential of the technology in analyzing more challenged forensic samples. The consistent advancements in the multiplex systems have enhanced the decipherability of STR markers in almost all sort of forensic samples including degraded samples which is

Table 4.5 Currently available web resources on STRs and STR-based databases

Sl. No.	Name	Web link
1	Federal DNA database – FBI	https://www.fbi.gov/services/laboratory/biometric-analysis/federal-dna-database
2	STRidER STRs for identity ENFSI Reference database	https://strider.online/
3	YHRD – Y Chromosome Haplotype Reference Database (online database on Y-STRs)	https://yhrd.org/
4	US-YSTR database (online database on Y-STRs)	https://www.usystrdatabase.org/
5	ChrX-STR.org 2.0 database (online database on X-STRs)	http://www.chrx-str.org/
6	Short Tandem Repeat DNA Internet Data Base (STRbase)	http://strbase.nist.gov/

a result of harsh environmental exposure, touch samples, and samples from mixed sources, leading to the more specific individualization. As CE-based STR technology has the limitation of undetectability of sequence variation in the STR repeats, it is unable to run multiple multiplex systems (autosomal, Y, and X) together in a single run. The next-generation sequencing (NGS) technology is almost ready with its multifaceted benefits for finding a place in forensic DNA examination to overcome with the limitations of CE-based STR technology. Through NGS, lacunae left by CE-based analysis can be filled as the sequence variation can be detected and various multiplex systems can be used together in a single run.

Since NGS is a new technique, not much of the scientists are well versed with the functioning of this technology. International collaborations will result into exchange of ideas on the inception of new technologies and will lead to the advancements in the forensic DNA testing procedures. Improved communication among the forensic scientists across the globe is the need of the hour. The STR analysis technology is being upgraded every day in terms of better sensitivity, increased number of markers for better discrimination, and increased resistance of multiplex systems toward PCR inhibitors; still there is so much which is yet to be done to make the process of human identification more specific and widely acceptable. Population databases of various sorts of STR markers are required to see the distribution of alleles among different populations. The presence of population databases will lead to a more concrete and specific identification. The NGS technology that has recently come into the picture has opened up new avenues of research in forensic DNA typing. The technology is yet to be explored in many aspects like identification of monozygotic twins, physical trait (eye color, hair color) identification, evolutionary and ancestry studies by analyzing a massive number of SNPs, and other informative markers in one run. The advent of NGS has given new grounds to DNA scientists to conduct more and more research by collaborating nationally and internationally and come up with more efficacious ways to solve various varieties of forensic cases with better exactitude.

References

1. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445
2. Britten RJ, Kohne DE (1968) Repeated sequences in DNA. Sci New Ser 161:529–540
3. Primose SB (1998) Principles of genome analysis: a guide to mapping and sequencing DNA from different organisms, 2nd edn. Blackwell Science, Malden
4. Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Physiol B Biochem Mol Biol 126:455–476
5. Tautz D (1993) Notes on the definition and nomenclature of tandemly repetitive DNA sequences. In: DJ Pena (ed) DNA fingerprinting: state of the science. Birkhauser Verlag, Basel, pp 21–28
6. Edwards AL, 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
7. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable minisatellite regions in human DNA. Nature 314:67–73
8. Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Appl 3:13–22
9. Roewer L (2013) DNA fingerprinting in forensics: past, present, future. Investig Genet 4:22
10. Gill P, Fereday L, Morling N, Schneider PM (2006) The evolution of DNA databases – recommendations for new European STR loci. Forensic Sci Int 156:242–244
11. Budowle B, Moretti TR, Niezgoda SJ, Brown BL (1998) CODIS and PCR-based short tandem repeat loci: law enforcement tools. Proceedings of the 2nd European Symposium Human Ident Madison, WI: Promega Corporation, pp 73–88
12. Butler JM (2011). Advanced topics in forensic DNA typing: methodology. 1st edition eBook ISBN: 9780123878236 hardcover ISBN: 9780123745132
13. Gill P, Urquhart A, Millican E, Oldroyd N, Watson S, Sparkes R, Kimpton CP (1996) A new method of STR interpretation using inferential logic -development of a criminal intelligence database. Int J Legal Med 109:14–22
14. Kimpton C, Fisher D, Watson S, Adams M, Urquhart A, Lygo J, et al (1994). Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Legal Med, 106: 302–311
15. Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM et al (1994) The validation of short tandem repeat (STR) loci for use in forensic casework. Int J Legal Med 107:77–89
16. Gill P, Sparkes R, Kimpton C (1997) Development of guidelines to designate alleles using an STR multiplex system. Forensic Sci Int 89:185–197
17. Kimpton CP, Oldroyd NJ, Watson SK, Frazier RR, Johnson PE, Millican ES, et al (1996). Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. Electrophoresis, 17: 11283–93, 1283
18. Sparkes R, Kimpton C, Gilbard S, Carne P, Andersen J, Oldroyd N, et al (1996a). The validation of a 7-locus multiplex STR test for use in forensic casework. (II), artefacts, casework studies and success rates. Int J Legal Med, 109: 195–204
19. Sparkes R, Kimpton C, Watson S, Oldroyd N, Clayton T, Barnett L, et al (1996b). The validation of a 7-locus multiplex STR test for use in forensic casework. (I). Mixtures, ageing, degradation and species studies. Int J Legal Med, 109: 186–194
20. Adrian L, Jennifer EL (2014) Forensic DNA profiling: state of the art. Res Rep Forensic Med Sci 4:25–36
21. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al (2002). Validation of a 16-locus fluorescent multiplex system. J Forensic Sci, 47: 773–785

22. Phillips C, Besada MG, Formoso LF, Magarinos MG, Santos C, Fondevila M, Ballard D, Court DS, Carracedo A, Lareu MV (2014) "New turns from old STaRs": enhancing the capabilities of forensic short tandem repeat analysis. Electrophoresis 35:3173–3187
23. Tan E, Turingan RS, Hogan C, Vasantgadkar S, Palombo L, Schumm JW, et al (2013). Fully integrated: fully automated generation of short tandem repeat profiles. Investig Genet, 4: 2041–2223
24. Jovanovich S, Bogdan G, Belcinski R, Buscaino J, Burgi D, Butts ELR, 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
25. Hennessy LK, Mehendale N, Chear K, Jovanovich S, Williams S, Park C, 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
26. 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
27. Moreno LI, Browna 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

Part II
Applications of DNA Fingerprinting

Chapter 5
Application of DNA Fingerprinting and Wildlife Forensics

Sandeep Kumar Gupta

Abstract Illegal wildlife trade has been detected as a global threat to the sustainability of biodiversity and the security of society. It invited the strict wildlife protection laws. In the past two decades, forensic science has evolved rapidly with its emerging branch "wildlife forensics" with great success. Investigation of wildlife crime is increasingly supporting the implementation of wildlife protection laws. Among two methods, viz., morphological examination and molecular analysis, the latter was proven as the most effective method for dealing such issues. The primers designed for conserved region located on the genes of the mitochondrial genome have enormous scope in dealing the DNA wildlife forensics. We describe the use of morphological examination, DNA typing, and genetic profiling in dealing the wildlife offenses.

Keywords DNA typing · Wildlife crime investigation · Conserved primers · Mitochondrial genome · Wildlife offense

5.1 Introduction

Wildlife crime is defined as "the taking, trading, exploiting, or possessing of the world's wild flora and fauna in contravention of national and international laws." Such offenses occur at a global scale and are a serious challenge for law enforcement. Transport of wildlife part under illegal international trade can have severe implications on other aspects such as invasive species introduction and disease transmission. Illicit trade of plant and wildlife product also affects the people's livelihoods and their resources. It is also threatening the declining populations of endangered species. Selective harvesting of species from the wild may disturb the subtle balance of nature in the ecosystems.

S. K. Gupta (✉)
Department of Animal Ecology and Conservation Biology & Wildlife Forensic and Conservation Genetics Cell, Wildlife Institute of India, Dehradun, India
e-mail: skg@wii.gov.in

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_5

Other than illegal trade of endangered species, wildlife poaching, and cruelty to wild and domesticated animals, falsification of wildlife products for economic gain is also a part of criminal activities and prohibited under the law. Poisoning and persecution of species, trade of endangered species such as tiger products for Oriental or Chinese medicine, and poaching of protected species such as rhino for the horn are common examples of wildlife crime. Thus, demand for the various parts and products derived from wildlife is a major cause for illegal hunting, which is a severe issue for wildlife management. It becomes more challenging when wildlife products are sold by masking its identity on common trading platform or online marketing [38]. Ornamental, superstitious use and other traditional uses of the body part of wildlife are the leading challenges in control of wildlife crime. Identification of the species from the available biological material is an effective method for fixing the accurate wildlife offense. Wildlife forensics has evolved as an essential branch of forensic science, which deals with the identification of the species from available biological remnants. Wildlife forensics has grown very rapidly in the past decade and has enormous scope in biodiversity conservation. When the morphological feature is available in the seized material, the examination of the morphological characteristics of the sized biological material is the rapid method of the delivering the result. Morphometry involves the comparison of morphological features of hair [3, 5, 6, 19, 25, 29, 30] and morphological osteology [28] with known reference specimens for validation of the source of the origin of unidentified case samples. Hence, morphological characterization has some extent of limitation in examining the unknown sample without intact reference material from the known origin.

5.2 Hair Morphology

Hairs are classified into primary guard hairs, secondary guard hairs, and fur hairs or underfur and over hair. Guard hairs are the largest, straightest, and robust among all the hair types. It is also called shield hairs [5] and divided into the primary and secondary types. In both primary and secondary guard hairs, the basal region is thinner and more flexible than the medial and distal regions. The secondary guard hairs are slightly shorter in length than the primary guard hairs. The larger or primary guard hairs are possessing the species-specific feature and most useful in species identification. Underhairs are shorter, softer, and superior than the guard hairs. Their thickness is almost uniform along the length. They have poor diagnostic value and are not preferred in species identification [5]. They required high-resolution microscopy for study, e.g., SEM. Over hairs are longer than the other [5]. They mainly exhibit a circular cross section and are not used in species identification. Vibrissae are rigid, large, and sensory in function. They are also known as whiskers, tactile hairs, or sinus hairs [5]. They are widest at the basal region and gradually taper toward the tip. Bristle hairs are found in some mammals such as the domestic and wild pigs; the hairs are rigid, thick, and uniform in diameter throughout the length. They contain a narrow medulla and are identified by the presence of bifurcated or trifurcated tips.

5.3 Hair Characteristics for Species Identification

Morphological feature such as the microscopic hair character has been widely used in the biological sciences for dietary habits and prey-predator relationships and in the identification of mammals inhabiting in a den or trees [25, 29, 30, 32]. It has also been used in dealing with wildlife offense cases such as those involving hide with fur [16], shahtoosh shawl made of Tibetan antelope's wool [1, 3], and paint brushes made of the hair of mongooses [7] and cats [6]. Hair for species identification has been used widely for more than one century [18–22, 31, 37]. Hausman [19–21] has presented written descriptions and simplified drawings on the configuration of cuticular scales and medullae for different hairs of 166 mammals. Mathiak [29] demonstrated a simple method for the identification of hairs of the mammals. Williams [43] described a protocol for the identification of hairs of moles and shrews. A manual was published on the animal furs of industrial importance [42] and described their origin and identification. Studies on the hair structure of Tasmanian marsupials and monotremes were carried out by Lyne and McMohan [27]. It indicated that a detailed microscopic examination of hair of these mammals would be of taxonomic value. Several studies were also conducted on the hairs of Indian mammals [6, 25, 33, 36]. Therefore, a need for a rapid and accurate method of identifying species from hair became evident.

Due to the variation of the cuticular pattern along the length of a hair and among hairs from different body parts, a strategy for identification of species on the basis of single characteristic of a hair could not be developed [2, 5, 25, 30, 35]. Therefore, an inclusive approach to combine all the characteristics, along with the external features, was developed for the identification of species. It is essential to conclude the investigation report based on the microscopic analysis, and there is no shortcut to reach a conclusion. In one case, it created identification problem when a veterinary expert from a veterinary medical college concluded the species based on the visible morphological feature. It was resolved by using DNA typing [16].

5.4 Setting Up of Infrastructure

Microscopic examination of hair for species testing can be performed using simple equipment. The main equipment required is a good compound microscope with a magnification range from 100X to 400X. If 1000X magnification can be added to the microscope, it can be utilized for the examination of underfur too. The microscope is necessarily required to be fitted with a digital/CCD camera and a computer with color printer. Hair comparison depends on the comparison of a large number of images; it is imperative that the computer should have a good capacity of the graphics card. Other than the microscope, microscopic slides, cover slips, tweezers, glassware (Petri dishes, beakers, glass rods), gelatin, and methylene blue are also required for preparing hair slide.

5.5 DNA Wildlife Forensics

In wildlife crime investigation, DNA typing is one of the vital protocols for molecular forensics, which is also known as DNA wildlife forensics [9, 12, 16]. The DNA typing encompasses some necessary basic steps, e.g., the DNA extraction, PCR amplification, DNA sequencing, and sequence verification. The different regions found on mitochondrial DNA (mtDNA) are the foundation of efficient molecular markers for phylogenetics and wildlife forensics. The mtDNA can be considered as the smallest chromosome coding for only 37 genes and consists of only about 16,600 base pairs. In most species, it is inherited from the mother. It represents a mainstay of phylogenetics, which allows biologists to confer the evolutionary history among species. Unlike nuclear DNA, the mtDNA is inherited from mother to offspring and is not involved in crossing over phenomenon; therefore, it is one of the unadulterated form of the DNA and has several conserved regions in its gene; these are the basis for the selection of conserved primers [17, 24, 26, 34, 39, 41], since conserved primers can be used in PCR amplification from the template DNA of majority of the species without prior information of the victim animal. Hence, the use of universal primers minimizes the effort of the researcher. The widely deployed mtDNA markers with descending order of similarity are 12S rDNA, 16S rDNA, cytochrome b, and control region (CR); hence, the 12S rRNA gene is highly conserved, and the CR is highly variable. The DNA sequence generated from the case property can be compared with the suitable database to get the exact identity of the source species. The gender-specific markers are also useful in dealing wildlife offense to identify the sex of the decomposed carcass in predicting the cause of the death of the individual as in cases of an Asian elephant, *Elephas maximus* [10]. Sometimes the species-specific primers can also be used for the confirmation of the species if prime suspect species is likely to be known [12]

5.6 Application of DNA Typing in Wildlife Forensics

The DNA typing is successfully used in wildlife crime investigation on routine basis. Due to the presence of multiple copies of mitochondria in a single cell, it becomes an extremely sensitive procedure, and hence it can easily be amplified from minute quantity of degraded biological samples. This protocol was helpful in unambiguous identification of the imperceptible biological samples collected from the wooden chopping board [9]. Furthermore, the accused tried to mislead the investigating team by artificially planting the samples of permitted domestic chicken (*Gallus gallus*) in the crime scene. The crime could be established by analysis of the DNA sequences obtained from the minuscule biological remnant from chopping wood [9]. Finally, DNA typing indicated that the biological remnant was of a peafowl (*Pavo cristatus*), which is listed as a Schedule-I protected species in the Wildlife (Protection) Act, 1973, of India.

Fig. 5.1 Suspected ivory idol recovered from handicraft shop

In other interesting cases, two suspected ivory idols were examined for the identification of source of the origin (Fig. 5.1). The identification was established by using species-specific primer developed for elephant [10]. Finally, comparison of DNA sequence confirmed that idols originated from Asian elephant, which has proven significant role of species-specific markers in wildlife forensics [12]. Species-specific marker has also been used in the identification of man-eater carnivore. In such cases, human-specific amplification was obtained from the fecal sample of suspected man-eater tiger/carnivore [11].

DNA typing has also been used in the identification of the species from the processed and dried internal organ of Himalayan black bear (*Selenarctos thibetanus*) and American Beaver (*Castor canadensis*). Several fake wild objects like tiger claw and skin were also traded in the market. Such a fake items can easily be differentiated from the original items by using DNA typing. Species of deer from the antlers are also identified with the help of DNA analysis [15]. However, in present scenario, it has constrained in identifying the species from the product made up of tanned skin including the skin of snake, which can be examined by morphological characteristics like cuticular feature, etc.

5.6.1 Application of Wildlife DNA Fingerprinting in Dealing Wildlife Offense

Microsatellite markers commonly known as short tandem repeats (STR) or simple sequence repeats (SSR) are noncoding repetitive DNA regions. It is a small motif of 2–6 nucleotide repeat in the genome of eukaryote and prokaryote [8, 40]. It is a codominant marker and used in kinship, population, and other studies like marker-assisted selection and fingerprinting. The main reason for the variability in microsatellites is due to a higher rate of mutation than the other neutral regions of genome. Mutation is also incorporated in the STR region during recombination phenomenon at the stage of meiosis [4]. Genomic microsatellite distributions are associated with sites of recombination as a consequence of repetitive sequences [23].

The use of microsatellite analysis comes into popularity in the mid-1990s for forensic investigation. It is widely used for the genetic profiling which is commonly referred as DNA fingerprinting of individuals. Shorter repeats frequently suffer from PCR stutter, PCR artifacts, and selective allele amplification. On other hand, longer repeats suffer from degradation and difficulties in amplification during PCR. Hence, the best suitable microsatellites used in forensic analysis are tri- to pentanucleotide (3–5 nucleotides) repeats, as these give an error-free data and can survive from degradation.

The DNA fingerprint of an individual/deceased can be compared with the biological relatives. Therefore, it has huge potential in dealing wildlife forensics. For human assassination, it has been practiced since its introduction to the forensic science. Subsequentally, it has been introduced for the wildlife crime investigation. A tiger (*Panthera tigris*) was killed in a zoological garden of India in the year 2000, and its skin and claws were taken away by the poachers. No biological sample was collected from the deceased tigress. Later on, in the year 2005, a claw in a pendant was recovered from a local boy. The multilocus genotype profile of the claw was matched with the alleged biological relatives of the deceased tigress. The genetic matching through computer-assisted analysis indicated that the claw was of a missing individual of the suspected family [13].

It has also assisted in scientific wildlife management. In a patrolling, an elephant tusk was found in a forest. It was collected and stored by forest staff [14]. Subsequently, forest official found one carcass of a male elephant with one missing tusk. It was suspected to be a hunting case for ivory trade. Consequently, small pieces of the ivory collected from both the tusks were examined for individual matching. The DNA profiling using microsatellite loci indicated that the ivory samples collected from two different incidences were of the same animal. It helped in proper wildlife management [14].

5.7 Scope of Further Improvements

Lack of robust database for wildlife is a major limiting factor in dealing wildlife forensics [17]. It can be supplemented by the generation of a reliable database for all range of the wild species through research and development for strengthening DNA database for endangered species [17]. Collected biological samples from a known carcass of wildlife from the reserve forest need to be utilized for such research and development; hence, it would also require the effort of forest officials including the patrolling party for routine collection of the known biological samples.

Due to the frequent mistake committed by the enforcement agencies, for proper guideline, this section has been divided into two parts, viz., what not to be done and what needs to be done.

5.7.1 What Not to Be Done?

Collection, preservation, and forwarding of biological samples are the most important procedures for forensic examination. Due to lack of proper knowledge of collection and storage of sample, enforcement agencies unknowingly commit some common mistake. As the seized sample likely to be of an animal species, these agencies contact to a local veterinary officer for the preservation and packaging of detained biological samples. It was often observed that veterinary expert preserves such biological samples in formaldehyde/formalin. It has been found that formalin-/formaldehyde-preserved sample is not suitable for the extraction of 400–500 bp PCR amplifiable quality of DNA; therefore, preservation of sample in formalin should be avoided. Several DNA forensic laboratories do not accept any biological sample preserved in or exposed to formaldehyde.

5.7.2 What Needs to Be Done?

If complete hide/skin or major bone including skull is recovered in a wildlife offense case, the same should be forwarded to the laboratory for the forensic analysis. If an excess of meat is recovered, only 10–30 g of biological samples is required to be packed for genetic analysis that can be stored in a vial of 50–100 ml capacity. The packaging of the meat samples should be as compact as possible. Ethyl alcohol (C_2H_5OH) is recommended as the best source for the preservation of meat sample for DNA forensics. The sample can be stored at room temperature for several years,

if it is stored in 70% to absolute ethyl alcohol in an airtight container. Alternatively, samples can also be stored in silica gel. However, it can degrade in the silica gel over a period; therefore, it needs a quick transportation to the laboratory within few days to week preferably by hand or by speed post. The sample can also be transported in cold chain either in the presence of dry ice, ice, or with coolant gel. If only blood (whole blood or dried stain) is recovered from the crime scene; it need not to be stored in alcohol but dried stain can be stored in silica gel or can be transported in dried form and whole blood in a blood collection vial following a cold chain.

5.8 Quick Guide for Sample Collection

It would be difficult to arrange all the essentials for the sample collection when a crime is detected. Hence, to avoid last-minute blunder, a sample collection kit can be kept ready for handling of crime scene and for the collection of biological evident. The component of a collection kit is as follows:

Surgical gloves, 2 pairs; zip lock plastic pouch, 4 no; screw-capped vial (50ml for meat/scat/skin), 4 no; screw-capped vial (5ml for blood), 4 no; injection syringe (5ml), 1 no; scalpel handle, 1 no; surgical blade, 4 no; forceps, 1 no; scissors (4 inch), 1 pair; slide case, 1 no; glass slide, 4 no; silica gel, 20 g; filter paper, 12 no; measuring tape (60inch), 1 no; paraffin roll (for sealing of vial); cello tape (1inch), 1 no; marker pen, 1 no; one bottle of ethyl alcohol; one pack of silica gel (5–8 mesh size); cotton cloth, stitching needle, thread; sealing wax, candle, and matchbox; and official metal seal

5.8.1 Stepwise Procedure for the Collection of Meat/Skin Piece/Scat

- Take a screw-capped vial.
- Fill half of the vial with ethyl alcohol.
- Dip a small piece of meat (10–20 g)/skin piece-3 × 3 cms (15–20 g) in the alcohol.
- Make airtight with the cap and seal with paraffin roll and cello tape.
- Write the species, place of collection, and date of sample on the vial. Put a cello tape on the writing to avoid erasure from alcohol spill.
- Stitch in cotton cloth and apply sealing wax and official seal to make it temper proof.

5.8.2 *Stepwise Procedure for the Collection of Blood*

- Take 2–5 ml blood collection vials.
- Inject 2–3 ml of blood in to the vial.
- Write the species, place of collection, and date of sample on the vial.
- For making smear, use the glass slide. Put a drop of blood on one slide.
- Use the edge of the other slide to smear the blood on the slide.
- Using marker pen, write necessary information on one side of the slide as mentioned above.

Now biological samples are ready for the forwarding to a forensic laboratory. A copy of all related case document, e.g., seizure memo, first information report (FIR), and sample seal, should also be supplied along with the forwarding letter of investigating officer or judicial magistrate.

5.9 Conclusion

Illegal wildlife trade has been detected as a global threat to the sustainability of biodiversity and the security of society. It invited the strict wildlife protection laws. In the past two decades, forensic science has evolved rapidly with its emerging branch "wildlife forensics" with great success. Investigation of wildlife crime is increasingly supporting the implementation of wildlife protection laws. Among the two methods, viz., morphological examination and molecular analysis, the latter was proven as the most effective method for dealing these issues. The conserved primers have significant role in dealing the wildlife DNA forensics. These primers are derived from various conserved regions located on the genes of the mitochondrial genome. Thus, use of morphological examination, DNA typing, and genetic profiling is the need of the time to deal with the increasing wildlife offenses.

References

1. Anon (1995) Characterization of *Panthalops hodgsoni* fibres. Corpo Forestale Dello Stato–Italia, Servizio CITES, Provisional draft
2. Appleyard HM (1954) Guide to the identification of animal fibers. Wool Industries Research Association, Leeds
3. Bahuguna A, Mukherjee SK (2000) Use of SEM to recognise Tibetan antelope Chiru hair and blending in wool products. Sci Justice 40:177–182
4. Blouin MS, Parsons M, Lacaille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. Mol Ecol 5:393–401
5. Brunner H, Coman B (1974) The identification of mammalian hair. Shanghai Printing Press Ltd, Hong Kong
6. Chakraborty R, Chakraborty S, De JK (1999) Identification of dorsal guard hairs of the species of Indian lesser cats Carnivora: Felidae. Mammalia 65:93–104

7. De JK, Chakraborty S, Chakraborti R (1998) Identification of dorsal guard hairs of five Indian species of mongoose, *Herpestes* Illiger (Mammalia: Carnivora). Mammalia 62:285–295
8. Field D, Wills C (1998) Abundant microsatellite polymorphism in Saccharomyces cerevisiae, and the different distributions of microsatellites in eight prokaryotes and S. cerevisiae, result from strong mutation pressures and a variety of selective forces. Proc Natl Acad Sci USA 95:1647–1652
9. Gupta SK, Verma SK, Singh L (2005) Molecular insight into a wildlife crime: the case of a peafowl slaughter. Forensic Sci Int 154:214–217
10. Gupta SK, Thangaraj K, Singh L (2006) A simple and inexpensive molecular method for sexing and identification of the forensic samples of elephant origin. J Forensic Sci 51:805–807
11. Gupta SK, Kumar A (2014) Molecular identification of man-eating carnivores from scat samples. Conserv Genet Resour 6:271–274
12. Gupta SK, Thangaraj K, Singh L (2011a) Identification of the source of ivory idol by DNA analysis. J Forensic Sci 56:1343–1345
13. Gupta SK, Bhagavatula J, Thangaraj K, Singh L (2011b) Establishing the identity of the massacred tigress in a case of wildlife crime. Forensic Sci Int Genet 5:73–74
14. Gupta SK, Kumar A, Nigam P (2011c) Multilocus genotyping of Asian elephant ivory: a case study in suspected wildlife crime. J Forensic Res S1:002. https://doi.org/10.4172/2157-7145.S1-002
15. Gupta SK, Kumar A, Hussain SA (2013) Extraction of PCR-amplifiable DNA from a variety of biological samples with uniform success rate. Conserv Genet Resour 5:215–217
16. Gupta SK, Sharma CP, Singh L (2014a) DNA typing established as an unambiguous tool for species identification in a dispute case. Forensic Sci J 13:9–14
17. Gupta SK, Kumar A, Hussain SA (2014b) Novel primers for sequencing of the complete mitochondrial cytochrome *b* gene of ungulates using non-invasive and degraded biological samples. Conserv Genet Resour 6:499–501
18. Hardy JI, Plitt P (1940) An improved method for revealing the surface structure of fur fibers. U.S. Department, International Fish and Wildlife Service, Washington, DC, Wildlife Circle 7
19. Hausman LA (1920) Structural characteristics of the hair of mammals. Am Nat 54:496–526
20. Hausman LA (1924) Further studies of the relationship of the structural characters of mammalian hair. Am Nat, New York 58:544–557
21. Hausman LA (1930) Recent studies of hair structure relationships. Sci Mon 30:258–277
22. Herzog (1927) Abdrucke tierischer Woolen und Haare in Harz. Melliand Textil Ber 6:341
23. Huang W, Si X, Hou Y, Miller-Antonio S, Schepartz LA (1995) Excavations at Panxian Dadong, Guizhou Province, southern China. Curr Anthropol 36:844–846
24. Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86:6196–6200
25. Koppikar BR, Sabnis JH (1975) Identification of hairs of some Indian mammals. J Bomb Nat Hist Soci 7:35–20
26. Kumar A, Hussain SA, Bhatt D, Gupta SK (2017) Conserve primers for sequencing complete ungulate mitochondrial cytochrome c oxidase *I* (CO*I*) gene from problematic and decomposed biological samples. Mitochondrial DNA Part B 2:64–66
27. Lyne AG, McMohan TS (1951) Observations on the surface structure of the hairs of Tasmanian monotremes and marsupials. Pap R Soc Tasman 1950:71–84
28. Margaret ES, Barry WB, Robert MH (2011) Tusk or bone? An example of ivory substitute in the wildlife trade. Ethnobiol Lett 2:40–44
29. Mathiak HA (1938) A key to the hairs of mammals of southern Michigan. J Wildl Mgmt 2:251–268
30. Mayer WVC (1952) The hair of Californian mammals with keys to the dorsal guard hairs of Californian mammals. Am Midl Nat 48:480–512
31. McMurtrie W (1886) Report upon an examination of wools and other animal fibers. U.S. Department Agriculture, Washington, DC, 613p.+illus

32. Mukherjee S, Goyal SP, Chellam R (1994) Refined techniques for the analysis of Asiatic lion *Panthera leo persica* scats. Acta Theriol 39:425–430
33. Nath S, Joseph (1981) Preparation of a key for identification of animals by the structure of their hair. FRI, 16pp. 3rd diploma course, Wildlife Institute of India, Dehradun, India
34. Parson W, Pegoraro K, Neiderstatter H, Foger M, Steinlechner M (2000) Species identification by means of cytochrome b gene. Int J Legal Med 114:23–28
35. Perrin MR, Campbell BS (1980) Key to the mammals of the Andries Vosloo Kudu Reserve (eastern cape), based on their hair morphology, for use in predator scat analysis. S Afr Tydskr Natuurnav 10(1):1–13
36. Sahajpal V, Goyal SP (2006) Microscopic hair characteristics: a tool for dealing wildlife offences in India. Annual meeting of the European hair research society (EHRS). London, UK
37. Saxinger G (1925) Eine neue methods zur Untersuchung des haarepithels Oberhautchens. Ztschr f Tiersucht U Zuchtungsbiol 5:379–388
38. Sharma CP, Kumar A, Vipin SV, Singh B, Kumar GC, Gupta SK (2018) Online selling of wildlife part with spurious name: a serious challenge for wildlife crime enforcement. Int J Legal Med, (In Press)
39. Singh A, Kumar A, Kumar RS, Bhatt D, Gupta SK (2017) Amplification of mtDNA control region in opportunistically collected bird samples belonging to nine families of the order Passeriformes. Mitochondrial DNA Part B 2:99–100
40. Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967–981
41. Verma SK, Singh L (2003) Novel universal primers establish identity of enormous number of animal species for forensic application. Mol Ecol Notes 3:28–31
42. Wildman AB (1940) Animal fibres of industrial importance: their origin and identification. Wool Industries Research Association, Leeds
43. Williams C S (1938) Aids to the identification of mole and shrew hairs with general comments on hair structure and hair determination. J Wildlife Mgmt 2:239–250

Chapter 6
Species Characterisation from Hair of Protected Mammals: Comparison of Molecular Methods

Vivek Sahajpal and S. P. Goyal

Abstract Human interference has increased exploitation of the wildlife subsequently leading to their extinction. India being a mega-biodiversity nation faces most of the heat due to rampant poaching and growing illegal wildlife trade. Hence it is very crucial to develop techniques which can characterise the species from artefacts seized under the wildlife laws. In this regard, keratin protein possesses a huge promise for its utility in species designation. Most of the land mammals produce keratin in either of the two forms, i.e. hair keratin or claw, bill, nail, hoof, etc. Keratins are considered to be the most heterogenous types of proteins after immunoglobulins due to their multiple gene loci coding. Thus, as synthesis of keratins is under genetic control and they are known to be polymorphic, the analysis of these proteins can be used for species identification; differentiating varieties, species, subspecies and even breeds; and perhaps even individualisation. In this chapter, the hair keratin protein of 20 selected artiodactyls species has been characterised to study their usability in species designation as well as application in wildlife forensics.

Keywords Keratin · Protein · Species designation · Wildlife · Forensics · Identification

6.1 Introduction

Excessive exploitation of the wildlife and natural resources by the mankind for amenities like eatables, clothes, trade, pharmaceuticals and aesthetic purposes has been present for centuries, but lately, the profit-centred manipulations have reached many species to the edge of extinction [44]. The unending trade in matters of wildlife and its products has made their conservation a major challenge [44]. The illicit trade in wildlife amounts to US $ 5 billion of the total $20 billion annual

V. Sahajpal (✉)
State Forensic Science Laboratory, Junga, Himachal Pradesh, India

S. P. Goyal
Wildlife Institute of India, Dehradun, Uttrakhand, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_6

international trade in wildlife and its products and fiscally stands second only to drugs [15]. It is tough to offer precise approximation about the volume of the illicit trade as large proportion is covert and the exact data maybe surprisingly elevated.

India has 8% of the world's total biodiversity consisting 60% of total tigers, 50% of Asian elephants, 70% of Asian rhinos and the last Asiatic lions in the wild, and this identifies India as one of the 12 mega-biodiversity nations [15]. There is an amalgam of Palaearctic, Oriental and Afrotropical fauna in India [37] having more than 400 species of mammals [45]. Widespread poaching and rising illicit wildlife trade have put us in the danger of being deprived of the wildlife.

In India the Wildlife (Protection) Act, 1972, protects the wildlife. The animals scheduled in I, II, III and IV schedules of the Wildlife (Protection) Act, 1972, are safeguarded, and this act amounts poaching as a punishable offence. India is a part of the *Convention on International Trade in Endangered Species* (CITES). Such strong legislation has failed to curb the humongous increase in poaching which has resulted in reducing the population of numerous wildlife species to a frightening low level. The conviction rate is extremely low in most poaching cases. Annon [2] held the lack of proper species characterisation and identification techniques from the seized items responsible for this. The lack of technical know-how, database and expertise in the wildlife scenario makes it impossible to know the species of the poached animal, which is of utmost importance for prosecuting a person under the wildlife protection laws. The task of identifying the finished products complicates the problem, and the fake products worsen the situation even more. Hence it is very crucial to develop techniques which can characterise the species from artefacts seized under the wildlife laws.

Mammals form a large proportion of poached animal species, and hair is consistently found as the physical evidence in most of the cases of mammalian poaching. On an average, evidences from hundreds of wildlife offence cases are received annually in the wildlife forensic laboratory at the Wildlife Institute of India, Dehradun, for species identification. About 50–60% of the total cases received are concerned with species identification using hair.

Hair, as a physical evidence, is well accepted and appreciated in criminal investigations [55]. Additionally, because of its chemical composition, hair resists degradation due to environmental and enzymatic factors. Hair can be used for species identification using the morphology and keratin structure. Additionally it also contains nuclear (when root is present) and mitochondrial DNA of an individual. Hence, with the help of morphometric and molecular methods, hair can be used for species identification.

The examination of physical properties of hair has been routinely used in forensic science for species identification; however, this method is often questionable especially when the sample is in very small amount, and it is always better to support and compliment the identification of species with other techniques. Gillespie and Marshall [21] observed that a variety of hair samples with different structures can be produced by an individual which, in the long run, further gets affected by the environmental factors. As observed in forensic case scenario, there is no control over quantity and quality of samples; it is worthwhile to focus on molecular meth-

ods for species identification from hair. Protein methods were regularly used in forensic science for species identification and individualisation before the dawn of DNA technology. Keratin profiling was advocated for species identification against microscopic hair characterisation by workers like Carracedo et al. [10].

Keratins are group of tissue proteins of ectodermal origin which are resistant to digestion by trypsin or pepsin and are insoluble in water, organic solvents and weak acids and bases [5]. Most of the land mammals produce keratins in two hard forms (1) hair keratin and (2) claw, bill, nail, hoof etc., and sometimes in the third form, a quill or a horn may be produced. Keratin is not a singular mixture but is actually a complex mixture of proteins and the sulphur-containing diamino acid cystine, which distinguishes it from collagen, elastin and myofibrillar proteins [8]. Possibly, keratins are the most diverse proteins after immunoglobulins. While several explanations for the heterogeneity have been given, multiple gene loci coding for keratins is the most likely reason [27]. As the synthesis of keratins is under genetic control and because keratins are known to be polymorphic, the analysis of these proteins can be used for species identification. Different species, subspecies, varieties and breeds and even individuals of the same species have considerable differences [8]. Although the keratins are insoluble in water, they can be solubilised using certain reducing agent-containing extraction buffer. They can then be separated by electrophoresis to get keratin patterns for species characterisation. The most specific way to achieve solubilisation is to break the disulphide bonds by reducing with reagents such as mercaptoethanol or dithiothreitol.

The efficacy of keratin profiling in species identification has been demonstrated by various workers: Budowle and Acton [7], Marshall [38], Marshall and Gillespie [40], Marshall [39], Marshall et al. [41], Carracedo et al. [9, 10], Butler et al. [8], Rodriguez-calvo et al. [52], Folin and Contiero [19, 20] and Sahajpal and Goyal [53], Espinoza et al. [17] and Paolella et al. [47].

Keratin-based methods, although very useful for characterisation of species, have certain limitations too. They cannot be used where the sample is minimal, highly degraded or very old. The histological origin of the tissue sample also has a bearing on the results of protein techniques. Hair samples are much more stable than other soft tissues as they are formed of keratin, but still they get affected by the environmental factors and slowly show degradation, and hence the obtained protein profiles have alterations. Therefore, the use of DNA-based techniques is very crucial to deal with wildlife offence cases considering their qualities like high sensitivity, consistent and precise results and the ability to be reproduced.

The commonly used approaches for species identification are RFLP (restriction fragment length polymorphism), PCR-RFLP (polymerase chain reaction- restriction fragment length polymorphism), RAPD (random amplified polymorphic DNA) and FINS (forensically informative nucleotide sequencing). RFLP (restriction fragment length polymorphism) is centred on the fact that different individuals and species have different nucleotide sequences, and if the DNA is treated with restriction, enzymes which target specific region will cut the DNA duplex into fragments which will essentially be polymorphic. They will be different for species [6], and when these fragments are resolved according to their size on a gel (agarose/PAGE),

they will produce distinct RFLP pattern for a species and even up to individual level. Most of the techniques used these days for species identification use polymerase chain reaction and are based on DNA. Conventional RFLP requires good quality genomic DNA which is very rare as in the case of hair samples, especially in forensically important hair samples; hence this technique is not of much use in species identification from hair samples.

PCR-RFLP makes the use of PCR to amplify a conserved gene of the species under scrutiny (mitochondrial cytochrome b, 16s rRNA, etc.) with the help of a universal primer to produce amplicons of an approximately same size. These amplicons are digested with specific restriction enzymes targeting specific restriction sites which are located at different positions in different species and hence producing fragments of different sizes (length polymorphism) for different species. These fragments upon resolving on the basis of their size, on a gel (agarose/PAGE) will produce distinct RFLP patterns specific for a species. Efficacy of PCR-RFLP has been demonstrated by various workers: Meyer et al. [42], Sironi et al. [59], Fairbrother et al. [18], Carrera et al. [12], Wolf et al. [63], Girish et al. [23], Malisa et al. [36], Chakraborty et al. [13], Khamnamtong et al. [31] and Akasaki et al. [1].

The other commonly used PCR-based technique is the random amplified polymorphic DNA (RAPD). This technique is based on random amplification of genomic DNA which uses a single short oligonucleotide primer that consists of arbitrary sequences. The PCS using this technique can generate several products which are species specific in a PCR, which can generate several species-specific products [25]. RAPD, although a very effective technique for species characterisation/identification, is not effective where the amount of DNA is very low. As RAPD requires good quality genomic DNA which is very rare in case of hair samples, this technique is not of much use in species characterisation from hair samples obtained as physical evidence in wildlife offence cases.

Perhaps the most talked about and useful technique in case of species identification from hair samples is *forensically informative nucleotide sequencing* (FINS) [3, 4]. This technique involves amplification of a species-specific, conserved gene (mitochondrial cytochrome b, 12s rRNA, 16s rRNA, D-loop, cytochrome oxidase I, etc.) using a universal primer set and its subsequent sequencing to obtain the sequence of the amplicon. BLAST (basic locus alignment search tool) is used to compare the sequence with the database of the National Center for Biotechnology Information (GenBank) for species identification or by aligning with sequences available for known species to calculate the similarity. Most of the mitochondrial genes which are conserved for a species can be used for FINS study. Extensive work has been done on mitochondrial cytochrome b gene by Parson et al. [48], Linacre and Lee [35], Pepe et al. [49] and Domingo-Roura et al. [16]. Additionally, 16s rRNA, 12s rRNA and D-loop have also been extensively used for species characterisation. Lee et al. [34] developed a nested PCR-based DNA sequencing for identifi-

cation of shahtoosh (*Pantholops hodgsonii*). The utility of FINS has also been demonstrated by Sahajpal and Goyal [54] for identifying species from a forensic sample. The limitations of conventional techniques like microscopy and usefullness of DNA techniques in species identifiation have been discussed by Sahajpal et al. [56], Zafrina and Panneerchelvam [64] and Pilli et al. [51]. Similarly, barcoding genes, like the mitochondrial cytochrome oxidase I (COI), can also be used for species characterisation [14].

In addition to FINS, PCR-based assays have also been reported for species identification. Guha and Kashyap [25] developed a hemi-nested PCR assay which is based on 16S rRNAgene of mitochondria for characterisation of some species. However, the PCR-based assays are useful when the spectrum of species to be identified is narrow. PCR assay-based techniques work well for species identification in meat samples where only a few species are used. However, in case of wildlife crime where a very wide spectrum of species is in trade, this technique does not hold much of a promise. FINS is a much better technique as compared to PCR-based assays as it can help to narrow down the search to one or few species depending upon the availability of the sequence database for suspected species. It is especially useful in wildlife forensics where a wide spectrum of illegally poached and traded species exists. Further, sequencing also helps to build up nucleotide sequence database which has a great relevance in phylogenetic analysis. FINS further has the advantage over other techniques like RAPD as it requires very little amount of DNA and hence is actually relevant to the forensic case scenario. In addition the nucleotide sequences obtained by the sequencing can be used to select restriction enzymes for restriction digestion of amplified fragments in PCR-RFLP technique. In the absence of such sequence database, restriction digestion will be just a hit and trial method to find out and appropriate restriction enzyme to study restriction fragment length polymorphism amongst species of interest. However, if sequencing is done, the sequence can be used to select restriction enzymes by locating absence or presence of enzyme-specific restriction sites. Thus FINS also helps in finding our restriction fragment length polymorphism. Hence, the sequence database developed for some species can be used to develop simple PCR-RFLP-based methods for species identification.

As mentioned earlier, in forensics there is not control over the quantity and quality of the samples received from analysis in crime cases. It is hence imperative that the techniques should be thoroughly optimised to obtain results from such minimal samples, and the best method essentially is to carry out a comparative study on various types of known samples prior to subjecting case exhibits to analysis. A comparative study of molecular methods for species characterisation was carried out on selected species listed under Schedule I of the Wildlife Protection Act, 1972. The list of species is given in Table 6.1. Out of the techniques discussed, keratin profiling (SDS-PAGE and IEF), PCR-RFLF and FINS were used to characterise species from hair sample of the selected species.

Table 6.1 Schedule I, artiodactyls of Indian Wildlife Protection Act, 1972

Sl. no.	Species	Common name
1	*Pseudois nayaur*	Bharal, blue sheep
2	*Antilope cervicapra*	Black buck
3	*Cervus eldii thamin*	Thamin or brow-antlered deer
4	*Gazella bennetti*	Chinkara
5	*Tetracerus quadricornis*	Four-horned antelope
6	*Bos frontalis*	Gaur or the Indian bison
7	*Capra ibex*	Himalayan ibex
8	*Tragulus meminna*	Mouse deer
9	*Moschus moschiferus*	Musk deer
10	*Hemitragus hylocrius*	Nilgiri tahr
11	*Ovis ammon hodgsonii*	Nyan, argali or the great Tibetan sheep
12	*Capricornis sumatraensis*	Serow
13	*Cervus duvauceli*	Swamp deer
14	*Budorcas taxicolor*	Takin
15	*Pantholops hodgsonii*	Chiru or Tibetan antelope
16	*Procapra picticaudata*	Tibetan gazelle
17	*Ovis vignei vignei*	Ladakh urial or shapu
18	*Bubalus bubalis*	Wild buffalo
19	*Bos grunniens*	Wild yak
20	*Hemitragus jemlahicus*	Himalayan tahr

6.2 Materials and Methods

6.2.1 Collection of Hair Sample

The reference collection of the Wildlife Forensic Laboratory of the Wildlife Institute of India was used to collect the reference samples of hair for the 20 selected artiodactyl species (from three to five individuals per species). Some samples were also obtained from other sources like ZSI (Zoological Survey of India) and BNHS (Bombay Natural History Society).

6.2.2 Keratin Profiling Techniques

6.2.2.1 Cleaning of Hair Samples

The hairs were washed with Milli-Q water to remove the dirt from the hair surface, and the inorganic dirt was removed using a mild detergent. They were rinsed with Milli-Q water for about 4–5 min, once or more depending upon the amount of dirt. After rinsing with water, carbon tetrachloride or isopropyl alcohol and other organic solvents were used to remove the organic dirt like grease oil and wax, from the hair surface. Samples were dried prior to subjecting them to extraction.

6.2.2.2 Buffers and Solutions

- The required buffers and solutions were prepared mol bio-grade reagents. Extraction buffer: 3.6 g urea, 0.045 g of Tris(hydroxymethyl)aminomethane and 60 mg of Cleland's reagent/dithiothreitol (DTT) in 4.8 ml of Milli-Q water (prepared freshly).
- Extraction buffer (SDS-PAGE specific): [30 mM Tris-HCl, 50 mM DTT, 8 M urea and 2% SDS].
- Sample buffer (for SDS-PAGE): 5.0 mg of Tris-HCl, 0.5 g of SDS and 5.0 g of sucrose in 10 ml of Milli-Q or double-distilled water.
- 30% acrylamide solution: (29.2 g acrylamide, 0.8 g bis-acrylamide, in 100 ml deionised water)
- Separating gel buffer (Laemmli's gel system): 1.875 M Tris-HCl (pH -8.8).
- Stacking gel buffer 0.6 M Tris-HCl (pH -6.8).
- Running buffer (Laemmli's gel system): (3 g Trizma base, 14.4 g glycine, 0.5 g SDS).
- Separating/spacer buffer (Schagger and Von gel system): 3 M Trizma base, 0.3% SDS, distilled H_2O 100 ml (pH 8.9 adjusted with HCl).
- Stacking gel buffer (Schagger and Von gel system): 1 M Tris (12.114 gm in 100 ml Milli-Q or double-distilled water (pH 6.8 adjusted with HCl).
- 10x Cathode buffer (Schagger and Von gel system): 1% SDS, 1 M Trizma base, 1 M glycine
- 10x Anode buffer (Schagger and Von gel system): 2 M Trizma base (pH 8.9 adjusted with HCl)
- Separating gel acrylamide, 48% (2X) acrylamide (Schagger and Von gel system): 48 g acrylamide, 3 g N,N-methylenebis-acrylamide, Milli-Q or distilled H_2O 100 ml (48% acrylamide solution should be wrapped in foil and stored at 4 °C).
- Separating gel acrylamide, 48% (1X) acrylamide (Schagger and Von gel system): 48 g acrylamide, 1.5 g, distilled H_2O 100 ml (48% acrylamide solution should be wrapped in foil and stored at 4 °C).
- 10% SDS.
- 10% ammonium persulphate.
- N, N, N′, N′-tetramethylenediamine.
- 1-Butanol
- Fixing solution (IEF): 20% trichloroacetic acid in Milli-Q or double-distilled water.
- Destaining solution: methanol, acetic acid, water [50:10:40 v/v/v].
- Staining solution: Coomassie Brilliant Blue R-250 0.1 g in 100 ml of fixing and destaining solution.
- Sensitisation solution (sensitive silver staining): 10% glutaraldehyde solution.
- Staining solution (sensitive silver staining): 21 ml 0.36% NaOH solution, 1.4 ml 35% ammonia solution, 4 ml 20% $AgNO_3$.
- Developing solution (sensitive silver staining): 0.5 ml of1% citric acid solution 0.5 ml and 0.05 ml 37% formaldehyde solution.
- Stopping solution (sensitive silver staining): 40% ethanol, 10% acetic acid, made up to 100 ml.

6.2.2.3 Extraction of Keratins

Razor blade was used to chop the cleaned and dried hair samples into small pieces of 0.5–1.0 cm. The chopped hair samples were used for the extraction of keratins and keratin-associated proteins (KAPS) with help of a basic protocol laid down by Marshall [38], Marshall and Gillespie [40] and later modified by Carracedo et al. [9–11] by removing the step of carboxymethylation. The extraction was carried out with the following extraction buffer as laid down by Carracedo et al. [9] and Butler et al. [8].

Initially samples were weighed, and 10 mg of hair sample was used for extraction with 200 µl of the extraction buffer. Four to seven hair samples were also used for extraction with reduced amount of extraction buffer (50–100 µl). The extraction was carried out over a period of 48 h in 1.5 ml Eppendorf vials at a temperature of 25 °C, maintained through a water bath. As the process involves reduction of the disulphide bonds by DTT, it is a must that the extraction should be carried out under inert atmosphere. The extraction mixture was centrifuged briefly for 5 min at 14,000 rpm, and the clear supernatant was pipetted out into an autoclaved Eppendorf vial of 1.5 ml capacity. To about 25 µl of extract, 5 µl of 0.1 *M* DTT was added. These samples were directly used for isoelectric focusing. For SDS-PAGE analysis, the extracts were treated with a denaturing sample buffer containing sodium dodecyl sulphate (SDS) detergent as a denaturant. The sample buffer was prepared by dissolving 5.0 mg of Tris-HCl, 0.5 g of SDS and 5.0 g of sucrose in 10 ml of Milli-Q or double-distilled water.

In addition to the above protocol for extraction of keratin, a second SDS-PAGE (denaturing)-specific protocol was also used. In this protocol 5 mg of chopped hair sample was extracted with 200 µl of extraction buffer comprising of 30 mM Tris-HCl, 50 mM DTT, 8 M urea and 2% SDS for 30 h. In case of both the protocols, the extracts used for SDS-PAGE were denatured at 96 °C for 3–5 min prior to running electrophoresis on a vertical SDS-PAGE gel.

6.2.2.4 SDS-PAGE

In case of SDS-PAGE, two gel systems were tried. These include Laemmli's gel system [33] and modified Schagger and Von gel system [30, 57]. Laemmli's gel system is basically designed to separate out a wide range of protein samples, whereas the Schagger and Von gel system is basically designed to separate out low molecular weight peptides and is more efficient in separating peptides having very little difference in their molecular weights [30]. In case of Laemmli's gel system, varying concentrations of gels were tried to obtain an optimum concentration, and similarly, attempt was made to optimise running conditions like current, voltage and temperature. Similarly, attempt was made with Schagger and Von gel system for optimisation of the process of keratin profiling.

SDS-PAGE (Laemmli's Gel System)

Vertical SDS-PAGE was performed on Hoefer SE- 260 (0.5 mm × 10 × 10.5 cm gel) using Laemmli gel [33]. The lower 8.5 cm of the gel cassette was filled with separating/resolving gel. The upper 2 cm were filled with the stacking gel. A range of resolving gel concentration (10%–22%) was obtained by the use of the following gel recipes (Table 6.2).

In addition to the gel concentration, running current/voltage, running time, temperature, etc. were also standardised by using gradients of all these parameters to obtain an optimised protocol for the separation of keratins.

SDS-PAGE (Modified Schagger and Von Gel System)

Vertical SDS-PAGE was also performed with modified Schagger and Von gel system [30, 57]. Resolving gels of varying concentrations of acrylamide were tried to obtain the most optimum gel concentration for separation of hair keratin of the selected species. The gel was prepared according to the protocol given by Schagger and Von Jagow [57] and Judd [30]. The gel recipe is mentioned in Table 6.3.

Table 6.2 Gel recipes for preparing different concentrations of resolving and stacking gels (optimised for Hoefer SE 280 0.5 mm × 10 cm × 10.5 cm gel)

	Concentration of Gel					
Components	10%	15%	18%	20%	22%	Stacking gel
1.875 M Tris-HCL (pH 8.8)	4 ml	4 ml	4 ml	4 ml	4 ml	0 ml
0.6 M Tris-HCL (pH 6.8)	0 ml	0 ml	0 ml	0 ml	0 ml	1.0 ml
Water (Milli-Q)	9.05 ml	5.7 ml	3.7 ml	2.37 ml	1.035 ml	7.0 ml
Stock acrylamide	6.65 ml	10 ml	12 ml	13.3 ml	14.66 ml	2.0 ml
10% SDS	0.2 ml	0.2 ml	0.2 ml	0.2 ml	0.2 ml	0.1 ml
10% ammonium persulphate	0.1 ml	0.1 ml	0.1 ml	0.1 ml	0.1 ml	0.05 ml
TEMED	7.0 µl	7.0 µl	7.0 µl	7.0 µl	7.0 µl	7.0 µl

Table 6.3 Gel recipe for preparing Schagger and Von gel (optimised for Hoefer SE 280 0.5 mm × 10 cm × 10.5 cm gel)

Components	Resolving gel	Spacer gel	Stacking gel
Distilled H_2O	3.5 ml	3.45 ml	5.2 ml
Separating buffer	5.0 ml	2.5 ml	0 ml
Stacking gel buffer	0 ml	0 ml	0.95
Stock acrylamide 30%	0 ml	0 ml	1.5
ETDA	0 ml	0 ml	75 µl
48% separating acrylamide (1x and 2x)	5.0 ml	1.5 ml	0 ml
Glycerol	1.6 ml	0 ml	0 ml
APS	50.0 µl	25.0 µl	75.0 µl
TEMED	5.0 µl	2.5 µl	3.45 µl

The lower 7.5 cm of the cassette was filled with resolving gel, and the next 1 cm was used for filling the spacer gel. The upper 2 cm were filled with the stacking gel along with the comb for well formation. In addition to the above-mentioned recipe, variation on cross-linker concentration was also studied for the optimisation of the protocol for separation of keratins. The running voltage/current, temperature, etc. were also studied by running a gradient of conditions. For all the gels, a low molecular weight ladder (LMW) (Amersham Biosciences, Sweden) was run with the samples for calculation of molecular weight.

6.2.2.5 Isoelectric Focusing Using Automated (Phast™) System

The separation of keratins using isoelectric focusing was done on Phast™ System (Amersham Biosciences, Sweden). Standard precast gels (PhastGel™ Amersham Biosciences, Sweden) were used which show separation with high level of reproducibility. The samples after extraction and a brief centrifugation at 14,000 rpm for 5 min were used for IEF. The loading of samples is automated in Phast system. The extracts were loaded on a loading tray having eight wells. The samples were transferred onto the comb which has perforations, which can hold 0.25 µl of the sample. The comb was loaded onto the Phast system after positioning the gel according to the standard protocol. Isoelectric focusing was done by following the standard protocol (Separation Technique File 100), but the run time was increased to 500 volt hours instead of 410 volt hours mentioned in the standard protocol (Table 6.4). The run take is about 40 min for completion.

A reference broad pI calibration kit (Amersham Biosciences, Sweden) was run alongside the samples for calculation of pI (isoelectric point) values.

6.2.2.6 Staining, Imaging and Statistical Analysis

Coomassie staining, conventional silver staining and sensitive silver staining (silver diamine method) were tried for staining. The IEF gels were fixed in 20% trichloroacetic acid (TCA) solution for 20 min. Statistical analysis was done using software packages like Statistica and SPSS (Release 12.0.0). Similarity the index between the various species was studied, and cluster analysis was done using the above-mentioned software to plot dendrograms based on UPGMA (unweighted pair group method with arithmetic mean) method.

Table 6.4 Modified separation file for isoelectric focusing of keratins

SAMPLE APPL. DOWN AT				2.2		0 Vh
SAMPLE APPL. UP AT				2.3		0 Vh
EXTRA ALARM TO SOUND AT				1.1		73 Vh
SEP	2.1	2000 V	2.0 mA	3.5 W	12 °C	75 Vh
SEP	2.2	200 V	2.0 mA	3.5 W	12 °C	75 Vh
SEP	2.3	2000 V	5.0 mA	3.5 W	12 °C	500 Vh

6.2.3 DNA Techniques

6.2.3.1 Buffers and Solutions

All the required buffers and solutions were prepared using molecular biology grade reagents, and the rest of the buffers and reagents were the part of commercially available standard kits for DNA isolation, amplification and sequencing. The reagents and buffers used commonly for most of the work are mentioned below.

- 0.39 M DTT (dithiothreitol)
- 0.5 M EDTA (ethylenediaminetetraacetic acid) pH 8.0
- Proteinase K (20 mg/ml)
- Tris-HCl [1 M, pH 7.5]
- TE buffer [10 mM Tris-HCl, 0.1 mM EDTA]
- 20% SDS (sodium dodecyl sulphate)
- 5 M NaCl
- 10 X TBE buffer [Tris base 27 g, boric acid 13.75 g, 10 ml 0.5 M EDTA, H_2O 240 ml]
- Phenol/chloroform/isoamyl alcohol [25:24:1 v/v/v]
- 5 M NaOH
- 10% ammonium persulphate
- 40% acrylamide solution: (39.2 g acrylamide, 0.8 g bis-acrylamide, in 100 ml deionised water)
- N,N,N′,N′-tetramethylenediamine (TEMED)
- 3 and 5 M sodium acetate
- HCl (11.6 M)
- Tris (2 M)
- pH indicator
- 5% and 10% Chelex-100 resin
- Ethidium bromide (EtBr)
- Stain extraction buffer [Tris 1.21 g, NaCl 5.84 g, 20 g SDS, 0.01 M EDTA]
- Liquid nitrogen
- Restriction enzymes

6.2.3.2 Kits for DNA Extraction, PCR, PCR Purification, Cycle Sequencing and Dye Removal

- Qiagen DNeasy® tissue kit along with proteinase K (Qiagen, Germany)
- Qiagen EZ1® (BioRobot) DNA kit (tissue) (Qiagen, Germany)
- Geneclean® kit (Qbiogene, USA)
- PCR kit (MBI Fermentas)
- Qiagen MinElute® PCR purification kit (Qiagen, Germany)
- BigDye Terminator kit® V 3.1 for sequencing (ABI)
- Qiagen DyeEx® spin kit (Qiagen, Germany)

6.2.3.3 Extraction of DNA

DNA extraction from hair obtained as physical evidence from the wildlife offence cases presents a typical problem. The hair samples obtained in such cases do not have intact roots with follicular cells that can be used to isolate good quality of DNA having high base pair size. In most of the cases, the hair samples lack intact root region, and further, the samples may be old or poorly preserved as such the follicular cells are rarely present. Adding further to the problem, the samples of skins from which hair samples may be obtained are chemically treated (tanning) as such the possibility of finding intact follicular cells with good DNA is very rare. The hair samples were mostly obtained from collections from museums where a recipe of strong preservative chemicals is used for keeping the samples intact for long periods of time. This makes the possibility of getting follicular cells from hair roots a rare possibility. Thus the findings should have good relevance and applicability in actual DNA-based wildlife forensic investigations. DNA isolation was attempted from such hair samples and with six (6) different protocols, and various modifications in the protocols were attempted to obtain a most optimum protocol. Further, the success rates for extraction and subsequent success/failure rates in PCR amplification were also noted.

Six methods were used for DNA extraction, i.e. phenol-chloroform (organic) method, Chelex extraction, Qiagen DNeasy® tissue kit (Qiagen, Germany), BioRobot EZ1® tissue kit (Qiagen, Germany), Geneclean® kit (Qbiogene, USA) and alkali digestion method.

Prior to extraction, it is imperative to clean the hair samples thoroughly so that all the organic and inorganic dirt gets removed. The samples for all the extraction procedures were washed thoroughly after cutting into small pieces of 5–10 mm and kept immersed in Milli-Q/double-distilled water overnight in a 2.0 ml tube, and the water was removed [16]. The hair samples were then immersed in isopropyl alcohol for 2 h to remove organic dirt. The alcohol was decanted and the hair samples were air-dried after two such washes with alcohol. Such samples were used for DNA extraction with the selected protocols. The protocols with detailed steps are discussed below. As hair samples of *Pseudois nayaur*, *Moschus* sp. and *Capricornis sumatraensis* were available in good amount, the studies to select the most appropriate protocol out of the six chosen protocols and its subsequent standardisation were carried out with hair samples of these species.

Phenol-Chloroform (Organic) Method

The method has been traditionally the most commonly used method for DNA extraction. The protocol developed by Higuchi et al. [28] for extraction of DNA from single hair is based on this method. A slight alteration was done, instead of using absolute ethanol for precipitation; isopropyl alcohol was used for precipitation. Initial extractions were carried out using large sample volumes (400 mg or 50–60 hairs) of hair samples in order to obtain sufficiently good amount of DNA to view on 0.8% agarose gel such that the size range of DNA present in hair could be visualised. Later extractions were carried out with 5–12 hair samples. 2–4 μl of isolated DNA was used for PCR.

Chelex Extraction

Chelex-100, basically a cation exchange resin, is used, which prevents degradation of DNA, and high temperature breaks down the rest of the component of the cell/tissue. In case of Chelex extraction, the boiling step denatures the DNA such that the DNA obtained through Chelex extraction is essentially single stranded. However, since most of the forensic DNA analysis is centred around the use of polymerase chain reaction (PCR) which utilises a single strand for DNA amplification, the extracted DNA can be used for PCR amplification.

Qiagen DNeasy® Tissue Kit (Qiagen, Germany)

Qiagen tissue kit is a silica column-based method for DNA isolation from tissue (meat, blood, etc.). The column kit unites the discriminatory binding properties of silica-gel-based membrane along with the speed of microspin or vacuum technology for DNA isolation. In the final step, the DNA is eluted into the vial with an elution buffer from the Qiagen spin column through a brief incubation and centrifugation. Domingo-Roura et al. [16] used Qiagen kit as such for DNA extraction from hair samples suspected to be those of Eurasian badger (*Meles meles*) without any modification in the components of the kit except for the incubation time which was increased to 24 h, and more proteinase K was added. However, Pfeiffer et al. [50] used a different lysis buffer [100 mM Tris-HCl, 100 mM NaCl, 3 mM $CaCl_2$, 2% SDS, 40 mM DTT (dithiothreitol) and 250 µg/ml of proteinase K]. However, in the present study, Qiagen tissue kit was modified by adding 20–40 µl of 1 M DTT to the lysis buffer (ATL). This procedure lysed the hair samples satisfactorily, and there was good success in subsequent PCR amplification. The modified protocol is adopted from the modified protocols provided on the website of Qiagen (www.Qiagen.com).

(BioRobot EZ1®) Tissue Kit (Qiagen, Germany)

BioRobot EZ1 uses a magnetic particle technology for isolation of DNA from lysates. Nucleic acids are isolated from the lysate in a single step through their binding to the surface of the magnetic particles. These particles are separated from the lysate using a magnetic source and are then washed with great efficiency. Finally the nucleic acids are eluted in water or buffer having low salt content. The system uses integrated circuit cards (EZ1cards) which contain the protocols for the running of the system. Instead of using a specific lysis buffer (Buffer G2 for tissue lysis) as recommended by the manufacturer, the lysate of hair samples as prepared by the modified Qiagen DNeasy® tissue kit was used. EZ1 forensic card was used for the isolation of DNA from the lysate, and DNA was obtained in about 20 minutes. The major problem regarding EZ1 BioRobot is that the minimum elution volume is

50 µl as such the DNA gets diluted. This was overcome by eluting the DNA in water and then concentrating it by a lyophiliser to a volume of 10–20 µl.

Geneclean® Kit (Qbiogene, USA)

Geneclean® kit is meant for isolation of DNA from ancient and preserved/museum samples of biological tissue like the bone, skin, etc. The protocol was used as recommended by the manufacturers. As the kit requires the tissue sample to be homogenised, mortar and pestle were used to homogenise the hair samples using liquid nitrogen.

The following steps were followed for the isolation of DNA. Initially 100 mg of hair sample was used and later on 5–12 hairs were tried.

Alkali Digestion Method

This method involves digestion of hair samples of other keratin-rich samples with a strong aqueous solution of an alkali. Hair samples are treated with a 5 M solution of sodium hydroxide, and they get lysed in about 2–5 h [24]. This strong alkali solution is neutralised with a solution of Tris base (2 M) and HCl (11.6 M). The final pH is brought in the range of 6–8 by the neutralisation. The lysate is then filtered through Centricon®/Amicon YM-30 (Millipore) columns. DNA is washed with three washes of TE, and the final retentate contains the DNA and is used for PCR.

6.2.3.4 PCR Amplification of Selected Mitochondrial Gene Fragments

PCR amplification for segments of four mitochondrial genes was attempted with universal primers. The most important gene amongst these was mitochondrial cytochrome b gene, which has been extensively used for identification of species and phylogenetic studies. Two primers for this region were tried, the first one was by Kocher et al. [32] which amplified a 359 base pair (approx.) fragment which will include a variable 307 bp region. In addition to this, a second primer [62] amplifying a 398 bp (approx.) region was used. The second region attempted for mitochondrial DNA analysis was 12s rRNA gene with one universal primer amplifying approximately 400 bp (approx.) used by Girish et al. [22] for species identification from meat samples. The third region attempted for amplification in the study was 16s rRNA gene using universal primer [43] that generated an amplicon of approximately 540 bp. In addition to these four regions, attempt was also made to amplify D-loop region of mitochondrial DNA for the successful extraction with a primer that targeted and generated 1 kb (approximately) amplicon. The details of the primer sequences used are mentioned in Table 6.5.

Table 6.5 Details of the primers used for the amplification of mitochondrial gene fragments

Gene	Size (bp)	Primer sequence (5′-3′)		References
Cytochrome b	381	Forward	CCATCCACATCTCCGCATGATGAAA	Kocher et al. [32]
		Reverse	CCCCTCAGAATGATATTTGGCCTCA	
Cytochrome b	398	Forward	TACCATGAGGACAAATATCATTCTG	Verma and Singh [62]
		Reverse	CCTCCTAGTTTGTTAGGGATTGATCG	
12s rRNA	400	Forward	CAAACTGGGATTAGATACCCCACTAT	Girish et al. [22]
		Reverse	GAGGGTGACGGGCGGTGTGT	
16s rRNA	540	Forward	CGCCTGTTTATCAAAAACAT	Mitchell et al. [43]
		Reverse	CTCCGGTTTGAACTCAGATC	
D-loop	1000	Forward	TGAATTGGAGGACAACCAGT	Shields and Kocher [58]
		Reverse	CCTGAAGTAGGAACCAGATG	

Table 6.6 Optimised conditions for the polymerase chain reaction of 36 cycles used for the amplification of various mitochondrial genes used in this study

	Denaturation		Annealing		Chain extension	
Primers	Temp. (°C)	Time (min)	Temp. (°C)	Time (min)	Temp. (°C)	Time (min)
Cytochrome b 381 bp	94	1.00	55	1.00	72	1.00
Cytochrome b 398 bp	94	1.00	51	1.00	72	1.00
12s rRNA 450 bp	94	0.50	51	0.50	72	1.00
16s rRNA 540 bp	94	0.450	49	0.45	72	0.45
D-loop 1 kb	94	1.00	55	1.00	72	1.00

PCR reagents from MBI Fermentas were used for the entire study. The standardisation of the PCR conditions for primers especially annealing temperature was done with help of gradient PCR (MJ Research), and samples of DNA extracted from muscle samples were used for the same as template. Once the protocols were standardised, the extracted DNA of selected species was tried for PCR amplification. Initially samples of *Pseudois nayaur* and *Moschus fuscus* were used for the studies as hair samples were available in significant amount. 6–8 µl of extracted DNA was used for a 25 µl PCR reaction. The amplifications were checked on a 2% agarose gel at a constant voltage of 120 volts and visualised using EtBr stain. The amplifications which showed a faint product of very low intensity were used as template for running a nested PCR so that a good concentration of amplicon could be obtained. 0.25 µl–0.5 µl of faint PCR product was used for the second PCR. The optimised PCR conditions are given in Table 6.6.

6.2.3.5 Cleaning of PCR Products and DNA Sequencing

Qiagen MinElute PCR purification kit was used to clean up the amplified products, and the unincorporated nucleotides and primes were removed prior to DNA sequencing. Only forward/reverse primers were selected out of the universal primers and used for DNA sequencing. The standard protocol for BigDye Terminator kit V3.1(ABI 3130 Genetic Analyzer) was used to perform the sequencing.

6.2.3.6 Data Analysis and Sequence Submission to NCBI/GenBank

The ABI Sequencing Analysis V5.2 software was used for data analysis. In addition to this, BioEdit version 5 [26] and MEGA (Molecular Evolutionary Genetics Analysis software version 4.0.) [52], Clustal W, etc. were used. The sequences were compared with the GenBank/NCBI (National Center for Biotechnology Information) through BLAST (basic locus alignment search tool) analysis. The interspecies similarity/variation was studied with help of MEGA, and the discriminatory power of the studied mitochondrial genes was assessed using MEGA. This was done by comparing the sequences for each gene across the species studied, to see the similarity between the sequences.

The overall data analysis for success rate of extraction procedures, extraction success with respect to species, PCR success rate, etc. was carried out with SPSS (Release 12.0.0).

6.2.3.7 PCR-RFLP Analysis

Although the technique is not required for the identification of species if the sequencing of mitochondrial DNA genes has been accomplished, this technique was pursued with a view to generate database of restriction enzymes that can be used for restriction digestion to study restriction fragment length polymorphism for the amplicons generated by using the primers that were used in the current study. This was also pursued to test the applicability of PCR-RFLP for species identification with the amplicons generated by the selected primers. This may be of good use for laboratories lacking DNA sequencing facilities. Restriction digestion was tried with some of the enzymes like HaeIII, AluI and RsaI for cytochrome b gene amplicons. As the technique is basically a hit and trial method if the sequence of PCR amplicon is not known, a web-based programme Restriction Mapper was used to generate restriction enzyme digestion map for the selected mitochondrial gene fragments that were successfully sequenced.

6.3 Results and Discussions

6.3.1 Keratin Profiling Techniques

6.3.1.1 SDS-PAGE

Extraction Protocol for Keratin Profiling

Although the first extraction protocol (SDS-free extraction buffer) is applicable to both SDS-PAGE and IEF techniques, it was observed that the second protocol (extraction buffer containing SDS) was much suited for the SDS-PAGE. The sample requirement is much lower and further, the extraction time is also less. A comparison of the two extraction protocols was done by comparing the keratin patterns obtained on a 20% Laemmli's gel system and overnight Coomassie staining. Reference samples of shahtoosh (*Pantholops hodgsonii*) and black buck (*Antilope cervicapra*) were used for the comparative study. It was observed that in case of the first protocol, the number of bands was less, and further the intensity of bands was also low as compared to bands observed in case of the second protocol after electrophoresis and Coommasie staining (Fig. 6.1). The number of bands observed and their visual intensity are shown in tabular form (Table 6.7). Therefore the second protocol (extraction buffer containing SDS) was followed throughout the study for keratin extraction for SDS-PAGE.

Efficacy of Gel Concentration, Running Voltage and Temperature

It was noticed in the trials with Laemmli's gel system that the optimised process was very slow and tedious and took about 8–10 h for separation of keratins. Further, the resolution of the bands was not so satisfactory, and hence the modified Schagger and Von gel system [30, 57] was tried. Two variants of the resolving gel were tried, first with 1x cross-linkage and the second one with 2x cross-linkage, which was

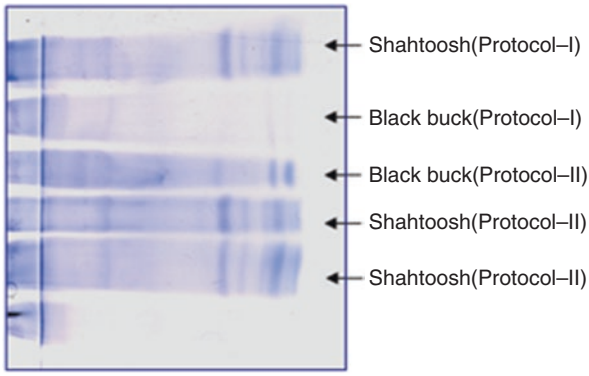

Fig. 6.1 Comparison of protocols for their efficacy of keratin extraction

Table 6.7 Similarity matrix for cytochrome b gene sequence for the selected species

	Bos grunniens	*Hemitragus jemlahicus*	*Bos gaurus*	*Pseudois nayaur*	*Procapra picticaudata*	*Ovis ammon*	*Gazella bennetti*	*Antilope cervicapra*	*Cervus duvauceli*	*Cervus eldi*	*Capricornis sumatraensis*	*Bubalus bubalis*	*Tetracerus quadricornis*	*Moschus fuscus*
Bos grunniens	ID	0.858	0.924	0.861	0.858	0.856	0.856	0.858	0.843	0.853	0.858	0.886	0.851	0.858
Hemitragus jemlahicus		ID	0.841	0.904	0.863	0.886	0.836	0.846	0.856	0.846	0.871	0.851	0.876	0.841
Bos gaurus			ID	0.833	0.861	0.846	0.858	0.861	0.846	0.853	0.851	0.866	0.848	0.838
Pseudois nayaur				ID	0.853	0.881	0.833	0.846	0.851	0.838	0.891	0.866	0.856	0.833
Procapra picticaudata					ID	0.838	0.866	0.874	0.853	0.836	0.846	0.846	0.843	0.848
Ovis ammon						ID	0.833	0.843	0.848	0.846	0.876	0.871	0.848	0.863
Gazella bennetti							ID	0.896	0.831	0.833	0.841	0.848	0.851	0.846
Antilope cervicapra								ID	0.853	0.848	0.869	0.841	0.853	0.841
Cervus duvauceli									ID	0.914	0.846	0.828	0.851	0.856
Cervus eldi										ID	0.841	0.838	0.846	0.863
Capricornis sumatraensis											ID	0.871	0.846	0.853
Bubalus bubalis												ID	0.869	0.874
Tetracerus quadricornis													ID	0.843
Moschus fuscus														ID

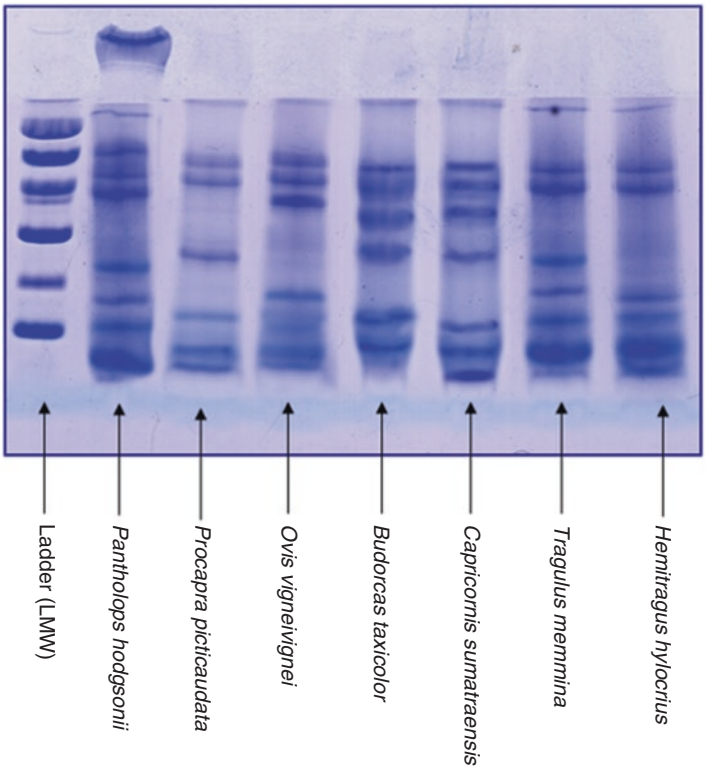

Fig. 6.2 Keratin profile of some selected species

attained by varying the quantity of the bis-acrylamide in the stock acrylamide solution. In comparing the efficacy of 1x and 2x cross-linkage, it became apparent that 2x cross-linkage, although being capable of separating out proteins with very little difference in their molecular weight, was not suitable for keratins. It was observed that the 1x cross-linked gels performed the best for resolving the keratins, to serve the purpose of species characterisation. The ladder as well as the keratins for the selected species separated satisfactorily with modified Schagger and Von gel system, having 1x cross-linkage or 1x concentration of bis-acrylamide. Therefore, 1x cross-linked, modified Schagger and Von gel system was used to generate keratin profiles of selected species. SDS-PAGE-based keratin patterns for all the selected species were obtained using modified Schagger and Von gel system. Figure 6.2 shows separated keratin bands for Tibetan antelope (*Pantholops hodgsonii*) which is the source of shahtoosh wool and for Ladakh urial/shapu (*Ovis vignei vignei*). It can be clearly seen that there are distinct bands for *Pantholops hodgsonii* and *Ovis vignei vignei*, and the ladder is also fully separated. The gel system was found to perform most optimally at a temperature at 12 °C. Similarly, the optimum voltage for the separation was 50 V (constant), and the electrophoresis run could be accomplished within 4–5 h.

Standardisation of Staining Procedures of SDS-PAGE Gels

Coomassie staining (Coomassie Brilliant Blue R 250) and conventional silver staining were used to standardise the staining procedure for SDS-PAGE gels. Additionally silver diamine staining protocol was also tried [46]. It was observed that the gels stained satisfactorily with Coomassie Brilliant Blue. The staining was initially done overnight at room temperature (25 °C ± 3). The excess of background blue-coloured gel staining was eliminated by destaining the gel for a minimum of 1 h. As the process of Coomassie Brilliant Blue staining is a lengthy process, where a minimum of 6–12 h were required to get proper staining of gels, an external heat-mediated path was used to hasten the process of staining and destaining. The gels were placed in a glass staining tray having the staining solution of Coomassie Brilliant Blue and heated to a temperature of 40–50 °C in an oven. This resulted in quick staining of gels and all the gels of 0.5 mm thickness stained in 30 min. Silver staining (conventional) and diamine-based staining stained the gels too heavily as such background noise was very high. Hence, Coomassie Brilliant Blue staining procedure is better suited for the study of keratin patterns with SDS-PAGE for the selected species.

Species Specificity and Reproducibility of Keratin Patterns Obtained by SDS-PAGE

The keratin patterns for all the species showed a great deal of species specificity. The keratin patterns using SDS-PAGE (Schagger and Von gel) for the selected species are given in Fig. 6.2. It is apparent from the visual examination of the keratin patterns that they are species specific as there is difference in the number of bands and their positions for each species. In addition to this, the keratin pattern also showed intra-species consistency with appearance of similar bands for different individuals of the same species. The dendrogram plotted using Statistica software, based on unweighted pair group centroid method, has been shown in Fig. 6.3. From the dengrogram it becomes evident that species can be characterised from the keratin patterns.

6.3.1.2 Isoelectric Focusing (IEF)

Efficacy of Keratin Extraction Protocol

The extraction protocol as laid down by Marshall [38], Marshall and Gillespie [40] and later modified by Carracedo et al. [9–11] was found to be apt for extraction of kerations from hair for isoelectric focusing (IEF). In an attempt to reduce the time of extraction, the samples extracted in just 28 h were tried, and it was observed that reducing the extraction time did not introduce any variation in results.

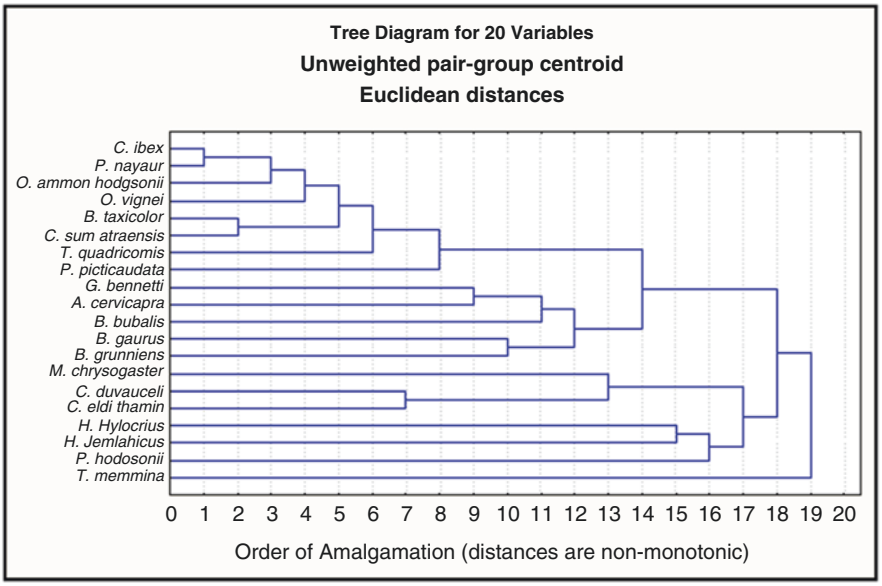

Fig. 6.3 Dendrogram for molecular weights of separated keratins in selected species, plotted using Statistica software, based on unweighted pair group centroid

Standardisation of Running Process

Following the standard protocol, it was observed that (Separation Technique File 100) at the recommended temperature (15 °C), the gels used to get heated up too much, and they used to get desiccated while the run was still on, as such the electrophoresis could not be completed properly due to interruption in the flow of current. This resulted in incomplete runs, and thus bands were not focussed properly. This problem was overcome by reducing the gel bed temperature to 12 °C that apparently helped in quicker dissipation of heat from the gel surface due to which the desiccation could be avoided. In addition to this, the final step in the standard protocol for isoelectric focusing was modified by increasing the run time from 410 Watt hours to 500 Watt hours. With this protocol the gels showed clear separation of keratins without any unacceptable distortions.

Standardisation of Staining

Like SDS-PAGE, Coomassie staining (Coomassie Brilliant Blue R 250) was initially tried on the IEF gels. However, there was no development of bands after overnight staining and subsequent destaining. Only the broad pI calibration ladder

appeared in the gel. The reason for this may be the low concentration of proteins in the extract. Conventional silver staining and silver diamine staining were able to detect even the low quantity of keratin proteins separated in the gels. However, conventional silver staining had lots of problems related to intense background staining which rendered the gels useless for any valid interpretation. Hence, silver diaminne [46, 60] staining was preferred over conventional silver staining method due the greater sensitivity and the relative ease in destaining of the gels. Figure 6.4 shows keratin patterns obtained by silver diamine staining.

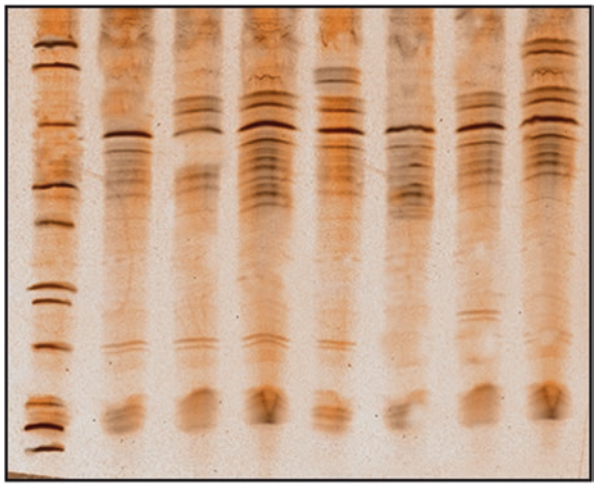

Fig. 6.4 Keratin patterns of some species with IEF and silver diamine staining

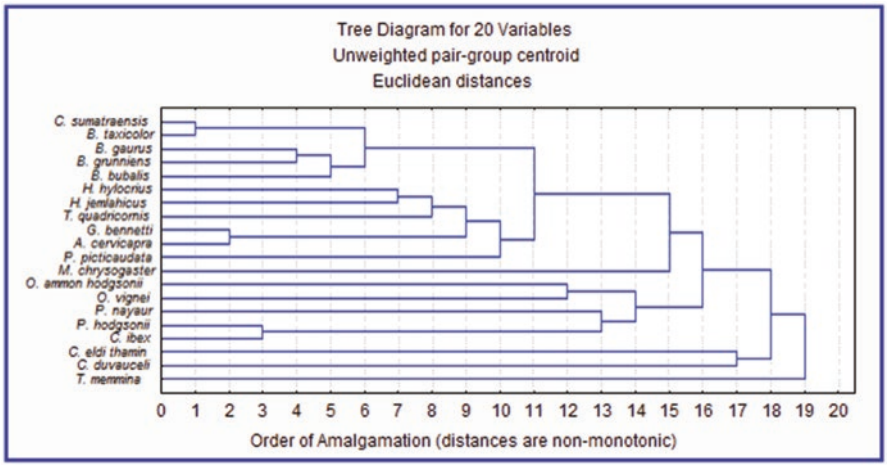

Fig. 6.5 Dendrogram for pI values of separated keratins in selected species, plotted using Statistica software, based on unweighted pair group centroid

Species Specificity and Reproducibility of Keratin Patterns Obtained by IEF

The keratin patterns for all the species showed a great deal of species specificity. Apparently for the visual examination, it becomes clear that the keratin patterns are species specific owing to the differences in the number and position of the bands obtained for each species. In addition the keratin pattern also showed intraspecies consistency with appearance of similar bands for different individuals of the same species. The Dendrogram plotted using Statistica software, based on unweighted pair group centroid method, is given in Fig. 6.5.

6.3.2 DNA Profiling

6.3.2.1 Efficacy of DNA Extraction with the Six Different Protocols Tried

The initial extraction trials with 400 mg (bulk) of hair samples with phenol-chloroform method indicated that the DNA present in hair samples was highly fragmented and most of it was of very low base pair length. It was observed in case hair samples of *Pseudois nayaur*, *Moschus fuscus* and *Capricornis sumatraensis* that most of the visible DNA after ethidium bromide staining was less than 100 base pairs in length (Fig. 6.6).

From the figure image of the gel (Fig. 6.6), it is apparent that no visible amount of DNA with ethidium bromide (EtBr) staining is present between the 300 bp and 1000 bp which was the targeted fragment length for the primers selected in this

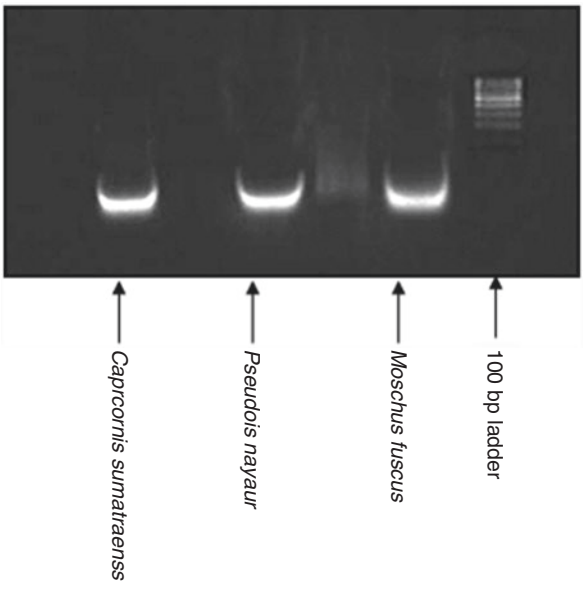

Fig. 6.6 DNA extracted in bulk from hair samples showing the low base pair size of DNA

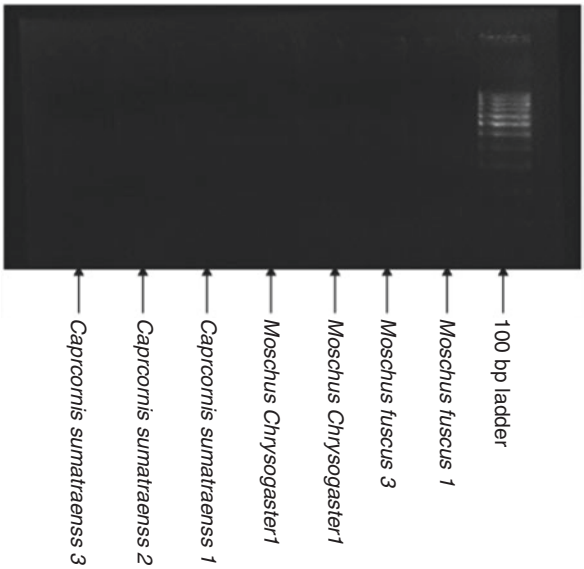

Fig. 6.7 Gel run with DNA extracted from five to ten hair samples showing no visible DNA in the lanes

study. The DNA extractions that were carried out later with lower amount of hair samples (5–12 hairs) of *Pseudois nayaur*, *Moschus fuscus* and *Capricornis sumatraensis* showed no DNA at all after staining with EtBr stain on 0.8% agarose gel in TBE (Fig. 6.7). It is apparent from the yield gel images that the concentration of DNA will always be low in hair samples and may fall below the detection limit of EtBr staining. Hence instead of yield, success rate of PCR would be more important with respect to the six extraction methods.

6.3.2.2 Polymerase Chain Reaction (PCR)

Standardisation of PCR Conditions

Through gradient PCR (MJ Research thermal cycler), the most optimum conditions for DNA amplification for each of the primer selected were standardised. It can be clearly observed in gradient PCR run for 12S rRNA gene; how intensity of amplified products as visualised on 2.0% agarose gels (EtBr stain) changes with the annealing temperature (Fig. 6.8). Similarly, for other primers, the most optimum conditions were selected.

Because the DNA extracted from hair was highly diluted, higher volume of DNA extract was required as template for successful amplifications, and normally 36–40 cycles were required to get a good amplification. It was always useful to keep the reaction volume to 25 µl instead of 50 µl to save the template DNA for further use. Primer concentration of 0.2 µM each, for forward and reverse primers, and the use of 6–8 µl of extracted DNA were found to be optimum for getting good ampli-

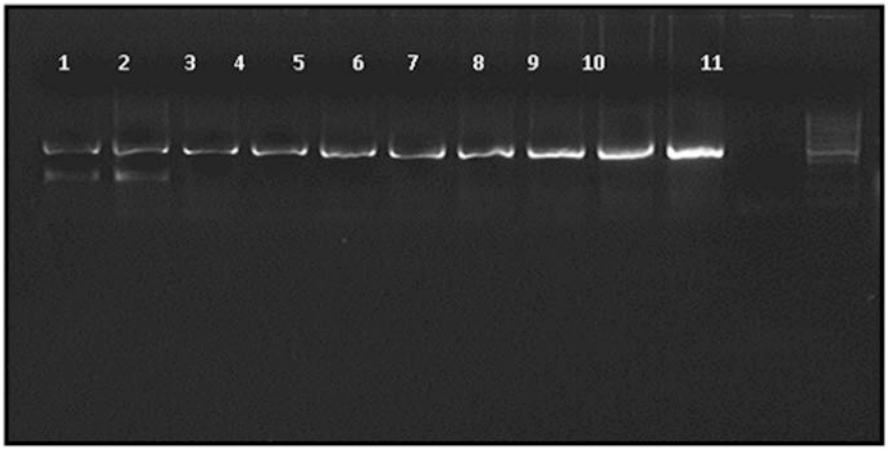

Fig. 6.8 Gradient PCR for 12S rRNA primer, with lane 10 showing the best intensity of the amplicon. Lane 1, 58 °C; Lane 2, 57.2 °C; Lane 3, 56.5 °C; Lane 4, 56 °C; Lane 5, 54.8 °C; Lane 6, 54 °C; Lane 7, 53.5 °C; Lane 8, 52.4 °C; Lane 9, 51.6 °C; Lane 10, 51.0 °C and Lane 11, 100 base pair ladder

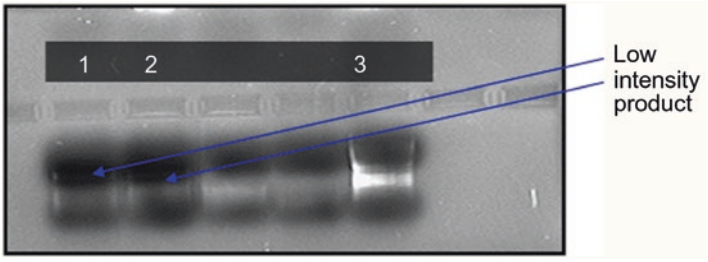

Fig. 6.9 Low-intensity PCR products for 16S rRNA gene amplified from DNA extracted from the hair sample of *Pseudois nayaur*. Lane 1 and 2 low-intensity products on 2.0% agarose gel by EtBr staining. Lane 3, 100 bp ladder

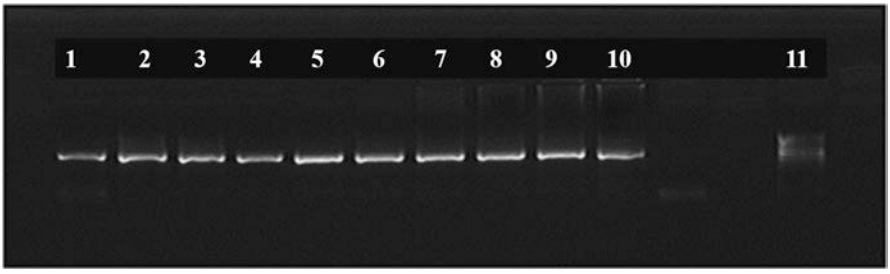

Fig. 6.10 Amplifications (16S rRNA for *Pseudois nayaur*) obtained by using PCR product as template for PCR. Lanes 1–10 (PCR product) and lane 11 (100 bp ladder)

fications. In most of the cases, owing to the low concentration of template DNA in the extract, the amplifications were very faint as visualised on 2.0% agarose gel by ethidium bromide staining. Figure 6.9 shows such PCR product for *Pseudois nayaur* for 16S rRNA gene where the intensity is very low as seen with ethidium bromide staining. In order to have a good quality PCR product in reasonable quantity for further analysis, it was found worthwhile to use the PCR product of low intensity as a template for nested PCR to get a PCR product with good concentration. Figure 6.10 shows PCR amplifications obtained by using faint PCR product shown in Fig. 3.9 for *Pseudois nayaur*.

PCR Success Rate with Respect to Extraction Protocol

As all the extraction protocols did not yield visible amount of DNA (visualised on a 0.8% agarose gel through ethidium bromide staining) from the limited amount of hair samples (5–12 hairs), their efficacy of extracting DNA could thus be rated only on the success rate of subsequent PCR for the DNA extracted through them. As the quality of DNA obtained from extraction has a great bearing on the success of PCR, the PCR success rates can be used to get an indication about the quality of DNA extracted by each protocol. A good success rate of PCR means that the extracted DNA is free of PCR inhibitors, and the protocol is good for DNA extraction even if the quantity of extracted DNA is low. It was observed in DNA extracted in case of phenol-chloroform method that even if DNA was extracted from a bulk amount of hair sample (400 mg of hair sample), the PCR success rate was roughly 40%. This was probably due to the high level of PCR inhibitors present in the extracted DNA.

In comparison with phenol-chloroform method, Geneclean® kit method was much efficient with respect to PCR amplification success rates. When DNA was extracted from a bulk amount of hair (100 mg), there was a good success rate in PCR amplification. The success rate was approximately 66% with respect to cytochrome b (398 bp) gene fragment amplification. The other four techniques, i.e. Qiagen DNeasy® tissue kit, Chelex extraction, BioRobot EZ1® and alkali digestion method, did not involve the extraction of DNA from a bulk amount of samples. Out of these methods, no success in PCR amplification was observed in case of Chelex and alkali digestion method.

Modified Qiagen DNeasy® tissue kit and Qiagen BioRobot EZ1® showed the best success rates with respect to PCR amplification. The average success rate for PCR (cytochrome b 398 bp, 16s rRNA 540 bp and 12s rRNA 450 bp gene fragments) from the DNA extracted from hair using these methods was about 70%. This indicates towards the efficiency of these methods to remove PCR inhibitors while extraction. Further, the advantage of concentrating DNA extracted through BioRobot EZ1® by lyophilisation was found to be very useful with the samples with low DNA concentration. As the Qiagen DNeasy® tissue kit is comparatively cheaper than Qiagen BioRobot EZ1® method and has almost similar efficiency with respect to PCR success, all the extractions for selected species were performed by Qiagen DNeasy® tissue kit. Figure 6.11 shows 398 bp pair cytochrome b gene amplified for three samples of *Pseudois nayaur*, one sample of *Capricornis sumatraensis* and one

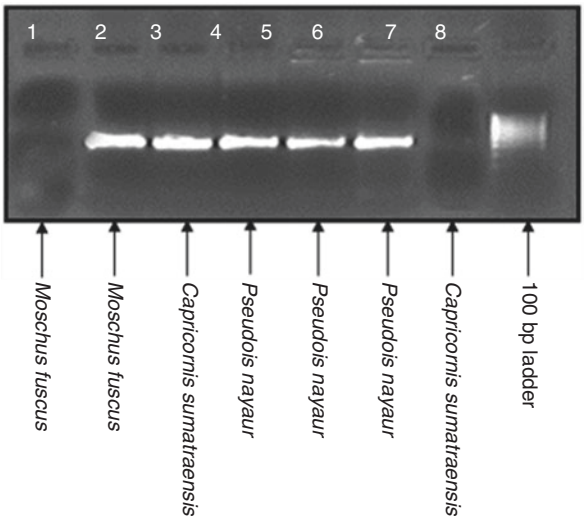

Fig. 6.11 PCR success rate with Qiagen DNeasy® tissue kit where about 70% of the samples is amplified (398 bp cytochrome b gene)

sample of Moschus *fuscus* out of the seven samples in all tried for these species through Qiagen DNeasy® tissue kit.

PCR Success Rate with Respect to Primers and Amplicon Size

Out all the primers selected for the study, the efficiency of all the primers was also compared, and it was observed that all the primers worked successfully and amplified the targeted sequence, except for the primer selected for D-loop region. As the D-loop primer is supposed to amplify a region of about 1 kilobase, which was perhaps not available in the DNA extracted from hair samples, no amplification was observed with any of the DNA samples tried for the selected species.

PCR Success Rate with Respect to Samples

As all the samples were old samples mostly collected from museums with little possibility of getting follicular cells with the root of the hair, it was not possible to amplify DNA for all the species. Out of the 20 selected species, it was possible to amplify DNA for fifteen (15) species only. For five species, viz. *Pantholops hodgsonii*, *Capra ibex*, *Hemitragus hylocrius*, *Ovis vignei vignei* and *Budorcas taxicolor*, it was not possible to extract/amplify DNA.

Table 6.8 Similarity matrix for cytochrome b gene sequence for the selected species

	Bos grunniens	*Hemitragus jemlahicus*	*Bos gaurus*	*Pseudois nayaur*	*Procapra picticaudata*	*Ovis ammon*	*Gazella bennetti*	*Antilope cervicapra*	*Cervus duvauceli*	*Cervus eldi*	*Capricornis sumatraensis*	*Bubalus bubalis*	*Tetracerus quadricornis*	*Moschus fuscus*
Bos grunniens	ID	0.858	0.924	0.861	0.858	0.856	0.856	0.858	0.843	0.853	0.858	0.886	0.851	0.858
Hemitragus jemlahicus		ID	0.841	0.904	0.863	0.886	0.836	0.846	0.856	0.846	0.871	0.851	0.876	0.841
Bos gaurus			ID	0.833	0.861	0.846	0.858	0.861	0.846	0.853	0.851	0.866	0.848	0.838
Pseudois nayaur				ID	0.853	0.881	0.833	0.846	0.851	0.838	0.891	0.866	0.856	0.833
Procapra picticaudata					ID	0.838	0.866	0.874	0.853	0.836	0.846	0.846	0.843	0.848
Ovis ammon						ID	0.833	0.843	0.848	0.846	0.876	0.871	0.848	0.863
Gazella bennetti							ID	0.896	0.831	0.833	0.841	0.848	0.851	0.846
Antilope cervicapra								ID	0.853	0.848	0.869	0.841	0.853	0.841
Cervus duvauceli									ID	0.914	0.846	0.828	0.851	0.856
Cervus eldi										ID	0.841	0.838	0.846	0.863
Capricornis sumatraensis											ID	0.871	0.846	0.853
Bubalus bubalis												ID	0.869	0.874
Tetracerus quadricornis													ID	0.843
Moschus fuscus														ID

Table 6.9 Similarity matrix for 12S rRNA gene sequence for the selected species

	Bos grunniens	*Hemitragus jemlahicus*	*Antilope cervicapra*	*Ovis ammon*	*Pseudois nayaur*	*Capricornis sumatraensis*	*Procapra picticaudata*	*Bos gaurus*	*Moschus fuscus*	*Tetracerus quadricornis*	*Gazella bennetti*	*Cervus duvauceli*	*Bubalus bubalis*
Bos grunniens	ID	0.896	0.893	0.884	0.905	0.893	0.878	0.961	0.919	0.913	0.922	0.91	0.926
Hemitragus jemlahicus		ID	0.91	0.955	0.955	0.931	0.904	0.913	0.928	0.904	0.925	0.913	0.913
Antilope cervicapra			ID	0.908	0.922	0.913	0.898	0.91	0.934	0.916	0.967	0.904	0.902
Ovis ammon				ID	0.952	0.926	0.896	0.902	0.916	0.902	0.91	0.91	0.902
Pseudois nayaur					ID	0.928	0.913	0.913	0.931	0.916	0.925	0.916	0.928
Capricornis sumatraensis						ID	0.893	0.91	0.922	0.916	0.928	0.905	0.911
Procapra picticaudata							ID	0.901	0.922	0.91	0.913	0.904	0.875
Bos gaurus								ID	0.937	0.925	0.928	0.928	0.931
Moschus fuscus									ID	0.94	0.949	0.934	0.916
Tetracerus quadricornis										ID	0.937	0.916	0.922
Gazella bennetti											ID	0.925	0.91
Cervus duvauceli												ID	0.913
Bubalus bubalis													ID

Table 6.10 Similarity matrix for 16S rRNA gene sequence for the selected species

	Pseudois nayaur	*Hemitragus jemlahicus*	*Ovis ammon*	*Tragulus meminna*	*Procapra picticaudata*	*Gazella bennetti*	*Tetracerus quadricornis*	*Bos gaurus*	*Cervus duvauceli*	*Antilope cervicapra*	*Capricornis sumatraensis*	*Cervus eldi*	*Moschus fuscus*	*Moschus chrysogaster*	*Bubalus bubalis*	*Bos grunniens*
Pseudois nayaur	ID	0.954	0.954	0.866	0.928	0.9	0.931	0.904	0.921	0.923	0.945	0.907	0.921	0.935	0.9	0.897
Hemitragus jemlahicus		ID	0.923	0.876	0.916	0.902	0.916	0.892	0.921	0.921	0.937	0.919	0.909	0.923	0.897	0.888
Ovis ammon			ID	0.862	0.914	0.897	0.909	0.897	0.902	0.909	0.923	0.897	0.916	0.916	0.897	0.902
Tragulus meminna				ID	0.859	0.85	0.855	0.845	0.855	0.862	0.855	0.85	0.862	0.867	0.843	0.85
Procapra picticaudata					ID	0.914	0.926	0.921	0.892	0.947	0.906	0.883	0.919	0.926	0.907	0.909
Gazella bennetti						ID	0.888	0.895	0.885	0.954	0.88	0.885	0.9	0.902	0.907	0.883
Tetracerus quadricornis							ID	0.905	0.912	0.904	0.907	0.895	0.928	0.924	0.905	0.905
Bos gaurus								ID	0.89	0.9	0.895	0.881	0.904	0.9	0.93	0.966
Cervus duvauceli									ID	0.892	0.909	0.964	0.893	0.904	0.912	0.893
Antilope cervicapra										ID	0.909	0.883	0.911	0.919	0.907	0.885
Capricornis sumatraensis											ID	0.902	0.909	0.914	0.897	0.89
Cervus eldi												ID	0.881	0.893	0.902	0.883
Moschus fuscus													ID	0.959	0.902	0.904
Moschus chrysogaster														ID	0.897	0.9
Bubalus bubalis															ID	0.93
Bos grunniens																ID

6.3.2.3 DNA Sequencing

All the successful amplifications were successfully sequenced using the BigDye Terminator kit® V 3.1 (ABI Prism 3130 Genetic Analyzer). The sequences were edited with the sequence editing softwares mentioned under the materials and methods section. Some of the sequences like the partial sequence for cytochrome b gene and 16s rRNA gene for *Tragulus meminna* were the first to be submitted to GenBank/NCBI.

6.3.2.4 Analysis of DNA Sequences

The study of similarity between the sequences of each gene (cytochrome b, 12s rRNA and 16s rRNA) for the species studied and successfully sequenced indicated that cytochrome b gene has maximum discriminatory power. The similarity index for the cytochrome b gene, 12S rRNA gene and 16S rRNA gene between the species studied is given in Tables 6.8, 6.9 and 6.10. From these tables it can be seen that cytochrome b gene sequences show least similarity across the species studied and hence have maximum discriminatory power in comparison with 16S rRNA and 12S rRNA gene. The neighbour joining (NJ) trees for the species studied and sequenced based on cytochrome b gene, 12S rRNA gene and 16S rRNA are given in Figs. 6.12, 6.13 and 6.14.

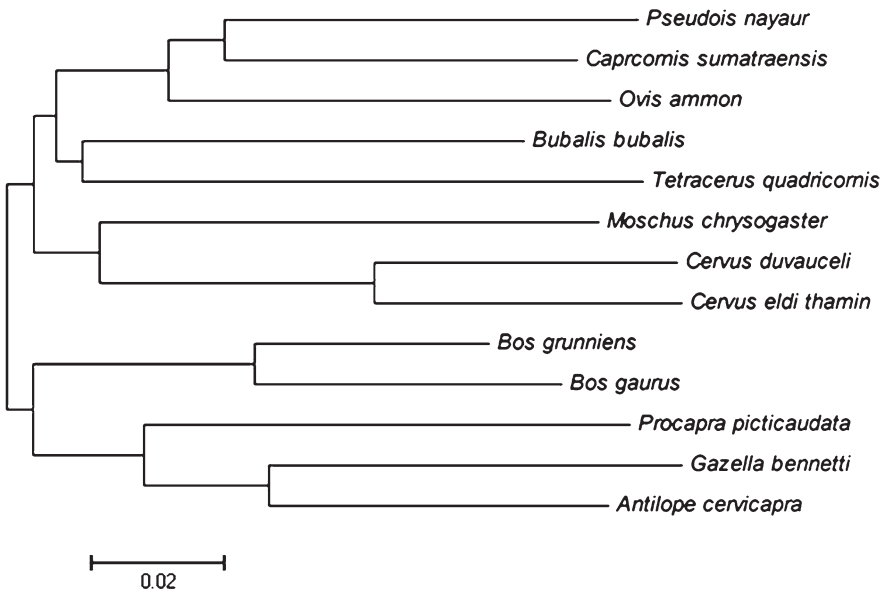

Fig. 6.12 Neighbour joining tree based on mitochondrial cytochrome b gene sequences (MEGA)

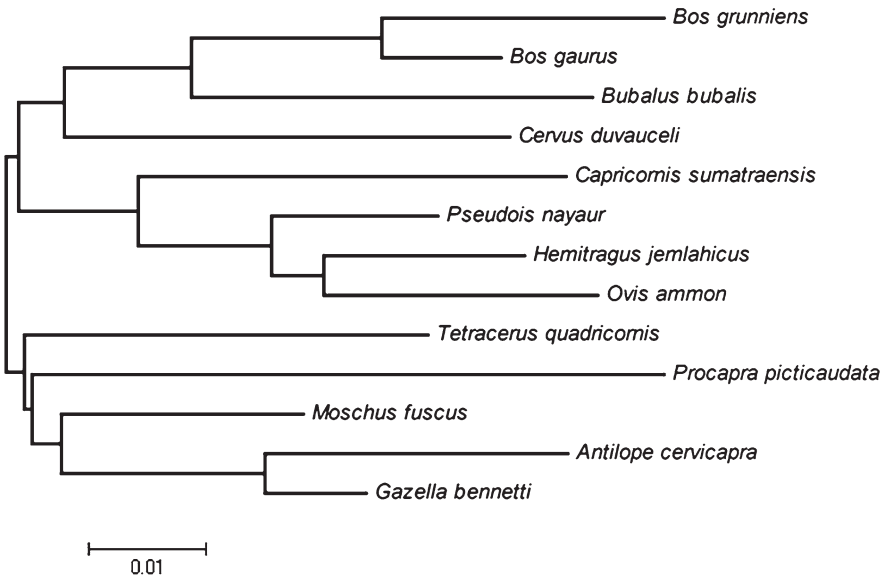

Fig. 6.13 Neighbour joining tree based on mitochondrial 12S rRNA gene sequences

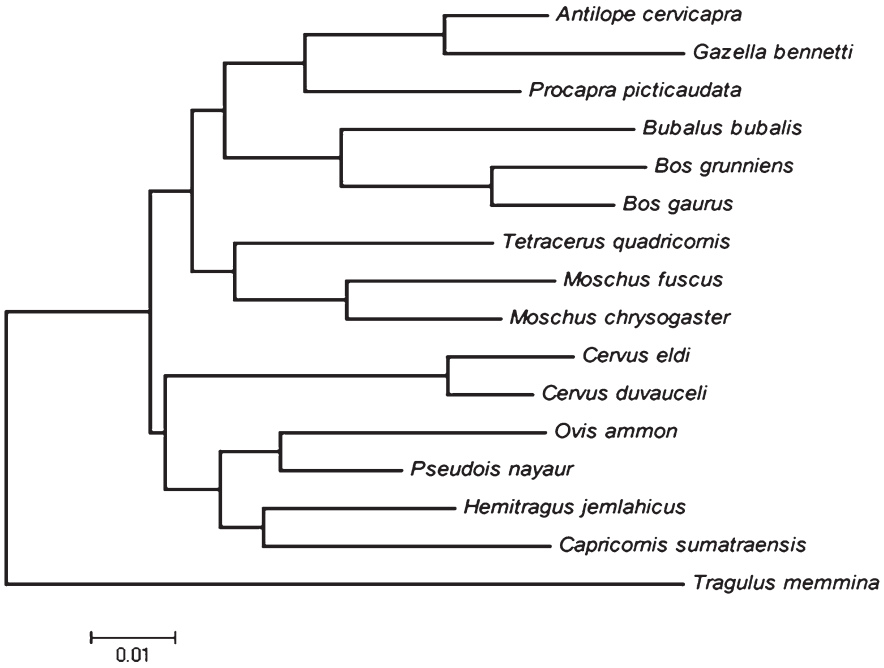

Fig. 6.14 Neighbour joining tree based on mitochondrial 16S rRNA gene sequences

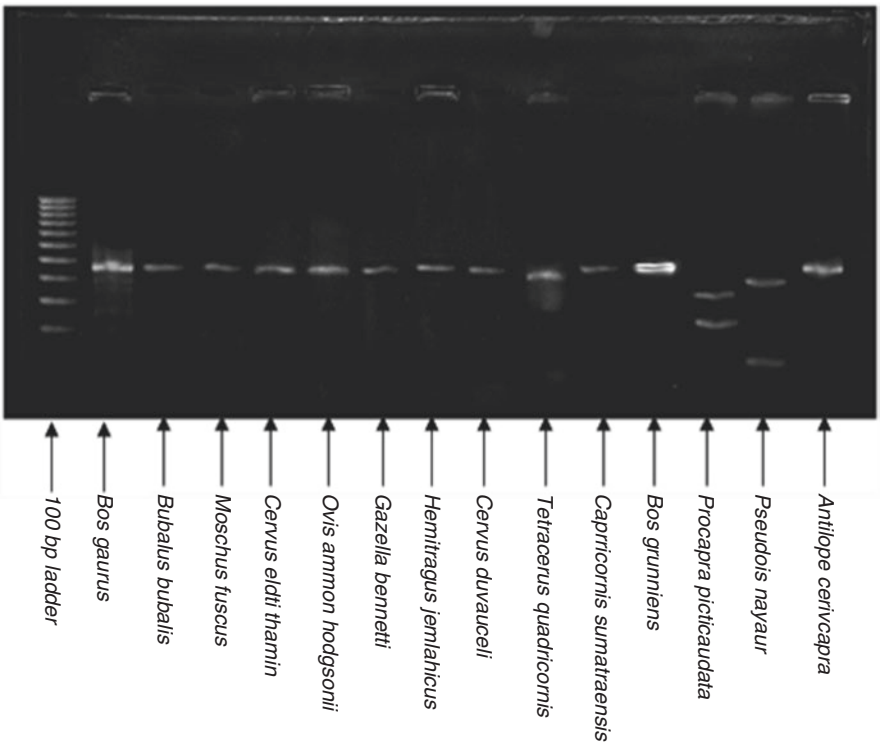

Fig. 6.15 Restriction fragment length polymorphism for cytochrome b gene amplicon (398 bp) using *HaeIII* enzyme (3% agarose gel)

6.3.2.5 Restriction Fragment Length Polymorphism (RFLP)

PCR-RFLP studies for cytochrome b gene fragment with three restriction enzymes (*HaeIII*, *AluI* and *RsaI*) showed that none of the enzymes is able to cut cytochrome b gene amplicons (398 bp approx.) in all the species where it was possible to amplify the cytochrome b gene. This was revealed by the electrophoresis of the digested amplicons on a 3% agarose gels (Fig. 6.15) and 10% polyacrylamide gel. Amplicons for many of the species were not cut even at a single site by two of the three selected restriction enzymes. These findings were confirmed by using Restriction Mapper software. It was quite evident that the restriction sites for single cutter enzymes are at quite similar positions for some of the species. For example, *Bos grunniens* and *Hemitragus jemlahicus*, where the enzyme *RsaI* cuts at positions 272 for the former and 271 for the latter as such, may not be suitable to study the restriction fragment length polymorphism easily on a gel. Similarly the analysis of sequences of 16S rRNA amplicons for the species studied with Restriction Mapper software revealed that *Hae III* enzyme cuts most of the species at almost similar positions. It cuts *Antilope cervicapra* (at positions 26 and 94), *Cervus eldi thamin* (at positions 26 and 94), *Gazella bennetti* (at positions 26 and 94), *Moschus fuscus* (at positions 27

and 95), *Tetracerus quadricornis* (at positions 24 and 92) and *Cervus duvauceli* (at positions 27 and 95). Finally, *RsaI* does not cut any of the amplicons for 16S rRNA for the species studied as the sequences lack the restriction site acted upon by *RsaI*.

In case of 12S rRNA amplicons, the analysis of sequences with Restriction Mapper revealed that Hae III cuts all the amplicons for species studied at very similar positions. *Bos gaurus* (at position 357), *Bos grunniens* (at position 355), *Tetracerus quadricornis* (at position 355), *Gazella bennetti* (at position 356), *Moschus fuscus* (at position 356) and *Bubalus bubalis* (at position 358). Thus it may not be of use in distinguishing between species studied as it cuts species like *Moschus fuscus* and *Gazella bennetti* and *Tetracerus quadricornis* and *Bos grunniens* at exactly the same positions. *Rsa I* does not cut 12S rRNA amplicons for *Antilope cervicapra*, *Gazella bennetti* and *Hemitragus jemlahicus*. For the rest of the species, it cuts the amplicons only once except for *Tetracerus quadricornis* and *Bubalus bubalis*, where it cuts twice at positions 24 and 199 for the former and at positions 26 and 178 for the latter.

It is apparent that *HaeIII*, *AluI* and *RsaI* cannot be used for distinguishing all the species studied on the basis of restriction fragment length polymorphism of PCR amplicons. Analysis of sequences with Restriction Mapper revealed that some enzymes may be useful for distinguishing between the species studied on the basis of PCR-RFLP. For cytochrome b gene amplicons (398 bp approx.), the restriction enzyme CviJI may be useful for distinguishing the species studied using PCR-RFLP. The enzyme is a blunt cutter and cuts all the cytochrome b amplicon sequences for all the species studied. Similarly, few other restriction enzymes can also be used to distinguish between the studied species. These include DdeI, TspEI and MnlI; however, these enzymes do not cut cytochrome b gene amplicons (398 bp approx.) for all the species. DdeI does not cut amplicons for *Antilope cervicapra* and *Gazella bennetti*, TspEI does not cut the amplicons for *Pseudois nayaur* and MnlI does not cut the amplicons for *Bos grunniens.* Though a single enzyme like CviJI may be useful to distinguish all the species, when a combination of enzymes is used, the discriminatory power increased. In case of cytochrome b (398 bp fragment) with the combination of the CviJI and DdeI, the species studied could be distinguished.

Similarly for 16s rRNA amplicons, most of the restriction enzymes cannot be used individually for distinguishing all species based on PCR-RFLP of 16S rRNA amplicon. However with a combination of CviJI and MseI, it was possible to distinguish between the species. Finally 12S rRNA analysis with Restriction Mapper revealed that CviJI restriction enzyme is one of the best blunt cutters for the PCR-RFLP studies. Though it can be used to distinguish many of the species, however, for some of the quite unrelated species like *Gazella bennetti* and *Moschus fuscus*, it cuts the amplicons at identical positions. This can lead to serious error in identifying species when a multitude of species are suspected. However combination of restriction enzymes worked satisfactorily to differentiate amongst the species studied. With a combination of CviJI and TspEI, it was possible to differentiate the studied species.

6.4 Conclusion

Extraction of keratin and keratin-associated proteins (KAPS) from hair samples should be carried out in accordance with the electrophoresis method to be employed in subsequent keratin typing. Rather than using a universal protocol for keratin extraction, in which the extract can be used for both SDS-PAGE and IEF, it was worthwhile to use two different protocols. For SDS-PAGE studies, it is concluded to use the sodium dodecyl sulphate-containing extraction buffer as it produces bands which are more intense as compared to bands appearing with extraction buffer that is used for extraction of keratins for IEF [8–11]. Another advantage of this method is that it takes very less time (30 h approx.), and the extract can be directly used for SDS-PAGE after centrifugation and heat denaturation. Perhaps the best advantage of the method is that it is a single-step extraction, where, as in case of SDS-free extraction buffer used for IEF, SDS-containing sample buffer has to be added to denature the extracted keratins after centrifugation. For IEF studies, the extraction time can be reduced by almost 20 h to about 28 h instead of 48 h, as recommended in the previous studies [8–11]. There was no difference in IEF-based hair keratin patterns obtained after 28 h (modified) and 48 h of extraction as seen in case of *Pseudois nayaur*. Staining of SDS-PAGE gels can be best carried out effectively with Coomassie staining (Coomassie Brilliant Blue R 250) and conventional silver staining, and sensitive silver diamine staining is not suitable to stain SDS-PAGE gels for keratin patterns. As Coomassie staining is time consuming (overnight), the process can be made faster if heated in an oven. Staining and destaining can be hastened by keeping the staining or destaining solution at 40–50 °C. For IEF studies, the low amount of sample (0.25 μl) which is loaded to the gel is a limitation as the concentration of keratin falls below the detection limit of Coomassie stain. Thus silver staining is most effective to detect focussed keratins on an IEF gel. Instead of conventional silver staining, sensitive silver diamine staining is much more effective to detect the bands without excess of background staining. Hair keratin patterns based on SDS-PAGE and IEF show intraspecies consistency and interspecies variation. The hair keratin patterns were consistent and highly reproducible under a standardised set of conditions. Individuals from the same species showed similar bands, whereas individuals of different species showed differences in the hair keratin pattern with respect to position and number of bands. Thus species can be effectively characterised from hair samples on the basis of keratin patterns. However, much care has to be taken as protein techniques are very sensitive to slightest changes in running conditions like pH, temperature, current/voltage, etc., and therefore a great deal of precaution needs to the taken to achieve consistently reproducible results.

In species identification/characterisation from hair through DNA-based techniques, the DNA extraction process is the most vital. The success of subsequent PCR amplification depends entirely on the quality of DNA extracted. Even if the quantity of the extracted DNA is less (observed in all the samples where extracted DNA could not be visualised using ethidium bromide stain), there can also be a

good success in PCR if the quality of extracted DNA is good (i.e. it is free of PCR inhibitors). Commercially available DNA extraction kits like Qiagen DNeasy® tissue kit (Qiagen, Germany) (with modification), BioRobot EZ1® (Qiagen, Germany) and Geneclean® kit (Qbiogene, USA) are far much better than the phenol-chloroform, Chelex and alkali digestion method. Amongst the commercial kits, modified Qiagen DNeasy® tissue kit (Qiagen, Germany) and BioRobot EZ1® (Qiagen, Germany) have the best success rate with respect to the success in PCR amplification. Qiagen DNeasy® tissue kit can be easily used for DNA extraction from hair by slight modification. BioRobot EZ1® has the advantage of eluting DNA in water, and hence it is possible to concentrate the DNA extract through lyophiliser. Qiagen DNeasy® tissue kit and BioRobot EZ1® were useful in extracting DNA from 5–12 hairs with a good success rate in PCR. Out of the three regions of mitochondrial DNA which were successfully sequenced, cytochrome b region has higher discriminatory power than 16S rRNA and 12S rRNA genes for species identification. With respect to the size of the amplicon, primers targeting fragments up to 540 base pairs can be successfully used for PCR amplification of DNA extracted from hair with the help of Qiagen DNeasy® tissue kit and BioRobot EZ1®. Primers targeting larger fragments like the one selected for D-loop region which targeted 1Kb fragment may not be of any use as there was not even a single amplification with this primer. Hence, it will be very practical to use primers that target regions approximately from 400 to 500 bp, as this will give good success rate for PCR and provide a sufficiently long fragment for a good sequence comparison and identification. In addition it would be worthwhile to use primers that target smaller DNA fragments. Sequencing of the amplified mitochondrial DNA amplicons and their comparison with the database of NCBI/GenBank is the most suitable and robust method for species identification when species identification has to be done through DNA. However, the availability of sequence database for the suspected species in the NCBI/GenBank database is a must for this. Hence it is imperative that workers from different regions of the globe provide DNA sequences of all possible flora and fauna of their regions to NCBI. A single enzyme was may not cut amplicons for a particular gene for all the species. Some enzymes like CviJI may work in case of cytochrome b gene, and similarly TspEI may work in case of 12S rRNA gene for species characterisation/identification. However mostly if a restriction enzyme cuts PCR amplicons for all the selected species for a particular gene, it cuts some of the species at very similar positions as such the restriction fragment length polymorphism may not be easy to detect on a gel. Further, for some unrelated species like *Moschus fuscus* and *Gazella bennetti* for 12S rRNA gene, enzymes like HaeIII, CviJI, AciI and TspDTI may cut at exactly the same position as such fragments of the same size are generated after restriction digestion. Such problems can be overcome by the use of a blend of restriction enzymes for PCR-RFLP studies.

However, keeping in view the wide spectrum of species in illegal trade, nothing can suffice sequencing of mitochondrial DNA and its comparison with reference sequences for species identification. Sequencing of mitochondrial DNA followed by BLAST search on NCBI/GenBank to find out the most probable species and

subsequent comparison with mitochondrial gene amplicon from the reference sample of the suspected species shall be the best approach for species identification.

Ethical Permission The research work was carried out as Ph.D work at the Wildlife Institute of India, and it was part of the mandate of wildlife forensic lab/cell of the Wildlife Institute of India to create protocols for the species identification from wildlife parts and artefacts in illegal trade. All the samples used were from the repository of wildlife samples established at wildlife forensic lab under the mandate.

References

1. Akasaki T, Yanagimoto T, Yamakami K, Tomonaga H, Sato S (2006) Species identification and PCR-RFLP analysis of cytochrome b gene in cod fish (Order Gadiformes) products. J Food Sci 71:190–195
2. Annon (1996) Establishment of wildlife forensic capacity at wildlife Institute of India. WII-US-FWS project report. Wildlife Institute of India, Dehradun
3. Bartlett SE, Davidson WS (1991) Identification of Thunnus Tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome b genes. Can J Fish Aquat Sci 48:309–317
4. Bartlett SE, Davidson WS (1992) FINS (forensically informative nucleotide sequencing): a procedure for identifying the animal origin ofbiological specimens. Biotechniques 12:408–411
5. Block RJ (1951) Chemical classification of keratins. Ann N Y Acad Sci 53:608–612
6. Bolstein D, White DL, Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–333
7. Budowle B, Acton RT (1981) A technique for the detection of variable electrophoretic patters of hair proteins. Electrophoresis 2:333–334
8. Butler DJ, Deforest P-R, Koblinsky L (1990) The use of isoelectric focusing to identify Rhinoceros keratins. J Forensic Sci 35(2):336–344
9. Carracedo A, Concheiro L, Requena I (1985) The isoelectric focusing of keratins in hair followed by silver staining. Forensic Sci Int 27:237–245
10. Carracedo A, Prieto JM, Concheiro L, Estefania J (1987) Isoelectric focusing patterns of mammalian keratin. J Forensic Sci 32:93–94
11. Carracedo A, Andrade-Vide C, Rodriguez-Calvo MS, Lareis MV (1988) Fast isoelectric focusing of some polymorphic proteins and enzymes in miniaturized gels using an automated system. J Forensic Sci 33(6):1379–1384
12. Carrera E, Garcia T, Cespedes A, Gonzalez I, Fernandez A, Hernandez PE, Martin R (1999) Salmon and trout analysis by PCR-RFLP for identity authentication. J Food Sci 64:410–413
13. Chakraborty A, Aranishi F, Iwatsuki Y (2005) Molecular identification of hairtail species (Pisces: Trichiuridae) based on PCR-RFLP analysis of the mitochondrial 16S rRNA gene. J Appl Genet 46(4):381–385
14. Chan A, Chiang L-P, Hapuarachchi HC, Tan C-H, Pang S-C, Lee R, Lee K-S, Ng L-C, Lam-Phua S-G (2014) DNA barcoding: complementing morphological identification of mosquito species in Singapore. Parasit Vectors 7(1):569. https://doi.org/10.1186/s13071-014-0569-4
15. Dey SC (1996) Wildlife trade: global perspective and the Indian scenario. Cent Bureau Investig Bull 4:6–8
16. Domingo-Roura X, Marmi J, Ferrando A, Lopez-Giraldez F, McDonald DW, Jansman HAH (2006) Badger hair in shaving brushes comes from protected Eurasian badgers. Biol Conserv 128:425–430

17. Espinoza EO, Baker BW, Moores TD, Voin D (2010) Forensic identification of elephant and giraffe hair artifacts using HATR FTIR spectroscopy and discriminant analysis. Endanger Species Res 9:239–246
18. Fairbrother KS, Hopwood AJ, Lockley AK, Bardsley RG (1998) Meat speciation by restriction fragment length polymorphism analysis using ά- actin cCNA probe. Meat Sci 50:105–114
19. Folin M, Contiero E (1996a) Electrophoretic analysis of mammalian keratins. Anthropol Anz 54:331–339
20. Folin M, Contiero E (1996b) Electrophoretic analysis of non-human primates hair keratin. Forensic Sci Int 83(3):191–199
21. Gillespie JM, Marshall RC (1977) Proteins of the hard keratin of echidna, hedgehog, rabbit, ox and man. Aust J Biol Sci 30:401–409
22. Girish PS, Anjaneyelu ASR, Visvas KN, Shivakumar BM, Anand M (2004) Sequence analysis of mitochondrial 12s rRNA gene can identify meat species. Meat Sci 66:551–556
23. Girish PS, Anjaneyelu ASR, Visvas KN, Shivakumar BM, Anand M, Patel M (2005) Meat species identification by polymerase chain reaction-restriction fragment length polymorphism of mitochondrial 12s rRNA gene. Meat Sci 70:107–112
24. Graffy EA, Foran DR (2005) A simplified method for mitochondrial DNA extraction from head hair shafts. J Forensic Sci 50:1–4
25. Guha S, Kashyap VK (2005) Development of novel heminested PCR assays based on mitochondrial 16S rRNA gene for identification of seven pecora species. BMC Genet 6:42
26. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl Acids Symp 41:95–98
27. Harris H (1969) Genes and enzymes. Proc R Soc (Biology) 174:1–31
28. Higuchi R, Beroldingen CHV, Sensabaugh GF, Erlich HA (1998) DNA typing from single hairs. Nature 332:543–546
29. Indian Wildlife (Protection) Act-1972. (with amendments) a hand guide with case law and commentaries. Natraj Publishers, Dehradun, pp 1–215
30. Judd RC (1996) SDS–Polyaccrylamide gel electrophoresis of peptides. In: Walker JM (ed) The protein protocols handbook. Humana Press, New Jersey
31. Khamnamtong B, Klinbunga S, Menasveta P (2005) Species identification of five Penaeid shrimps using PCR-RFLP and SSCP analyses of 16S ribosomal DNA. J Biochem Mol Biol 38(4):491–499
32. Kocher TD, Thomas WK, Meyer A, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86:6196–6200
33. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of Baceriophage T4. Nature 227:680–685
34. Lee JC, Tsai LC, Yang C, Huang LH, Linacre A, Hsieh HM (2006) DNA profiling of Shahtoosh. Electrophoresis 27:3359–3362
35. Linacre A, Lee JC (2004) Species determination: the role and use of the cytochrome B gene. Methods Mol Biol 297:45–52
36. Malisa A, Gwakisa P, Balthazary S, Wasser S, Mutayoba B (2005) Species and gender differentiation between andamong domestic and wild animals using mitochondrial and sex-linked DNA markers. Afr J Biotechnol 4:1269–1274
37. Mani MS (1974) Ecology and biogeography in India. Dr. Junk Publisher, The Hague, p 727
38. Marshall RC (1981) Analysis of the protein form single wool fiber by two dimensional polyacrylamide gel electrophoresis. Text Res J 51:106–108
39. Marshall RC (1983) Characterization of the protein of human hair and nail by electrophoresis. J Investig Dermatol 80:519–524
40. Marshall RC, Gillespie JM (1982) Comparison of samples of human hair by two-dimensional electrophoresis. J Forensic Sci Soc 22(4):377–385
41. Marshall RC, Gillespie JM, Klement V (1985) Methods and future prospects for forensic identification of hairs by electrophoresis. J Forensic Sci Soc 25(1):57–66

42. Meyer R, Hofelein C, Luthy J, Candrian U (1995) Polymerase chain reaction – restriction fragment length polymorphism analysis: a simple method for species identification in food. J AOAC Int 78:1542–1551
43. Mitchell SE, Cockburn AF, Seawright JA (1993) The mitochondrial genome of Anopheles quadrimaculatus species a: complete nucleotide and organization. Genome 36:1058–1073
44. Mukherjee SK (1996) Some thoughts on wildlife trade. Cheetal 2:30–33
45. Nameer PO (2000) Checklist of Indian mammals. Kerala State Forest Department and Kerala Agricultural University, pp 1–90
46. Oakely BR (1980) A simplified ultra sensitive silver stain for detecting proteins in polyacrylamide gels. Anal Biochem 105:361–363
47. Paolella S, Bencivenni M, Lambertini F, Prandi B, Faccini A, Tonetti C, Vineis C, Sforza S (2013) Identification and quantification of different species in animal fibres by LC/ESI-MS analysis of keratin-derived proteolytic peptides. J Mass Spectrom 48(8):919–926
48. Parson W, Pegoraro K, Niederstatter H, Foger M, Steinlechner M (2000) Species identification by means of the cytochrome b gene. Int J Legal Med 114:23–28
49. Pepe T, Trotta M, Di Marco I, Cennamo P, Anastasio A, Cortesi ML (2005) Mitochondrial cytochrome b DNA sequence variations: an approach to fish species identification in processed fish products. J Food Prot 68:421–425
50. Pfeiffer I, Volkel I, Taubert H, Brenig B (2004) Forensic DNA-typing of dog hair: DNA-extraction and PCR amplification. Forensic Sci Int 141(2–3):149–151
51. Pilli E, Casamassima R, Vai S, Virgili A, Barni F, D'Errico G, Berti A, Lago G, Caramelli D (2014) Pet fur or fake fur? A forensic approach. Investig Genet 5(1):7
52. Rodriguez-calvo MS, Carraceelo A, Muoz I, concheiro L (1992) Isoelectric focusing of human hair keratin: correlation with SDS-PAGE and effect of cosmetic treatment. J Forensic Sci 37(2):425–431
53. Sahajpal V, Goyal SP (2005) The use of SDS-PAGE for identification of Shahtoosh and other wools. Proceedings of the 11th international wool research conference, University of Leeds, U.K
54. Sahajpal V, Goyal SP (2010) Identification of a forensic case using microscopy and forensically informative nucleotide sequencing (FINS): a case study of small Indian civet (Viverricula indica). Sci Justice 50(2):94–97
55. Sahajpal V, Singh K, Goyal SP, Thakur V (2009) Dealing wildlife offences in India: role of the hair as physical evidence. Int J Trichology 1(1):18
56. Sahajpal V, Goyal SP, Thakar MK (2012) Forensic examination of hair of protected Indian wildlife species: microscopy, protein and DNA methods. Lambert Academic Publishing, Germany, pp 1–256
57. Schagger H, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166:368–397
58. Shields GF, Kocher TD (1991) Phylogenetic relationships of north American ursids based on analysis of mitochondrial DNA. Evolution 45:218–221
59. Sironi M, Bandi C, Novati S, Scaglia M (1997) A PCR-RFLP method for the detection and species identification of human microsporidia. Parasitologia 39(4):437–439
60. Switzer CR, Merril CR, Shilfrin S (1979) Anal Biochem 98:23–237
61. Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
62. Verma SK, Singh L (2003) Novel universal primers establish identity of an enormous number of animal species for forensic application. Mol Ecol 3:28–31
63. Wolf C, Rentsch J, Hübner P (1999) PCR-RFLP analysis of mitochondrial DNA: a reliable method for species identification. J Agric Food Chem 47(4):1350–1355
64. Zafarina Z, Panneerchelvam S (2009) Analysis of hair samples using microscopical and molecular techniques to ascertain claims of rare animal species. Malays J Med Sci: MJMS 16(3):35–40

Chapter 7
Molecular Basis of Identification Through DNA Fingerprinting in Humans

Moumita Sinha, I. Arjun Rao, and Mitashree Mitra

Abstract DNA, the operating element of the genes, brings the coded notes of inheritance in every single surviving thing: animals, plants, bacteria, and other microorganisms. Within human beings, the information-bringing DNA arises in each cell having nucleus, including cells surrounding hair roots, spermatozoa, white blood corpuscles, and salivary cells. These would be the cells of utmost significance in forensic investigations. DNA testing has countless prospective advantages for civil and criminal justice; in spite of this, because of the likelihood for its mishandling or abuse, vital issues have been mentioned about trustworthiness, authenticity, and privacy. The methods of DNA testing are results of the innovation in molecular biology that is generating an increase of knowledge about human genetics. The greatly personal and complex info that can be developed by DNA testing involves firm and meticulous knowledge of genetic basis of testing methods.

Keywords DNA fingerprinting · Chromosome · STRs · VNTRs · RFLP · PCR · Genetics

7.1 Introduction

Science of biological variation is genetics which is the underlying core of genetics and the quintessence of Mendel's breakthrough in 1865 is that inheritance is particulate and that the visible traits determined through inherited factors, i.e. genes, exist in allelic pairs (i.e. the alternative forms of a gene at a particular site). These allele forms are present on each chromosome which are inherited from the father and mother on respective chromosomes. Chromosomes containing genes are present in the nucleus of a cell in the form of thread or rod-like structures. The

M. Sinha (✉) · I. Arjun Rao
Department of Forensic Science, Guru Ghasidas University, Bilaspur, Chhattisgarh, India

M. Mitra
School of Studies in Anthropology, Pt. Ravishankar Shukla University, Raipur, Chhattisgarh, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_7

combination of alleles of an organism is unique to that individual which is called as the genotype of that individual [16]. The genotype of an individual governs its corresponding phenotypes. A fragment of DNA not only harbours the genes responsible for the characteristic phenotypical feature of the organism but also contains many noncoding regions [17]. Many markers, the identifiable locations on a chromosome, have been designed in the noncoding region of DNA which are of wide use in forensic science and criminal justice system [5]. The DNA markers for forensic DNA typing have been designed in such a way that they are not a part of the expressed genes and hence are not associated to any phenotype of the individual. The observed different traits within a genetic marker are called polymorphism [18]. Advancement in DNA technology in the past few decades has led to the detection of the variation in DNA sequence directly in the form of polymorphism, thus obsoleting the techniques used in pre-DNA era such as blood groups variation, serum proteins, and tissue HLA types [7].

7.2 Structure and Function of DNA

Nuclear DNA, found in all nucleated cells, is the hereditary unit of an organism. The DNA molecules are present in the form of a packed structure along with certain protein molecules, mostly the histone proteins, which are called as chromosomes. A typical human cell contains 22 pairs of non-sex chromosomes/autosomes and two sex chromosomes. For male individual the sex chromosomes are one X chromosome and another Y chromosome, whereas for female individuals, a pair of X chromosomes is found [12]. Each chromosome, which may be autosome or sex chromosome, is comprised of a long strand of molecule of DNA in the form of double-helical structure (Fig. 7.1). Each of the component strand of DNA consists of a chain of nucleotides designated as adenine (A), thymine (T), cytosine (C), and guanine (G). These nucleotides follow a pattern of bonding between them as ‘A’ can pair only with ‘T’ by a double bond, whereas ‘G’ can bond with ‘C’ through a triple bond. The bond formation between these nucleotides occurs among the two strings of DNA twist in pecking order. In this case, a pair of complimentary nucleotides is called base pairs [14].

A DNA molecule possesses a unique property of self-replication during the S phase of the cell cycle [15]. During this process of DNA replication, two strings of a double coil DNA split-up followed by the creation of a new string is complementary to each string. A gene is a segment of DNA macromolecule that comprises the blueprint for the organization and function of a cell as well as the organism. A human contains 50,000–100,000 genes harbouring in each nucleated body cell, out of which chromosome 1 contains the highest numbers of genes [21].

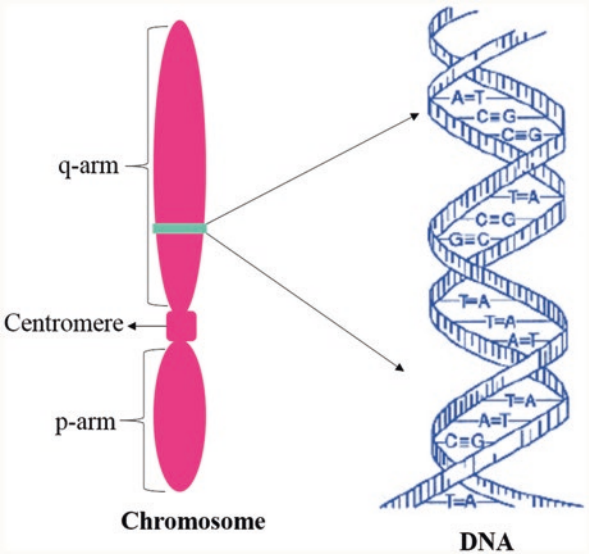

Fig. 7.1 Schematic presentation of the double-helical structure of DNA in a chromosome

7.3 Individualistic Variation in DNA Sequence

The human genome has DNA tools that expose differences in the complete genetic architecture of the affiliates of a species: single-nucleotide variations, insertions, and deletions. The noncoding segments of DNA, which are not as much of pressure by forces of selection, it is likely that an average of no less than one nucleotide per 300–1,000 differs among two individual [4].

These differences in nucleotide could alter the detection site for an individual site-restricted endonuclease so as to sustain the DNA have being excise at that position by that enzyme for example, in Fig. 7.2 the annotation that a single change in nucleotide from C to T has abolished a cutting site of restriction endonuclease. Additionally, certain sections of DNA comprise repetitive elements, several alike series of nucleotides organized in tandem. From person to person, VNTRs (variable number tandem repeats), the repetition numbers of an order, can alter. Currently in forensic DNA typing, VNTRs are the most frequent form of variation used. The restating unit can be as minute and as large as 30, or even more, nucleotides or (TG)n polymorphism as a dinucleotide. Inside the noncoding segments of DNA, tandem repeats are not limited to, while they are less recurrent in coding segments of DNA. When DNA is incised with a restriction endonuclease, they potentially recognize change in the lengths of fragments for the two main types of alterations—single-nucleotide varia-

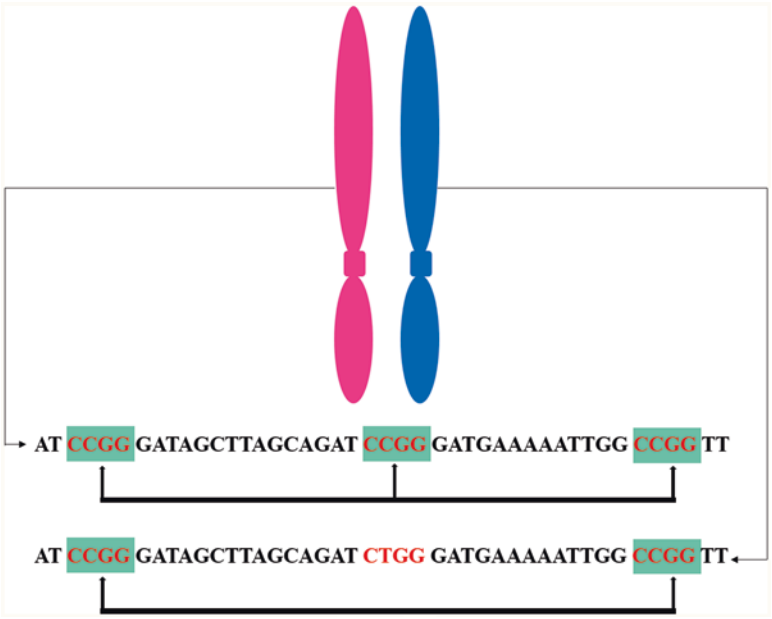

Fig. 7.2 Fundamentals of restriction fragment length polymorphism (RFLP). Figures indicating heterozygosity at a particular restriction site; CCGG sequence present on left chromosome is recognized and cut by specific restriction enzyme, whereas single nucleotide modification results in the formation of CTGG, which is not recognized by the enzyme

tions and VNTRs. The differences in the parts of sections can result from an alteration in the gathering of 4, 5, or 6 nucleotides that is the restricted incising position of the specific restriction enzyme (Fig. 7.3), or the alteration can arise, not from a transformation in the excising position of the endonuclease but from the presence of varied figures of tandem repeats among two incised locations. Figure 7.3 deciphers the key features of the two types of alteration and their utilization in DNA typing.

In Fig. 7.2, alteration at the incising position of a restriction enzyme can product in two different allelic forms: the enzyme nicks or it does not which created diallelic system causes three possible genotypes are shown. If someone obtained the identical allelic form at a specific chromosomal location from both parents, the person's genotype is supposed to be homozygous for that allelic form; if unlike allelic forms were received from each parent, the person held to be heterozygous at that chromosomal location. In the same way as shown in Fig. 7.3, when the alteration is in the number of tandem repeats, there can be many allelic forms, of which a particular person can have only two. This is called a multiallelic system. The count of possible genotypes is the figure of the assured numbers from 1 to the number of alleles like a three-allele system, as shown in Fig. 7.3, which are expected to be 6 genotypes (1 + 2 + 3). The formula n (n + 1)/2 is a representation of the number of genotypes, where n is the number of alleles. Inheritance of alteration in the noncoding sections

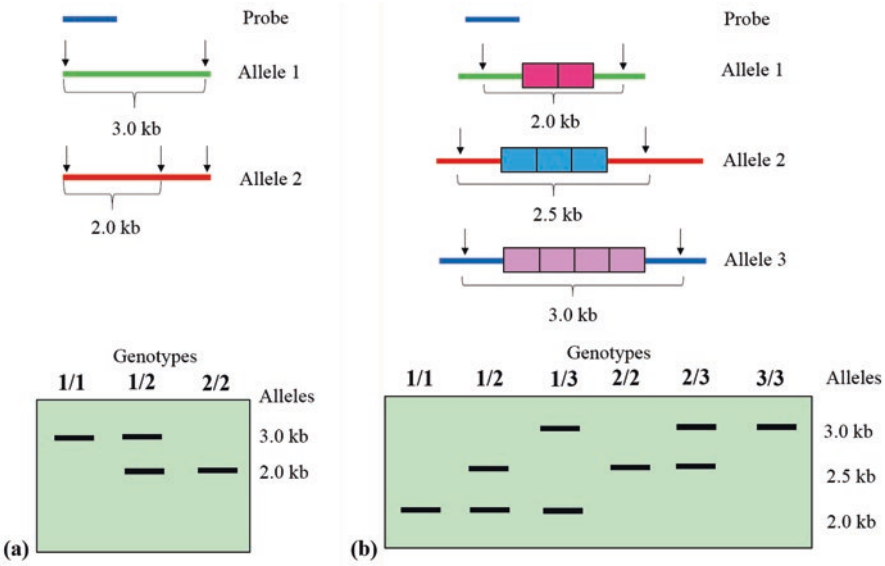

Fig. 7.3 Two types of RFLPs. Structure of alleles with arrows indicating the respective cutting sites of the enzyme. Electrophoretic patterns (**a**) diallelic system of RFLP rising from single nucleotide change and (**b**) multiallelic VNTR system with three alleles and six possible genotypes as demonstrated by electrophoresis

of DNA obeys the similar rules that Mendel concluded for expressed genes. A particular individual receives one of the alleles from father's two alleles and one of the alleles from mother's two alleles. Once the two alternate sites, for each on a separate chromosome, are studied, the inheritance at one location is neutral of that at the other. If they are sufficiently far apart, then the two sites, even on the same chromosome, could also be passed on neutrally. While if the two sites on the same chromosome are very close, then the event of linkage disequilibrium occurs—a departure from neutral inheritance in which specific alleles at the two sites manage to be transferred jointly.

7.4 Technical Basis of DNA Typing

Forensic typing of DNA customarily includes matching 'DNA sample', i.e. DNA isolated from biological substances left behind at a crime scene together with 'questionable DNA' (i.e. DNA isolated from the suspected blood sample). The techniques of DNA typing involve restriction enzyme digestion, electrophoretic analysis, probing and the polymerase chain reaction (PCR) [6, 13, 20].

7.4.1 Restriction Fragment Length Polymorphism (RFLP)

The RFLP is the approach in which DNA is exposed to restriction enzymes which cuts double-stranded DNA at specific recognition sites and creates fragmentation in regulated manner. The long-stretched DNA fragments are in that way summarized to an imitable set of small portions called restriction fragments (RFs), which are generally some hundred to thousand base pairs elongated.

Along with a separate restriction enzyme, individual fragment of DNA has a different order and extent. For examination of restricted fragments to determine RFLPs, the fragmented portions are parted in electrophoresis on the base of size. Electrophoretic analysis, usually achieved on acrylamide or agarose gels, is observed in larger fragments at one end and smaller fragments at the other, the smaller fragments moving farthest in the electric field. The fragment portions are then denatured as single-stranded DNA (ssDNA) and defused and moved from the gel to a nylon membrane, in which they are fixed; this enables recognition of particular RFLPs and VNTRs. RFLPs that are marked by particular sequences are revealed by hybridization with a probe, a smaller fragment of single-stranded DNA labelled with radioactive phosphorus, that is employed to determine a specific sequence of complementary DNA. The nylon membrane is employed in a bath that comprises the probe, and the probe hybridizes to the focused denatured restricted fragments. Non-tagged attached probe is cleansed off. Hybridized probes are traditionally being tagged with radioactive isotopes, but consideration is gradually being specified to non-isotopic tagging. When isotopically tagged probes are utilized, the array of probe attaching is imagined with autoradiography.

The whole procedure, digestion of DNA, electrophoresis analysis, transfer to membrane, and hybridization with probe, was established in 1975 by Edwin Southern; [19] and the currently modified form still in use is described to as Southern blotting. These techniques are normally utilized in genetics, molecular biology, biochemistry, and clinical DNA analysis; there is no differentiation in their application in forensic. Differentiations between persons are stated as variances in the lengths of restricted fragments [1]. These RFLPs results from various manners of distinctions at the genomic level.

The mutations that modify the sequence of base at a recognition site of restriction enzyme or incision spot can give rise to a deficit of the incision site or the creation of an incision site that did not exist previously. RF lengths also changes with insertion or deletion of nucleotides between two incision sites. Alteration of these kinds is usually related with a minor frequency of alleles such as, the deficit or acquire of a specific incision location possibly will be responsible for only two alleles.

Certain segments of DNA hold several portions of repeats of short sequences. Subsequently, there is a category of RFs that vary in the numeral of repeated portions exist. Particular VNTR polymorphisms have a minor frequency of alleles, and the arrays of RFs that characterize each of the alleles at a specified locus can be voluntarily distinguished. On the other hand, extremely polymorphic VNTR loci have 50–100 alleles or still more. In that condition, the sharing of RF size is effectively uninterrupted; alleles with RFs nearby in size could not be determined

with electrophoresis, and the constraint of perseverance must be demarcated effectively. The widespread alterability, the VNTR grade of RFLPs has confirmed its high utility in differentiating among individuals.

The RFLP examination with single-locus probes is commonly suggested to be brought about by simple array of one or two RFLP bands, liable on whether the individual is homozygous or heterozygous, correspondingly. The extent of alteration presented in the arrays from different individuals depends on how numerous various alleles occur at the specific target locus, e.g. how several distinctive tandem repeats are in the population as an entity (Fig. 7.4). A substitute to the usage of a single-locus probe is to practicing the multilocus probe that hybridizes to several distinctive VNTR spots in the genome. The resultant arrays in a single individual comprise various bands of diverse intensity; the arrays have been matched with barcodes. The method was established by Jeffreys and colleagues [8–11]. As the difficulty of the arrays, explanation can be challenging. Consequently, the use of multiple single-locus probes is preferred.

7.4.2 Polymerase Chain Reaction (PCR) for Amplification of DNA

PCR or polymerase chain reaction is a technical procedure in molecular biology to augment a target region, single or a few replicas of a fragment of DNA through numerous series of extent, produce thousands to millions copies of an exact sequence of DNA. Polymerase chain reaction was given by American biochemist, Kary Mullis, in 1984. Mullis was rewarded with a Nobel Prize and Japan Prize for his invention in 1993. PCR in present day is popular and a frequently needed method used in biological and medical research labs for a range of wider applications. The PCR is an influential technical method that has speedily turned out to be one of the most extensively used technical methods in molecular biology, since it is fast, rapid, economical, and trouble-free. The method involves amplification of particular DNA fragments from microscopic amounts of DNA source material, even when the source DNA is of extreme poor quality.

7.4.2.1 Basic Concept of PCR

The key principle of PCR is easy. PCR is abbreviated for Polymerase Chain Reaction, in which one part of DNA molecule is employed to yield multiple copies through constant amplification achieved by particular proteins known as polymerase enzymes that are capable of twining together individual DNA nucleotide sequences to form elongated molecular strings. To perform this, polymerases entail a resource of DNA-assembling material, i.e. the nucleotides, comprising of four bases: adenine (A), guanine (G), thymine (T), and cytosine (C). It also demands a minor piece of DNA, identified as primer, where they fix the nucleotides as well as an extended

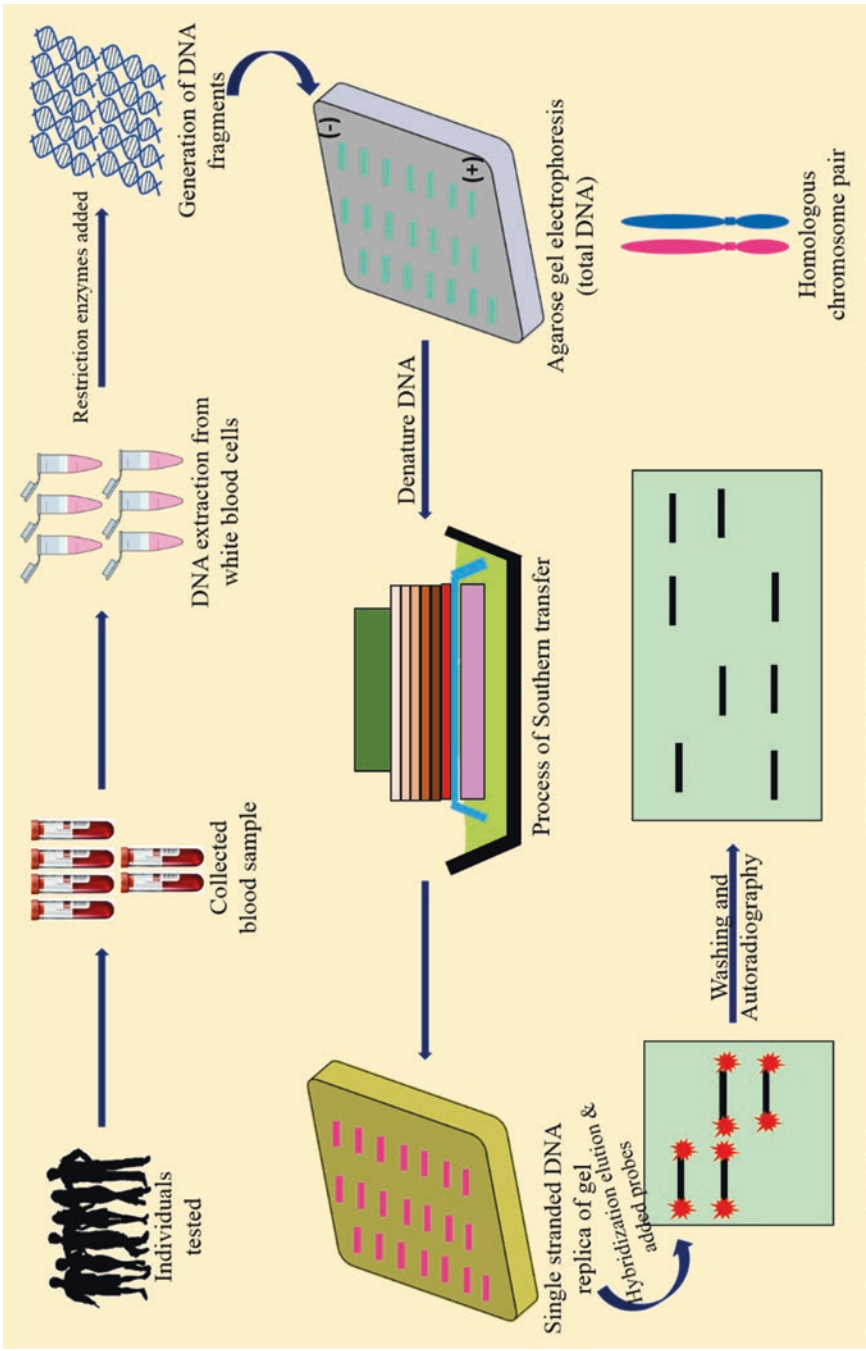

Fig. 7.4 Schematic representation of Southern blotting of single-locus, multiallelic VNTR

molecule of DNA to provide a template for building the original string. If these three components are provided, the enzymes will expand the precise replicas of the DNA templates. PCR is a technique employed to obtain numerous replicas of any exact string of DNA molecule. It's a way of carefully intensifying an exact fragment of DNA. The fragment may symbolize a tiny portion of a substantial and complex mixture of DNAs.

7.4.2.2 Procedure of PCR

The procedure comprises three major phases: denaturation, annealing, and extension. This process efficiently amplifies the DNA amount all through these phases in the PCR cycle. To amplify a small fragment of DNA using PCR, the fragment is first heat up at certain temperature so the DNA denatures or detaches into two different pieces of single-stranded DNA. In the next process, an enzyme called 'Taq polymerase' creates or constructs two new DNA strands, using the master strands as templates. This process results in the amplification of the master DNA strand, with a portion of the new molecules comprising one former and one fresh DNA strand. Then strands individually can be used up to build two fresh copies. The annealing stage takes place at a lower temperature, 50–60 °C, which permits the primers to hybridize to their particular complementary DNA template strands. The newly produced strand of DNA (of the primer) attached to the template is then used to create similar copies of the new template strings required. Taq polymerase adjoins up the obtained building blocks to the last part of the primers annealed. The lengthening of the primers by Taq polymerase takes place at approximately 72 °C for 2–5 min. DNA polymerase cannot be utilized for elongation of primer as it is expecting that it is not steady at higher temperatures required for the PCR. The magnificence of the PCR cycle and procedure is that it is extremely fast, when judged with another molecular biology methods, and every single cycle multiplies the sum of replicas of the required fragment of DNA (Fig. 7.5).

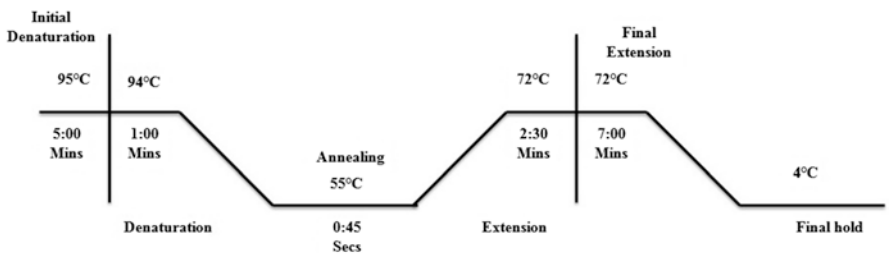

Fig. 7.5 Basic steps in polymerase chain reaction

7.5 Features of an Optimized Forensic DNA Typing System

The techniques of DNA testing stay to go forward as novel means to reveal individual alterations. DNA sequencing possibly will be the prime way of personal identification, although that is still now faraway from practical. It is significant that the suppleness to take on novel techniques be maintained as calibration of DNA technology is established and databanks are generated. Any technique of forensic DNA testing, like processes for medical DNA and additional testing, should be fast, precise, and reasonable in price. Furthermore, to accomplish highest discrimination between persons, forensic DNA testing needs the utilization of markers with a great level of polymorphism. Preferably, the great degree of alterability would be established in all the populations of the world. The probes and the markers utilized to determine them must have an innovative sequence, so that individual probe hybridizes with specific part of the genome. Probes of single-locus had better to be employed. The chromosomal location of the markers must be on distinct chromosomes, i.e. they should be independent. The DNA markers should also originate from non coding part of DNA and thus probably nonfunctional portions of the genome and also to elude allegations, specious or else, of relation of specific markers with particular behavioral characters or diseases. The computerization of DNA testing could facilitate to minimize the time and cost.

The benefit of rapid and cost-effective technique is one that can examine other portions of the genome. Similarly, if a locus is simply humbly polymorphic, its usage in DNA testing might have added advantages, such as comprehensive obviousness of scoring; utilized in arrangement, such loci could reveal that the probability of a random compatibility is very low. It should be highlighted that novel techniques and expertise for determining uniqueness in each individual's DNA continue to be progressed. The techniques summarized in this chapter are probable to be replaced in efficacy, computerization, budget, and other qualities by new technologies. Precaution should be procured to confirm that DNA testing methods used for forensic applications do not turn out to be 'inaccessibly' too early. Or else, people and the criminal justice organization will not be competent to develop utmost advantage from developments in the science and technology [3].

7.6 Future Perspectives

Refining current techniques through scaling of STR typing will be a foreseeable way that forensic biology will follow as a trend seen in many other industries. This will allow a greater quantity output; conserves reagents, resources, and time; and thus reduces the inclusive expenses. Points at which this can be attained are at the stage of the PCR by means of miniature thermal cyclers, which are presently advanced. A minor thermal cycler utilizes less reagents but also accomplishes its cycles faster due to rapid heating and cooling. Additional means to boost output is using capillary

array electrophoresis instruments (CAE). Capillary electrophoresis is nowadays achieved sequentially instead of parallel. The utilization of the currently improved 96-array capillary electrophoresis systems for STR typing has previously been described to be efficacious, and it is probable that this CAE will be employed gradually. One more approach to pace up DNA separations is using microchip CE devices. These chips are regular 30 cm capillary that is substituted by one of several cm in length, ensuing in a 10–100 quicker separation. When these chips are analogous in a microplate of 96-channel, a greater output can be attained [2]. In the future, new applications within forensic biology like Matrix-assisted laser ionization time-of-flight mass spectrometry (MALDI-TOF MS), use of ethnic markers, SNPs, and markers for commonplace characteristics can improve the future perspectives of DNA testing.

References

1. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331
2. Butler J (2005) Forensic DNA typing: biology, technology, and genetics of STR markers. 2nd ed. Academic Press, Amsterdam, p 688, ISBN: 9780080470610
3. Committee on DNA Technology in Forensic Science, National Research Council (1997) DNA Technology in Forensic Science. National Academy of Sciences, Washington, DC
4. Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J (1985) An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet 69:201–205
5. Fan H, Chu JY (2007) A brief review of short tandem repeat mutation. Genomics Proteomics Bioinformatics 5:7–14
6. Farley MA, Harrington JJ (1991) Forensic DNA technology. Lewis Publishers, Chelsea
7. Geserick G, Wirth I (2012) Genetic kinship investigation from blood groups to DNA markers. Transfus Med Hemother 39:163–175
8. Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA "fingerprints". Nature 318:577–579
9. Helminen P, Ehnholm C, Lokki ML, Jeffreys A, Peltonen L (1988) Application of DNA "fingerprints" to paternity determinations. Lancet 1:574–576
10. Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific "fingerprints" of human DNA. Nature 316:75–79
11. Jeffreys AJ, Wilson V, Thein SL, Weatherall DJ, Ponder BAJ (1986) DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. Am J Hum Genet 39:11–24
12. Karkanaki A, Praras N, Katsikis I, Kita M, Panidis D (2007) Is the Y chromosome all that is required for sex determination? Hippokratia 11:120–123
13. Kirby LT (1990) DNA fingerprinting: an introduction. Stockton Press, New York
14. Lively TN, Jurema MW, Shields GC (1994) Hydrogen bonding of nucleotide base pairs: application of the PM3 method. Int J Quantum Chem 52:95–105
15. Nishitani H, Lygerou Z (2002) Control of DNA replication licensing in a cell cycle. Genes Cells 7:523–534
16. Orgogozo V, Morizot B, Martin A (2015) The differential view of genotype–phenotype relationships. Front Genet 6:179
17. Patrushev LI, Kovalenko TF (2014) Functions of noncoding sequences in mammalian genomes. Biochemistry 79:1442–1469

18. Roychoudhury AK, Nei M (1988) Human polymorphism genes: world distribution. Oxford University Press, New York
19. Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–527
20. U.S. Department of Justice, Federal Bureau of Investigation (1991) Proceedings of the international symposium on the forensic aspects of DNA analysis. U.S. Government Printing Office, Washington, DC
21. White PS, Sulman EP, Porter CJ, Matise TC (1999) A comprehensive view of human chromosome 1. Genome Res 9:978–988

Chapter 8
Genetic Fingerprinting for Human Diseases: Applications and Implications

Inusha Panigrahi

Abstract DNA fingerprinting traditionally refers to the identification of individuals from blood and/or tissue samples for forensic purposes. But genetic fingerprinting can also include characterization of the genetic basis of human diseases, especially the inherited disorders. Some of the variants or haplotypes identified may run in families and thereby also have pathological or phenotypic connotations. The DNA sequencing technologies have evolved over the years, and nowadays, high-throughput techniques and applications are available with increased automation. Thus, genetic fingerprinting can have various connotations in relation to human diseases. The genetic testing done would depend on the clinical situation or phenotype, and what we are looking for in a specific patient or individual. Pretest and posttest counseling are important to facilitate decision-making.

Keywords Genetic fingerprinting · Human diseases · Inherited disorders · Methylation · Mutations · Microarray · NGS

8.1 Introduction

DNA fingerprinting traditionally refers to the identification of individuals from blood and/or tissue samples for forensic purposes [1]. But genetic fingerprinting can also include characterization of the genetic basis of human diseases, especially the inherited disorders. Some of the variants or haplotypes identified may run in families and thereby also have pathological or phenotypic connotations [2–6]. The DNA sequencing technologies have evolved over the years, and nowadays, high-throughput techniques and applications are available with increased automation. The traditional Sanger sequencing using fluorescently labeled terminating nucleotides followed by gel electrophoresis remains the gold standard sequencing technology. The following write-up describes the genetic abnormalities underlying human diseases or human variation and their clinical applications with some illustrative cases.

I. Panigrahi (✉)
Genetic and Metabolic Unit, Department of Pediatrics, Advanced Pediatric Centre (APC), Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_8

II. Types of Genetic or Genomic Disorders The human diseases can result from various abnormalities in the genome. These are covered here under (1) VNTRs and disease (minisatellites and microsatellites); (2) genomic disorders and low-copy repeats; (3) disorders with abnormal methylation and imprinting; (4) genetic identification/characterization in human infections; (5) race, ethnicity, and founder mutations; and (6) chromosomal microarray and next-generation sequencing in genetic disorder diagnosis.

8.2 VNTRs and Human Diseases

These include minisatellites and microsatellite sequences in DNA. Minisatellite is a DNA sequence of size varying from 1 to 20 kb, and VNTR usually refers to sequences with repeat units of 9–80 bp, located in noncoding regions. Since this number differs from individual to individual, this is used for forensic purpose in identification [1]. The human telomeres also contain tendency repeated DNA sequence GGG TTA. The length of telomere becomes shorter with age. Microsatellites or short tandem repeat polymorphisms can result from DNA recombination during meiosis or may be result of replication slippage. The alphoid DNA at centromere has also repeat units and constitutes 3–5% of DNA or the chromosome. These repeat sequences across the genome do not encode a protein. Multiple VNTR studies have been reported in relation to human diabetes susceptibility. The eNOS VNTR polymorphism may contribute to microvascular complications of type 2 diabetes [2].

8.3 Genomic Disorders and Low-Copy Repeats

The term "genomic disorders" was proposed in 1998 and usually refers to disorders in which DNA rearrangements lead to loss or gain of a dosage-sensitive gene or lead to disruption of the gene. These mostly result from homologous recombination between regions of low-copy repeats (LCRs). The genomic disorders show high frequency of new mutations [3]. This is because of large regions of sequence similarity and instability at LCRs. These LCRs or segmental duplication regions are classically more than 1 kb in size. The LCRs contain genes, certain gene fragments, pseudogenes, and/or other nonallelic fragments delivered from duplication event. Disorders with abnormalities of LCRs include Charcot-Marie-Tooth disease (CMTD) type, Smith-Magenis syndrome, and Potocki-Lupski syndrome.

8.4 Disorders with Abnormal Methylation and Imprinting

The diseases with abnormal methylation and imprinting include Prader-Willi syndrome (PWS) and Angelman syndrome. Abnormal imprinting during human development can also lead to growth disorders like Beckwith-Wiedemann syndrome (BWS) and Silver-Russell syndrome (SRS). Around 150 candidate imprinted genes across 115 chromosome bands have been identified in the human genome. Apart from fetal growth and development, abnormalities in imprinted genes can affect language and behavior and contribute to cancers or malignancies [5, 6]. Insulin-like growth factor2 (IGF2) gene is involved in BWS, and the GRB10 gene regulating gene expression is involved or affected in SRS. The loss of paternally expressed genes or maternal uniparental disomy results in PWS. PWS presents with infantile hypotonia, feeding problems, early onset obesity, short stature, behavioral problems, and sleep disturbances. The genes or transcripts affected include SNURF-SNRPN, NDN, MKRN3, MAGEL2, and small nucleolar RNAs (snoRNAs). The candidate genes when inherited from mother are silenced but active when inherited from father. The maternally expressed genes in this region include UBE3A and ATP10C. The GRB10 gene implicated in SRS acts as a suppressor of growth and two maternal copies resulting from uniparental disomy (UPD) of this chromosome7 region also leads to SRS.

Mutations affecting imprinting center at 11p15 region can also lead to SRS phenotype, in around 60% patients. The chromosome 11p15.5 region contains a number of imprinted genes which have maternal or paternal expression. The abnormal methylation of H19 and IGF2 genes in this region results in BWS phenotype. BWS is an overgrowth syndrome with macroglossia, hypoglycemia, omphalocele, characteristic ear creases, and an increased risk of tumors like Wilms' tumor and hepatoblastoma. Imprinting disturbances have been reported following pregnancies conceived with the use of assisted reproductive technology (ART) or in vitro fertilization (IVF) pregnancies. These include the classical imprinting disorders BWS and PWS.

8.5 Genetic Identification/Characterization in Human Infections

Several techniques are used to identify the different pathogenic strains of bacteria and viruses, and typing of the isolates from blood or other body fluids like saliva, urine, synovial fluid, or bone marrow helps in proper characterization of the infection and development of appropriate treatment strategies. The PCR methods used

include randomly amplified polymorphic DNA (RAPD) fingerprinting, the enterobacterial repetitive intergenic consensus (ERIC)-PCR, repetitive extragenic palindromic (REP) PCR, and BOX repeat-based (BOXAIR) PCR. RAPD can be used to evaluate genetic and epidemiological relationships of community-acquired methicillin-resistant *Staphylococcus aureus* (MRSA) isolates [7]. Nowadays with widespread use of antibiotics and inappropriate antibiotic policies, there has been emergence of resistant organisms, which require more aggressive or alternative therapies. Carbapenem-resistant pseudomonas is considered one such organism or super bug. Methods for identification of these organisms are being developed which include PCR-based fingerprinting methods. The majority (94.4%) of the carbapenem-resistant isolates from a hospital in Costa Rica carried the intI1 gene, in contrast to 26% of carbapenem-susceptible isolates [8]. The isolates were divided into 29 clusters by RAPD analysis. The resistant isolates were overrepresented in samples obtained from surgical emergency and intensive care units. Resistance to carbapenems has also been reported in *Klebsiella pneumoniae*. A study from Romania identified that majority of these were carbapenemase-producing strains. The major genotype identified was OXA-48 (n = 51) out of 75 non-susceptible isolates [9].

Characterization of isolates by repetitive-sequence-based PCR and study of antimicrobial susceptibility helps in making judicious choice of antibiotic for clinical use. Presence of blaOXA-51 and blaOXA-23 in isolates contributed to imipenem resistance in *Acinetobacter baumannii* in a multicenter study [10].

8.6 Race, Ethnicity, and Founder Mutations

There is an intricate interlink between genetics, genealogy, and geography. The mtDNA makers and Y-chromosome-related DNA markers can help determine the lineage or ancestry of an individual. These are also some ancestry informative SNPs (AISNPs) and mini-haplotypes in which <10 SNPS span small molecular region (<10 kb) and can determine the ancestry of particular individual. The unusual non-recombinant mode of inheritance in Y and mtDNA markers is efficient for identification studies after mass disasters. The forensic Y-chromosome database is YHRD, and the largest mtDNA database is EMPOP. There are also population-specific DNA databases in some countries.

Some of the genetic diseases have a high prevalence in some populations like Gaucher disease and Tay-Sachs disease in the Ashkenazi Jewish population. Cystic fibrosis is common in European American population, whereas beta-thalassemia is common in Mediterranean population. Nowadays, in disorders involving multiple genes, the mutations can be tested by next-generation sequencing (NGS), and specific mutations can be identified or confirmed by Sanger sequencing (Fig. 8.1). Other techniques used include ARMS-PCR for beta-thalassemia, multiplex ligation probe amplification (MLPA) to detect deletions/duplications, and triplet-primed

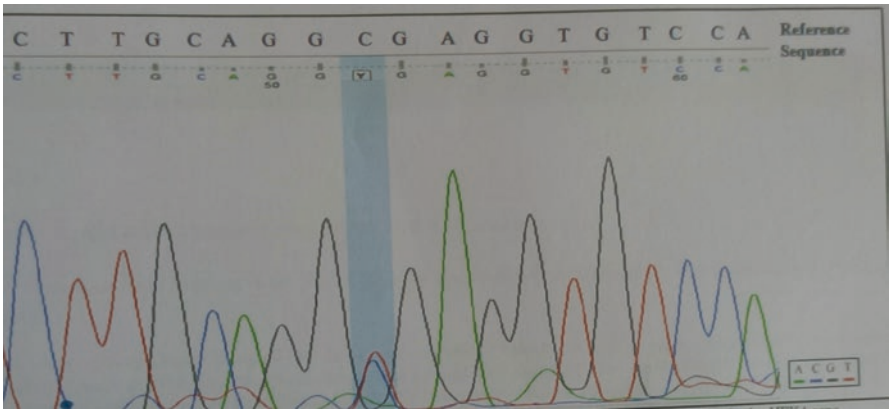

Fig. 8.1 Sanger sequencing graph of a Tay-Sachs disease carrier individual showing heterozygosity for the c.1528 C > T (p.Arg510Ter) mutation in *HEXA* gene

PCR for triplet repeat disorders. In India there is high frequency of inbreeding in ethnic groups and consanguinity in South Indian population. So, some specific mutations in some diseases achieve high prevalence, called founder effect. A study in the Narikuravar nomads of Tamil Nadu found a splicing mutation c.87 + 1 G > A responsible for alkaptonuria that abolishes the intron 2 donor splice site [11]. The Agarwal community has high prevalence of certain genetic diseases including megalencephalic leukoencephalopathy with cysts (MLC). A recent study found two founder mutations in the *CALPN3* gene in the Agarwal community, responsible for one form of limb-girdle muscular dystrophy (LGMD) [12]. The mutations are c.2338G > C (p.D780H), and the splice site mutation c.2099-1G > T mutation. We have also found a founder mutation 215_216 insA in *PANK2* gene, implicated in Hallervorden-Spatz disease, in Garg-Agarwal community. In another study, 6 patients from the Agarwal community harbored a mutation in *ALDOB* gene, of a total of 11 patients of hereditary fructose intolerance (HFI) in the study [13]. In beta-thalassemia, there are over 350 globin gene mutations that have been described worldwide. However, some common mutations account for majority of beta-thalassemia alleles in the endemic or high prevalence areas. The intronic mutation IVS 1–1 G > T is common in Punjabis from North India, and the promoter mutation -88C > T is common in Jat Sikhs [14]. Similarly the abnormal hemoglobins HbD and HbQ India are also seen predominantly in people of Punjabi origin. In Gaucher disease which presents with anemia and hepatosplenomegaly, the c.1448 T > C (Leu444Pro) mutation is common in Indian children with the disorder [15], and this has been associated with neurological symptoms of the disease. Knowledge of the common or founder mutations can help in planning molecular testing in affected patients in the particular ethnic group.

8.7 Chromosomal Microarray and Next-Generation Sequencing in Genetic Disorder Diagnosis

Advanced array technologies have revolutionized molecular diagnosis. Chromosomal microarray and CGH array are used for detection of chromosomal deletions/duplications in unexplained intellectual disability or autism. Next-generation sequencing (NGS) can simultaneously amplify multiple genes, and the genetic defect underlying phenotype can be identified in 6–8 weeks' time. This also enables identification of novel variants and new genes responsible for clinical phenotype. A thorough bioinformatics analysis using available online databases and/or indigenously developed software can help in characterizing the variants identified on NGS testing.

Family 1 A 5-year-old girl presented with mild developmental delay and abnormal eye movements and a family history of genetic disorder. She had partial ocular albinism, and there was history of proven oculocutaneous albinism (OCA)1 in paternal uncle. The uncle had c.1255 G > A or p. G419R mutation in OCA1A gene in homozygous state. She did not have any history of seizures, constipation, or any neurocutaneous markers. There was no white forelock and no deafness. The OCA1 and X-Linked OA mutation testing (Sanger sequencing) was negative in the child. Clinical exome sequencing was performed on DNA of the child, and the data was analyzed for variants in the genes *MC1R*, *OCA2*, *SLC45A2*, *TYR*, *TYRP1*, *GPR143*, *EDN3*, *SOX10*, *EDNRB*, *MITF*, *PAX3*, *SNAI2*, *EDNRB*, *SMOC1*, *GPR143*, *LYST*, *MC1R*, *MITF*, *MYO5A1*, *RAB27A*, *AP3B1*, *BLOC1S3*, *BLOC1S6*, *C10orf11*, *DTNBP1*, *GPR143*, *HPS1*, *HPS3*, *HPS4*, *HPS5*, *HPS6*, *LYST*, *OCA2*, *SLC24A5*, *SLC45A2*, *TYR*, *TYRP1*, and *SHEP4*. This revealed three variants in three genes. One was a stop-gain mutation c.823G > T in homozygous state in the *HPS3* gene with predicted deleterious effect. Another was heterozygous c.913C > T mutation in OCA2 gene and classified as variant of unknown significance (VUS), possibly polymorphic. The third one was in *MYO5A1* gene c.2635C > T in heterozygous state also classified as VUS but can be deleterious in homozygous state.

Thus, we identified a novel homozygous deleterious variant in *HPS3* gene in this family consistent with Hermansky-Pudlak syndrome (HPS). There are seven types of oculocutaneous albinism (OCA1–7) caused by mutations in seven different genes usually inherited as an autosomal recessive manner. HPS is a multisystem disorder characterized by tyrosinase-positive oculocutaneous albinism; a bleeding diathesis resulting from a platelet storage pool deficiency; and, in some cases, pulmonary fibrosis, granulomatous colitis, or immunodeficiency. *HPS3* encodes a protein that plays a role in organelle biogenesis associated with melanosomes, platelet dense granules, and lysosomes. Mutations in this gene are associated with HPS type 3 [16]. The girl had predominantly ocular manifestations of the disorder and is likely to develop the bleeding manifestations later in life. Initial workup was biased since the uncle had mutation for OCA type 1; however further evaluation using NGS technology clarified the genetic abnormalities in the child. The c.2635C > T variant

in heterozygous state in *MYO5A1* gene was also classified as VUS but unrelated to phenotype. Mutations in *MYO5A1* gene and additional genes like *RAB27A* are known to cause Griscelli syndrome [17]. This disorder results in pigmentary dilution of the skin and hair, the presence of large clumps of pigment in hair shafts identified on microscopy, and also abnormal accumulation of melanosomes in melanocytes. Patients manifest with hypotonia, marked motor development delay, and mental subnormality.

Family 2 A non-consanguineous couple presented with two previous infantile deaths; one died at 2 ½ year age who had jaundice, cholestasis, and coagulopathy. Galactosemia was suspected, but the galactose-1-phosphate uridyltransferase (GALT) enzyme assay report was not available. The second child died at 26 days age, with bluish skin lesions over the body-suspected bleeds. Keeping possibility of metabolic liver disease, galactosemia, or inherited disorder of cholestasis, a targeted inherited disorder panel was done by NGS. The variant prioritization and bioinformatics analysis revealed heterozygosity for a variant in the *ABCB11* gene in husband and wife, involved in progressive familial intrahepatic cholestasis (PFIC). The variant was in exon 25 of the gene c.3382C > G (p.Arg1128Gly). The parents were counseled about the likely pathogenic nature of the variant in homozygous state and 25% risk of recurrence. *ABCB11* gene encodes for the bile acid salt export pump (BSEP), which is located in the liver. Mutations in *ABCB11* cause familial liver disease PFIC2 [18], presenting with cholestasis, jaundice, and hepatomegaly, and liver failure leads to coagulopathy, bleeding, and early death. The parents opted for prenatal diagnosis in subsequent pregnancy. A chorionic villus sampling was performed at 12 weeks gestation and DNA analysis was done. The fetus was negative for the variant and hence likely to be unaffected.

Family 3 In a non-consanguineous family with three previous affected fetuses with encephalocele and postaxial polydactyly, NGS testing identified homozygosity for two pathogenic variants in the *CCD2A* gene. Meckel syndrome (MKS) is characterized by triad of *cystic renal dysplasia*, *occipital encephalocele*, or other CNS defects and polydactyly. However, all three components of the triad may not be present in every family. *CCD2A* mutations are associated with Meckel syndrome 6 [19]. A related disorder is Joubert syndrome. Meckel syndrome is a cause of neural tube defects (NTDs) and is common in Gujarati Indians [20]. The carrier frequency of MKS is high 1 in 30, in Bedouins of Kuwait. The risk of isolated NTDs in baby can be reduced to over 70% by periconceptional folic acid therapy, but folate administration may not be beneficial in autosomal recessive disorders leading to NTD. Meckel syndrome shows genotypic heterogeneity (Table 8.1). Some of the genes are involved in ciliogenesis or influence ciliary function. Some genes also cause other disorders like Joubert syndrome and nephronophthisis. Thus, the same gene can cause multiple phenotypes due to tissue pleiotropic effects.

Family 4 A 3-year-old child presented with developmental delay and facial abnormalities like prominent eyes, squint, and large ears. Keeping a possibility of chro-

Table 8.1 Genes involved in causation of Meckel syndrome, a cause of neural tube defects

Sl No.	Meckel type	Chromosome location	Gene MIM No.	Gene role
1	MKS 1	17q22	609,883	Encodes for a component of flagellar apparatus/ basal body proteome
2.	MKS 2	11q12.2	613,277	The gene TMEM216 is necessary for the assembly and functioning of cilia
3.	MKS 3	8q22.1	609,884	The gene affects centrosome and cilia number and renal tubulogenesis
4.	MKS4	12q21.32	610,142	It encodes for a protein responsible for ciliary assembly and trafficking in some tissues
5.	MKS5	16q12.2	610,937	Role in interactions in basal bodies and ciliary axonemes, affects ciliary protein functions
6.	MKS 6	4p15.32	612,013	Functions in transition zone at base of the cilia and is part of protein complex in basal body
7.	MKS 7	3q22.1	608,002	Encodes a component of ciliary proteins, needed for renal and cardiac development
8.	MKS 8	12q24.31	613,846	Encodes protein TCTN2 expressed in the brain, neural tube, heart, and kidneys
9.	MKS 9	17p11.2	614,144	Encodes for protein that associate with basal bodies and primary cilia
10.	MKS 10	19q13.2	614,951	Encodes for B9D2 which associates with basal bodies and primary cilia
11.	MKS 11	16q23.1	614,949	Encodes for component for protein complex which acts as barrier to restrict diffusion through plasma and ciliary membranes
12.	MKS 12	1q32.1	611,279	Encodes for protein of kinesin superfamily and has important role in intracellular transport and cell division

mosomal syndrome, a chromosomal microarray was advised. This revealed a hemizygous deletion of 8.7 Mb at 5q14.3 q15. The deleted region includes several genes including *MEF2C*, *LYSMD3*, and *GPR98*. The *MEF2C*related disorders are associated with severe intellectual disability, hypotonia, seizures, and facial abnormalities [21]. The diagnosis made was of chr.5q14.3q15 microdeletion syndrome. The parents were counseled regarding the report, the prognosis, and prenatal diagnosis options in next pregnancy. Chromosomal deletions or duplications can lead to intellectual disability or autism, and chromosomal microarray can be nowadays done as first-line test in these patients if it is available locally or from private laboratories. However genetic counseling may require expertise of a trained person.

8.8 Conclusion

DNA fingerprinting traditionally refers to the identification of individuals from blood and/or tissue samples for forensic purposes. But genetic fingerprinting can also include characterization of the genetic basis of human diseases, especially the

inherited disorders. Some of the variants or haplotypes identified may run in families and thereby also have pathological or phenotypic connotations. The DNA sequencing technologies have evolved over the years, and nowadays, high-throughput techniques and applications are available with increased automation. Thus, genetic fingerprinting can have various connotations in relation to human diseases. The genetic testing done would depend on the clinical situation or phenotype, and what we are looking for in a specific patient or individual. Pretest and posttest counseling are important to facilitate decision-making.

Acknowledgments Thanks to all families whose clinical details have been discussed as cases in this write up. Appropriate consent was taken for publication as per ethical guidelines from parents/guardian. Also, thanks to the individuals/labs for the molecular testing in the patients.

References

1. Roewer L (2013) DNA fingerprinting in forensics: past, present, and future. Investig Genet 4:22. https://doi.org/10.1186/2041-2223-4-22
2. Mehrab-Mohseni M, Tabatabaei-Malazy O, Hasani-Ranjbar S, Amiri P, Kouroshnia A, Bazzaz JT, Farahani-Shrhabi M, Larijani B, Amoli MM (2011) Endothelial nitric oxide synthase VNTR (intron 4 a/b) polymorphism association with type 2 diabetes and its chronic complications. Diabetes Res Clin Pract 91(3):348–352
3. Dittwald P, Gambin T, Szafranski P, Li J, Amato S, Divon MY et al (2013) NAHR-mediated copy number variants in a clinical population: mechanistic insights into both genomic disorders and mendelizing traits. Genome Res 23(9):1395–1409
4. Goidts V, Armengol L, Schempp W, Conroy J, Nowak N, Müller S, Cooper DN, Estivill X, Enard W, Szamalek JM et al (2006) Identification of large scale human specific copy number differences by inter species array comparative genomic hybridization. Hum Genet 119(1–2):185–198
5. Butler MG (2009) Genomic imprinting disorders in humans. J Assist Reprod Genet 26:477–486
6. Poole RL, Docherty LE, Al Sayegh A, Caliebe A, Turner C, Baple E, Wakeling E, Harrison L, Lehmann A, Temple IK, Mackay DJ, International Clinical Imprinting Consortium (2013) Targeted methylation testing of a patient cohort broadens the epigenetic and clinical description of imprinting disorders. Am J Med Genet A 161A(9):2174–2182
7. Mobasherizadeh S, Shojaei H, Havaei SA, Mostafavizadeh K, Davoodabadi F, Khorvash F, Ataei B, Daei-Naser A (2016) Application of the random amplified polymorphic DNA (RAPD) fingerprinting to analyze genetic variation in community associated-methicillin resistant staphylococcus aureus (CA-MRSA) Isolates in Iran. Glob J Health Sci 8(8):53822. https://doi.org/10.5539/gjhs.v8n8p185
8. Toval F, Guzmán-Marte A, Madriz V, Somogyi T, Rodríguez C, García F (2015) Predominance of carbapenem-resistant Pseudomonas aeruginosa isolates carrying blaIMP and blaVIM metallo-β-lactamases in a major hospital in Costa Rica. J Med Microbiol 64(Pt 1):37–43
9. Lixandru BE, Cotar AI, Straut M, Usein CR, Cristea D, Ciontea S et al (2015) Carbapenemase-producing Klebsiella pneumoniae in Romania: a Six-Month Survey. PLoS One 10(11):e0143214
10. Salimizand H, Menbari S, Ramazanzadeh R, Khonsha M, Saleh VM (2016) DNA fingerprinting and antimicrobial susceptibility pattern of clinical and environmental Acinetobacter baumannii isolates: a multicentre study. J Chemother 28(4):277–283
11. Sakthivel S, Zatkova A, Nemethova M, Surovy M, Kadasi L, Saravanan MP (2014) Mutation screening of HGD gene identified a novel alkaptonuria mutation with significant founder effect and high prevalence. Ann Hum Genet 78(3):155–164

12. Ankala A, Kohn JN, Dastur R, Gaitonde P, Khadilkar SV, Hegde MR (2013) Ancestral founder mutations in calpain-3 gene in the Indian Agarwal community: historical, clinical and molec ular perspective. Muscle Nerve 47(6):931–937
13. Bijarnia-Mahay S, Movva S, Gupta N, Sharma D, Puri RD, Kotecha U, Saxena R, Kabra M, Mohan N, Verma IC (2015) Molecular diagnosis of hereditary fructose intolerance: founder mutation in a community from India. JIMD Rep 19:85–93
14. Garewal G, Das R, Ahluwalia J, Marwaha RK, Varma S (2005) Nucleotide -88 (C-T) promoter mutation is a common beta-thalassemia mutation in Jat Sikhs of Punjab. India Am J Hematol 79(4):252–256
15. Panigrahi I, Kalra J, Goyad P, Khetarpal P, Munshi A (2016) Mutational analysis in Gaucher disease: implications in genetic counseling and management. J Genet Disor Genet Rep 5(2). https://doi.org/10.4172/2327-5790.1000132
16. Torres-Serrant M, Ramirez SI, Cadilla CL, Ramos-Valencia G, Santiago-Borrero PJ (2010) Newborn screening for Hermansky-Pudlak syndrome type 3 in Puerto Rico. J Pediatr Hematol Oncol 32:448–453
17. Panigrahi I, Suthar R, Rawat A, Behera B (2015) Seizure as the presenting manifestation in Griscelli syndrome type 2. Pediatr Neurol. 52(5):535–538
18. Van der Woerd WL, van Mil SW, Stapelbroek JM, Klomp LW, van der Graaf SF, Houwen RH (2010) Familial cholestasis: progressive familial intrahepatic cholestasis, benign recurrent intrahepatic cholestasis and intrahepatic cholestasis of pregnancy. Best Pract Res Clin Gastroenterol 24(5):541–553
19. Tallila J, Jakkula E, Peltonen L, Salonen R, Kestilä M (2008) Identification of CC2D2A as a Meckel syndrome gene adds an important piece to the ciliopathy puzzle. Am J Hum Genet 82:1361–1367
20. Young ID, Rickett AB, Clarke M (1985) High incidence of Meckel syndrome in Gujarati Indians. J Med Genet 22:301–304
21. Engels H, Wohlleber E, Zink A, Hoyer J, Ludwig KU, Brockschmidt FF et al (2009) A novel microdeletion syndrome involving 5q14.3-q15: clinical and molecular cytogenetic characterization of three patients. Eur J Hum Genet 17(12):1592–1599

Chapter 9
Molecular Diagnosis of Enteric Bacterial Pathogens

Amita Shrivastava, Pradeep K. Singhal, and Pankaj Shrivastava

Abstract Bacterial strains belonging to family *Enterobacteriaceae* are well-established enteric pathogens of humans and animals which are the major cause of mortality worldwide. So to deal with these infections, it is necessary to identify the disease-causing pathogens along with their virulence mechanism. The characterization and detection of these pathogens rely on conventional culturing and biochemical techniques. But these conventional techniques are low sensitive and time intense. This led to search for more rapid, sensitive, and advanced technique for their detection. In the present scenario, the molecular methods are most commonly applied to identify the pathogenic bacterial strains. The molecular methods include DNA-based methods, such as restriction endonuclease analysis of genomic and plasmid DNA, plasmid profiling, chromosomal DNA profiling using pulsed field gel electrophoresis, polymerase chain reaction (PCR)-based methods, 16S rRNA sequencing methods, polyphasic taxonomic approaches, etc. These are the more rapid and more sensitive techniques which overcome the issues of identification by conventional techniques. Hence, these techniques are considered the most promising for bacterial identification.

Keywords Conventional identification techniques · Enterobacters · Molecular techniques

A. Shrivastava · P. K. Singhal (✉)
Department of P. G. Studies and Research in Biological Sciences, Rani Durgavati University, Jabalpur, Madhya Pradesh, India

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

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_9

9.1 Introduction

Microorganisms are ubiquitous and extraordinarily diverse and accomplish various roles in the environment. They occurred broadly in the biosphere, foods, water, and soils and exogenously as well as endogenously in humans and animals. The microbial population of human intestinal tract comprises of a huge diverse of bacterial, viral, and parasitic forms including aerobic and anaerobic organisms. Majority of these organisms belonging to family *Enterobacteriaceae*, including *Escherichia coli*, *Shigella* sp., and *Salmonella enterica*, but some others such as *Pseudomonas* sp. also inhabit human gastrointestinal tract. Many of these organisms have beneficial effects on their hosts, yet certain conceivably harmful microorganisms can have significant negative consequences on both animals and human. It is accounted for that pathogenic infections cause around 40% of the roughly 50 million aggregate yearly passing around the world [1, 2].

Among enterobacters, widely dispersed species in nature is *Salmonella*. They are gram-negative, rod-shaped, facultative anaerobic, nonspore-forming motile organisms that can grow best at temperature extended from 7 to 48 °C and pH ranged from 6.5 to 7.5. They can dwell as common commensal in the gastrointestinal tracts of animals and humans which may be the reason of several infections such as diarrhea, bacteremia with enteric fever, or invasion of vascular structures, bone, or other localized sites [3]. They are additionally a reason for foodborne sickness around the world and a cause of livestock infections that can be transmitted from animals to humans. Another enterobacter, *Escherichia coli*, is a gram-negative, rod-shaped bacterium which typically resides in the lower intestinal tract of humans and animals. Its pathogenic strains may be the reason of a few diseases such as gastroenteritis, urinary tract infection, meningitis, peritonitis, and septicemia [4].

Enteric contamination-related illness can be caused by live enteric pathogens (such as bacterial dysentery), by toxins which they produced (such as staphylococcal food poisoning), or by combined both toxins and live agents (such as cholera). From these, a few pathogens act primarily through infections at the mucosal surfaces of the gastrointestinal tract (such as *Vibrio cholerae*), whereas others are able to cause systemic infections (such as *S. enterica* subspecies *enterica* serovar *Typhi* or *S. typhi*). Along these lines, the pathogenesis, etiology, and thusly resistance of these diseases are varied. These diseases are transmitted by contaminated food or water and usually lead to symptoms like diarrhea. Many sound peoples will encounter a self-constrained disease that endures a couple of day, yet in others, severe dehydration, bacteremia, chronic symptoms, and serious complications can develop, including malnutrition and death [5].

Identification of these infectious microbes is necessary to evaluate their association and sources of infection for treatment of diseases they cause. There are various types of methods that have been used for this purpose. The conventional methods including antibiotic resistance patterns, biochemical reactions, bacteriocin typing, and phage typing are normally less efficient, time-consuming, and costly. Serotyping may be considered as the reference method for strain characterization [6], but it may not discriminate the organisms originating from diverse regions [7]. These methods

for detection of enteric pathogens have several limitations and require technical expertise and subjective interpretation. A maximum number of pathogenic bacterial strains are identified by culture methods. Routine stool cultures are carry out to identify the existence of *Salmonella* sp., *Shigella* sp., *Campylobacter* sp., and *E. coli* O157, consuming at least 3–4 days. Due to this delayed diagnosis, patients are at risk for untreated infections, and the diseases can be transmitted to others. Further, *Campylobacter* species need specific conditions in the culture medium to regain viability, but non-culturable forms of *Shigella* failed to regain viability in the medium, adding to limitations of sensivity of culture techniques [8]. The closely related pathogenic and nonpathogenic strains of *E. coli* could also not be reasonably differentiated by culture techniques. Generally adopted techniques to detect and identify *E. coli* are culture, fermentation, enzyme-linked immunosorbent assay, and PCR. The conventional culture techniques, therefore, has many disadvantages such as biased, laborious, long and resource-intensive assay protocols, highlighting necessity of quick, relaible and sensitive techniques to identify and manage pathogenic enterobacters [9, 10].

Molecular techniques such as multilocus polymerase chain reaction (PCR), DNA sequencing, enzyme electrophoresis, biotyping, restriction endonucleases analysis, ribotyping, pulsed field gel electrophoresis (PFGE), nucleotide sequence analysis, protein analysis, and plasmid profiling have been successfully adopted in the recent past to supplement culture techniques. These are the most successful techniques that are rapid and sensitive. Among these molecular techniques, plasmid profile analysis [11, 12], random amplified polymorphic DNA analysis (RAPD) [11], repetitive extragenic palindromic sequences analysis by PCR (rep-PCR) [13], and pulsed field gel electrophoresis (PFGE) [10, 14, 15] are generally adopted and considered promising for bacterial identification.

We endeavor to acquire a multidisciplinary approach for the detection of virulent agents belonging to family *Enterobacteriaceae* and control infectious diseases. The main focus of this chapter is to compress the quick and sensitive identification techniques for pathogenic strains which concentrated specially on the identification of species-specific DNA sequences (molecular techniques).

9.2 PCR-Based Typing of Enteric Pathogens

It is one of the most commonly used molecular technique that can discriminate microbial strains in the analysis and epidemiological studis of pathogenic diseases. It is an accurate, reliable, and short time taking method to amplify a specific gene or DNA fragment within the genome of a species [16]. This is a powerful molecular strategy as it is able to amplify the smallest amount of particular microbial DNA sequences. PCR offers assortment of techniques for with numerous analytical and epidemiologic investigations such as PCR-RFLP, PCR-ribotyping, random amplification of polymorphic DNA (RAPD), repetitive extragenic palindromic (Rep)-PCR, enterobacterial repetitive intergenic consensus (ERIC)-PCR, PCR sequencing, and

other techniques [17–20]. In PCR, the targeted genes have been increased logarithmically in the number of copies which can be later detected on ethidium bromide-stained gels. On the basis of size (mass) and migration of the PCR product (amplicon), we can detect the expected target in the sample which later confirmed by the sequencing of the amplicon [21].

The advanced methods to detect enteric pathogens are based on specific nucleotide primers designed for PCR. This is a very fast and quick technique to detect and identify the enteropathogens; however, the conventional identification techniques are considered prolonged, too variable, and labor intensive and have several limitations. Previous studies also showed that PCR is quite high sensitive and specific strategy to identify enterobacters in stool sample as compared to that by the other conventional methods [22–24].

PCR techniques has a few constraints, for example, it is not capable to distinguish between viable and nonviable microorganisms, and the PCR inhibitors can lead to false-negative outcomes [25]. Other than the development of direct hybridization methods, there are several non-PCR-based techniques to amplify DNA like nucleic acid sequence-based amplification (NASBA) and strand displacement amplification [26, 27].

9.3 Gel Based Typing of Enteric Pathogens

The method of pulsed field gel electrophoresis (PFGE) had been introduced in 1984 [28] and has since been considered as a gold standard method to identify microorganisms and epidemiological investigations [29]. Among molecular techniques, it is the most reliable, precise, and reproducible method as well as presently being utilized as a part of international surveillance programs. It is a DNA-based subtyping method that creates DNA banding patterns. Initially, some rare restriction enzymes cut the DNA into short segments, generating 8–25 large bands. Later, the DNA segments are subjected to electrophoresis on agarose gel for their separation on the basis of their size, and to diminish sheering of large DNA segments, the current have been applied at alternating angles. The selection of restriction enzymes ensures the generation of adequate size DNA segment for electrophoretic separation (1000–15,000 base pairs) along with suitable number of bands. Comparative study of multiple enzyme patterns explains the existence of new species, thus enhanced the power of discrimination of this method [30].

PFGE has been used successfully for characterizing *E. coli* [31, 32], *Salmonella*, *Listeria*, and other pathogens. In PFGE, the restriction pattern concerning the entire bacterial genome is studied barring the use over probes. However, point mutations, deletions, insertions, and loss or acquisition of plasmids may elucidate minute variations in profiles inside a subtype or between epidemiologically homologous strains. These progressions for the most part result in a few fragment contrasts in PFGE banding patterns. PFGE is a very successful method to identify enteric pathogens besides the fact that it has some disadvantages too. It is distinctly expensive and

needs more than 3 days getting the outcome. The level of separation likewise relies upon the selection of restriction enzymes. A further drawback of this technique is poor portability. Specialized issues with PFGE change in their unpredictability and simplicity of analysis. They can incorporate the feeble power of banding designs because of low cell concentration, art factual bands because of inadequate processing of DNA, skewed lanes because of faulty electrodes or uneven gel thickness, and additionally buffer height because of uneven surfaces utilized for gel casting or electrophoresis. Some serotypes, especially those with certain distinct phage types, can be so genetically homogeneous that multiple genotypic techniques fail to discriminate outbreak from non-outbreak strains [33].

9.4 Ribotyping

For the bacterial identification, the most commonly used technique is PCR, and it can be utilized successfully by focusing on signature sequences of the target DNA. In any case, it is typically hard to discover particular primers that have a solitary target and do not yield any non-specific amplification. This technical issue can be overwhelmed by another technique termed as PCR-ribotyping which is generally an adopted technique to characterized bacterial strains. The diverse bacterial species comprises highly conserved ribosomal gene in respect to other bacterial strains. This feature is very beneficial to describe the genetic relationship of diverse bacteria species. The rRNA coding sequence fingerprinting defines about the hybridization of restriction-digested DNA fragments along with probes specific for rDNA [34]. In this method, a multiple band pattern is generated which is extremely prejudicial and can be utilized for affirmation of microorganisms up to species level as well as in many cases even beyond.

rRNA gene sequencing has been an exceptionally authentic technique to characterize bacterial species, yet it has insufficient heterogeneity for more advance classification. PCR-ribotyping relies upon the spacer regions or IV (intervening sequence) amplification between 16S and 23S RNA genes. The inconstancy in length and copy number may facilitate classification of several bacterial strains and mycoplasma [35, 36]. For instance, in a previous report, PCR-ribotyping had identified seven serovars of *Salmonella* sp. [37]. This method has replaced the conventional PCR for affirmation of divergent species as it subsequently convoluted by non-specific amplifications and hence required supporting morphological and biochemical information.

Ribotyping is actually capable to subtype a few microbial isolates that fall inside some basic serotypes and phage types [38]. Lin et al. [39] distinguished seven diverse ribotypes among *S. enteritidis* isolates when chromosomal DNA was treated with SphI. Fernandes et al. [40] demonstrated that ribotyping is more appropriate for following the diversity of *S. enteritidis* and the restriction endonuclease SphI distinguished finest within subtypes of this serotype. Dambaugh et al. [41] obtained discrete patterns for the most widely recognized serovars of Salmonella species by using PvuII (restriction enzyme).

Automated ribotyping has been shown as being useful for the identification of various bacterial spp. [42–44] and successfully applied for the identification of enterobacters. As compared to conventional PCR, this technique is more reliable as it requires a single primer as opposed to setting up reactions for every species. Thus, odds of false-negative outcomes because of non-specific amplification are decreased. Comparison study of ribotyping with PFGE has been to some degree flighty and frequently relies upon the enzymes utilized for digestion and also the idea of the population nature.Various researches have found that PFGE has more discrimination power as compare to that by ribotype analysis [45] whereas others have found these two techniques equivalent [46] or ribotype analysis superior [47]. Moreover, this technique is fast, more reliable, and highly reproducible and can be successfully applied for bacterial characterization. However, it has some disadvantages too. This method has more influence in epidemiological applications yet has low influence in the quick identification of pathogenic microbes [48]; however, ribotyping is considered as not suitable for local epidemiological applications or surveillance studies in a restricted region [49]. Further, important disadvantages of automated ribotyping are the need of expensive reagents per isolate and the high cost of the automated RiboPrinter itself.

9.5 DNA Sequence-Based Typing Methods

The DNA sequencing provides data regarding the degree of genetic diversity and population structure of the bacterium of interest. This strategy is considered as reproducible and easy transposable between laboratories [50, 51]. The following are the commonly used DNA sequence-based typing methods.

9.5.1 Variable Number of Tandem Repeat (VNTR) Analysis

Genomes of several bacterial species have regions with repetitive DNA sequence motifs extending from a couple of bases to more than 100 base pairs in length. The repetitions of sequences termed as tandem which means a number of copies of each of the repeat motifs are clustered together and arranged in the similar direction. These repeats in a tandem can be highly variable, even between strains of the same species. It is referred as PCR-based method that depends upon the DNA amplification that includes short tandem repeats (STR) of a DNA sequence. PCR primers are intended to anneal non-repetitive sequences simply outside the repeat region, and amplified products are isolated and estimated to detect the number of repeats present in the amplicon.

In this technique, variations in the number of repeated copies at particular loci were used to differentiate the microbial isolates. As a result of moderately high mutation rate, strains can consolidate specific patterns in a comparatively small

duration of time [52]. Several regions of repeated motif may be investigated in the meantime to increase expanded differentiation to analyze genetic diversity. The most widely recognized strategy utilizing multiple VNTR loci for typing is termed as multiple locus VNTR analysis (MLVA) or MLV fingerprinting (MLVF) [53].

9.5.2 Multilocus Sequence Typing (MLST)

The MLST technique, a molecular typing approach, introduced since 15 years is for the most part performed on seven housekeeping genes. It analyzes DNA sequences from regions of housekeeping or virulence genes or potentially rRNA sequences which changes because of mutation or recombination events [54]. In a particular gene, the nucleotide variations are merged and used in the determination of discrimination between the strains [55]. MLST gives information like those got by multilocus enzyme electrophoresis, yet in substantively more noteworthy detail, since it can survey particular nucleotide changes as opposed to screen for changes in the overall charge and expression of the enzyme under study [54]. In MLST, a sequence type (ST) can be defined as sequencing the allelic group including the isolate for each of the genes. Relatedness between STs can be disclosed through various techniques of clustering: (a) the unweighted pair group method with arithmetic mean, using distance matrices containing the pairwise differences of allelic profiles; (b) the minimum spanning tree approach, constructing a tree that interfaces all sections such that the summed distance of all connections of the tree is the shortest, i.e., minimum; and (c) the base upon related sequence types (eBURST or the more recent global optimized version, goeBURST), algorithm, inferring patterns of evolutionary descent among isolates by a model of clonal expansion and diversification and assigning isolates to clonal complexes [56].

This technique is to a great degree valuable for protracted epidemiological or phylogenetic analyses. More than 230 *Salmonella* isolates were lately described by MLST in light of sequences from the 16S RNA, *pduF*, *glnA*, and *manB* genes [57]. These outcomes were contrasted with PFGE and serotype analysis. MLST could discriminate strains superior to others like PFGE; however, not all genes performed similarly. For MLST to be compelling as an epidemiological tool, the criteria for the selection of genes and their number should be satisfactory for the discriminations of the isolates with later genetic diversity. In this case, genes under greater selective pressure, such as virulence genes, might be giving a superior outcome. For this situation the technique is commonly termed as multivirulence-loci sequence typing (MVLST) [58].

MLST demonstrates incredible guarantee for exact strain differentiation with information that can be precisely shared between research labs. Be that as it may, general interest of this approach will be enhanced when automated sequence apparatus turns out to be more moderate and labs can generate familiar environment with complicated DNA sequencing and statistical software sequencing. This technique is very tedious and exorbitant yet can be exceptionally oppressive if the genes are

accurately chosen. It has been effectively utilized worldwide in the study of disease transmission and population genetic studies of many gram-positive and gram-negative bacteria. It is less appropriate for routine typing in outbreak investigations or local surveillance studies on account of its moderately low discriminative power, high cost, and workload.

9.5.3 Sequencing of 16S rRNA Gene

The 16S rRNA gene sequencing, a universal method, has been widely proposed to identify microorganisms and detect new species [59]. In all living cells, ribosomes are present which are involved in protein synthesis mechanism. They consist of two subunits, and each subunit is composed of protein and ribosomal RNA (rRNA). In prokaryotes 70S ribosome is present which consists of 30S subunit (small subunit) and 50S subunit (large subunit). The 30S subunit is made out of 16S rRNA and 21 polynucleotide chains, while 50S subunit is made out of 5S and 23S rRNAs. The 16S rRNA contains 16S rRNA gene which gives a species-specific signature sequences, and this signature sequences are very important in bacterial identification. The signature sequences are unique DNA sequences (five to ten bases long) to numerous major groups of prokaryotes, archaea, and eukarya which are found specifically in the 16S rRNA location in all groups of organisms. The average lengths of the structural rRNA genes are 1522 bp, 2971 bp, and 120 bp, respectively, for 16S, 23S, and 5S rRNAs.

There are numerous focal points of utilizing ribosomal RNA in identification methods as they are present in all cells; RNA genes comprise of highly conserved nucleotide sequences, mixed with variable regions that are genus or species specific, and there is no need to culture the microbial cells [60]. All bacterial species contains 16S rRNA gene [61], so this technique can be used as gold standard for bacterial identification up to species level. This technique may ensure reclassification of bacterial species into into new genera [62, 63]. This method might be likewise valuable in studying the relationship of unidentified species to known ones and in the discovery of new species.

The members of *Enterobacteriaceae* family could be identified up to 80.1–94.4% through the Vitek GNI+ and API 20E systems. Some members of this family like *E. coli* are difficult to distinguish; in such condition 16S rRNA sequencing should be employed which is reported as a significantly better technique compared to that by others [64]. Subsequently, bacterial strains can be characterized accurately by nucleotide sequencing of the PCR product and comparing it with known sequences put away in a database [65]. It can likewise be utilized to distinguish microorganisms that are hard to culture or be applied to samples from post-antibiotic treatment [66–68].

9.5.4 Whole Genome Sequencing

The sequencing of the entire genome of a microbe and comparing the sequenced genome with known references are referred as whole genome sequencing (WGS). This is a key approach in molecular methods for genome mapping of novel species, finishing genomes of known organisms, to compare genomes over diverse samples, to generate more precise reference genomes, and to identify microbes and other genomic analyses. Presently, this technique is considered as reference microbial typing method to enhance knowledge and understanding of infectious diseases and clinical microbiology [69].

The members of family *Enterobacteriaceae* are well-known human and animal pathogens including *Escherichia coli*, *Salmonella enterica*, *Yersinia*, *Shigella*, etc. They cause various diseases in humans specially diarrhea which is the second biggest killer of children globally and most common problem in human beings. WGS of multiple strains of these bacterial pathogens has contributed immensely to our understanding of the high level of similarities and differences among closely related organisms at sequence level. By WGS of closely related species, we can distinguish genes, characterize gene structure, highlight fast as well as slow evolutionary change, recognize regulatory elements, and disclose combinatorial limit about gene regulation. Hence, it can be concluded that the WGS approach offers the potential to understand and differentiate the basic pathogenic mechanisms employed by different pathogenic bacteria.

9.6 Polyphasic Taxonomic Approach

In modern technology a new technique has been employed for classifying microorganisms taxonomically and phylogenetically to distinguish bacterial species which is known as polyphasic taxonomy approach. This technique is first proposed by Rita Colwell [70]. The phenotypic, genotypic and chemotypic, technologies to elucidate the taxonomic position of microbes together are termed as polyphasic taxonomy approach. Recently, this is the most prominent technique for identification and classification of microorganisms. The microorganisms as were partitioned beneath invalid taxa have recently been partitioned under new genera and species kind through this strategy. This approach is coupled with various phenotypic, genotypic, and chemotypic characters and is more advanced than other techniques. The main advantage of this approach is that it reduces the burden of misidentification as well as incomplete ID and offers authenticate identification of microbes.

It was observed that the phenotypic characters often among genetically different species and molecular methods alone are not capable to distinguish phylogenetically related species. So this technique is more reliable to discriminate microbes accurately.

9.7 Pros and Cons of Molecular Typing Methods

Previously used cultural and biochemical methods for microbial typing are too variable, labor intensive, and time-consuming and provided unreliable results, but the recently used molecular methods for detecting and typing of bacterial pathogens are fast, labor reducing, more touchy, particular, and proficient which provides reliable data for tracing the bacterial pathogens.

In spite of the fact that the utility of molecular approaches will emphatically influence the diagnosis of diseases caused by enterobacters, there are a few impediments. One burden of any non-culture strategy is the unavailability of the microbial isolates for antibiotic susceptibility test or any other tests by public health laboratories. This will frustrate the public health laboratories' capacity to give data concerned subtyping to the studies of epidemiology, thusly restricting their capacity for determining and characterizing outbreaks. A further impediment of this single-stool work process includes the identification of certain ova or parasites that are not accessible on molecular testing menus. Even though these parasites are uncommon, they will show the necessity of microscopy for their identification. One final aspect in utilizing molecular procedures for enteropathogens is the noteworthiness of the discoveries when the approach can't discriminate viable and nonviable pathogens [71]. Obvious indicators of viability have been the isolation of enteropathogens from stool sample. However, the importance of the nucleic acid of enteropathogens presented in stool samples and their concern with infections are not understood so far. Additionally the researches are expected to address the relationship of these discoveries. Molecular approaches can likewise recognize other bacterial species that are not identified with conventional culture techniques. The expanded affectability of this approach leaves open inquiries with regard to the hugeness of the discoveries. Present information about intense gastroenteritis including analysis and the study of disease transmission depends on traditional cultures and microscopic examination. The pattern might change as we push far from culture toward fast nonculture techniques.

9.8 Conclusions

Numerous approaches are accessible to distinguish bacterial strains; however, each of them has its focal points and impediments that make it valuable in a few examinations and prohibitive in others. A perfect strategy ought to satisfy the accompanying six criteria are the following: type ability, reproducibility, discriminatory power, simplicity of elucidation, simple to utilize, and minimal effort. It is clear that any technique utilized as of now for the identification of enteric bacterial strains is a perfect strategy alone as far as these criteria, yet all strategies display benefits and furthermore confinements. Clearly it is hard to locate a solitary approach, which is most appropriate to identify and characterize microbial strains.

Amid most recent couple of decades, the quality of molecular science has altered biological sciences. The molecular approaches have opened another part to characterize diverse bacterial strains. These approaches can distinguish to a superior degree than phenotypic techniques and enhance our insight into genetic and epidemiological relationships. Conventional research center testing techniques depending on microbial morphology and growth factors are progressing to a work process structure where molecular testing is fit for distinguishing pathogenic strains. The molecular testing industry is promoting extended menus and instruments that empower research facilities of any size to solidify testing on a solitary stage. The capacity to automate sample preparation, extraction, and amplification on a solitary framework spares time and enhances lab effectiveness while enhancing institutionalization of test outcomes. The decision of a molecular typing method will rely on the expertise level and assets of the research center/laboratory and the point and extent of the investigation.

A change from the conventional research center testing approaches to molecular approaches will substantially affect the lab too. The workload in conventional methods is relentless and tedious. Molecular techniques will decrease workload and material prerequisites as well as dispose of the subjective elucidation concerned with culture methods. Albeit molecular approaches are expensive, their capacity to identify various pathogenic strains with one test board is more savvy than carrying out numerous tests utilizing conventional techniques.

Reference

1. Kramer MN, Coto D, Weidner JD (2005) The science of recalls. Meat Sci 71:158–163
2. Invitski D, Abdel-Hamid I, Atanasov P, Wilkins E (1999) Biosensors for the detection of pathogenic bacteria. Biosens Bioelectron 14:599–624
3. Hook EW (1990) Salmonella species (including typhoid fever). In: Mandell GL, Douglas RG, Bennett JE (eds) Principles and practices of infectious diseases. Churchill Livingstone, New York, pp 1700–1715
4. Vidic J, Manzano M, Chang CM, Jaffrezic-Renault N (2018) Advanced biosensors for detection of pathogens related to livestock and poultry. Vet Res 48:11
5. Dougan G, Huett A, Clare S (2002) Vaccines against human enteric bacterial pathogens. Br Med Bull 62:113–123
6. Koneman EW, Allen SD, Janda WM, Schreckenberger PC, Winn WC (1997) Diagnostic microbiology, 5th edn. Lipponcott Raven Publishers Philadelphia, New York, pp 50–56
7. Wong HC, Ho CY, Kuo LP, Wang TK, Lee CL, Shih YC (1999) Ribotyping of the *Vibrio parahaemolyticus* isolates obtained from food poisoning outbreaks in Taiwan. Microbiol Immunol 43:631–636
8. de Boer RF, Ott A, Kesztyus B, Kooistra-Smid A (2010) Improved detection of five major gastrointestinal pathogens by use of a molecular screening approach. J Clin Microbiol 48:4140–4146
9. Hobson NS, Tothill I, Turner AP (1996) Microbial detection. Biosens Bioelectron 11:455–477
10. Fournier-Wirth C, Deschaseaux M, Defer C, Godreuil S, Carrière C, Bertrand X, Tunez V, Schneider T, Coste J, Morel P (2006) Evaluation of the enhanced bacterial detection system for screening of contaminated platelets. Transfusion 46:220–224

11. Seno M, Sakaki M, Ogawa H (2004) Genotypic diversity of *Salmonella enteritidis* isolates from sporadic patients in limited area during one year. J Infect 49:291–296
12. Tekeli A, Erdem B, Sahin F, Koyuncu E, Karasartova D et al (2006) Plasmid profiles and randomly amplified polymorphic DNA analysis of *Salmonella enterica* serotype Enteritidis strains from outbreaks and sporadic cases in Turkey. New Microbiol 29:251–260
13. Kilic A, Bedir O, Kocak N, Levent B, Eyigun CP et al (2010) Analysis of an outbreak of Salmonella enteritidis by repetitive-sequence-based PCR and pulsed-field gel electrophoresis. Intern Med 49:36–46
14. Us E, Erdem B, Tekeli A, Gerceker D, Saran B et al (2011) Investigation of *Salmonella* serotype enteritidis isolates by plasmid profile analysis and pulsed-field gel electrophoresis. Microbiyol Bul 45:210–222
15. Merino LA, Ronconi MC, Navia MM, Ruiz J, Sierra JM et al (2003) Analysis of the clonal relationship among clinical isolates of *Salmonella enterica* serovar Infantis by different typing methods. Rev Inst Med Trop Sao Paulo 45:119–123
16. Almeida PF, Almeida RCC (2000) A PCR protocol using inl gene as a target for specific detection of *Listeria monocytogenes*. Food Control 11:97–101
17. Singh A, Goering RV, Simjee S, Foley SL, Zervos MJ (2006) Application of molecular techniques to the study of hospital infection. Clin Microbiol Rev 19:512–530
18. Towner KJ, Cockkayne E (1995) Molecular methods for microbial identification typing, 1st edn. Chapman & Hall, Ltd, London, pp 3–10
19. Ranjbar R, Hosseini MJ, Kaffashian AR, Farshad S (2010) An outbreak of shigellosis due to serotype 3a in a prison in Iran. Arch Iran Med 11:413–416
20. Vaneechoutte M, Verschraegen G, Struelens MJ (1997) Typing of *Staphylococcus aureus* and *Staphylococcus epidermidis* strains by PCR analysis of inter-IS256 spacer length polymorphisms. J Clin Microbiol 35:2580–2587
21. Khan AA, Nawaz MS, Khan SA, Cerniglia CE (2000) Detection of multidrug-resistant *Salmonella typhimurium* DT104 by multiplex polymerase chain reaction. FEMS Microbiol Lett 82:355–360
22. Islam M, Hossain M, Hasan M, Rahman M, Fuchs G, Mahalanabis D et al (1998) Detection of *Shigella* from stools of dysentery patients by culture and polymerase chain reaction techniques. J Diarrhoeal Dis Res 16:248–251
23. Kulkarni SP, Lever S, Logan JM, Lawson AJ, Stanley J, Shafi MS (2002) Detection of *Campylobacter* species: a comparison of culture and polymerase chain reaction based methods. J Clin Pathol 55:749–753
24. Lampel K, Orlandi P, Kornegay L (2000) Improved template preparation for PCR-based assays for detection of food-borne bacterial pathogens. Appl Environ Microbiol 66:4539–4542
25. Radstrom P, WA AL-S (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR inhibiting samples. Appl Environ Microbiol 64:3748–3753
26. Schweitzer B, Kingsmore S (2001) Combining nucleic acid amplification and detection. Curr Opin Biotechnol 12:21–27
27. Christine CG (2004) Life beyond PCR: alternative target amplification technologies for the diagnosis of infectious diseases, part II. Clin Microbiol Newsl 26:129–136
28. Schwartz DC, Cantor CR (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37:67–75
29. Finney M (2000) Pulsed field gel electrophoresis. Curr Protoc Mol Biol 51:2.5B.1–2.5B.9
30. Liebisch B, Schwarz S (1996) Molecular typing of *Salmonella enterica* sub sp. *enterica* serovar *Enteritidis* isolates. J Med Microbiol 44:52–59
31. Mitsuda T, Muto T, Yamada M, Kobayashi N, Toba M, Aihara Y, Ito A, Yokota S (1998) Epidemiological study of a food-borne outbreak of enterotoxigenic *Escherichia coli* O25: NM by pulsed-field gel electrophoresis and randomly amplified polymorphic DNA analysis. J Clin Microbiol 36:652–656

32. Philippe B, Lalande V, Salauze B, Burghoffer B, Avesani V, Delmée M, Rossier A, Barbut F, Petit JC (2000) Comparison of PCR-Ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. J Clin Microbiol 38:2484–2487
33. Ahmed R, Soule G, Demczuk WH, Clark C, Khakhria R, Ratnam S, Marshall S, Ng LK, Woodward DL, Johnson WM, Rodgers FG (2000) Epidemiologic typing of *Salmonella enterica* serotype enteritidis in a Canada-wide outbreak of gastroenteritis due to contaminated cheese. J Clin Microbiol 38:2403–2406
34. Mendoza MC, Landeras E (1999) In: Saeed AM, Gast RE, Potter ME, Wall PG (eds) Molecular epidemiological methods for differentiation of *Salmonella enterica* serovar *Enteritidis* strains in humans and animals: epidemiology, pathogenesis, and control. Iowa University Press, Ames, pp 125–140
35. Bidet P, Barut F, Lalande V, Burghoffer B, Piet JC (1999) Development of a new PCR-ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. FEMS Microbiol Lett 175:261–266
36. Sechi LA, Leori G, Lollai SA, Dupre I, Muloicotti P, Fadda G (1999) Different strategies for the molecular characterization of the *Mycobacterium bovis* strain isolated in Sardinia Italy. Appl Environ Microbiol 65:1781–1785
37. Lagatolla C, Dolzani L, Tomin E, Lavenia A, Michele M, Tommasini T (1996) PCR-Ribotyping for characterization *Salmonella* isolates of different serotypes. J Clin Microbiol 34:2440–2443
38. Landeras E, onzalez-Hevia MA, Alzugaray R, Mendoza MC (1996) Epidemiological differentiation of pathogenic strains of *Salmonella enteritidis* by ribotyping. J Clin Microbiol 34:2294–2296
39. Lin AW, Usera MA, Barrett TJ, Goldsby RA (1996) Application of random amplified polymorphic DNA analysis to differentiate strains of *Salmonella enteritidis*. J Clin Microbiol 34:870–876
40. Fernandes SA, Ghilardi ACR, Tavechio AT, Machado AMO, Pignatari ACC (2003) Phenotypic and molecular characterization of *Salmonella enteritidis* strains isolated in São Paulo, Brazil. Rev Inst Med Trop 45:59–63
41. Dambaugh TR, Mangiaterra E, Fritschel S (1997) Ribotype characterization of *Salmonella* and *E. coli* O157:H7 with PvuII on the RiboPrinter microbial characterization system. Am Soc Microbiol (Washington, DC) 439:14
42. Busse HJ, Kainz A, Tsitko IV, Salkinoja-Salonen M (2000) Riboprints as a tool for rapid preliminary identification of sphingomonads. Syst Appl Microbiol 23:115–123
43. Satokari R, Mattila-Sandholm T, Suihko ML (2000) Identification of pediococci by ribotyping. J Appl Microbiol 88:260–265
44. Clermont D, Harmant C, Bizet C (2001) Identification of strains of *Alcaligenes* and *Agrobacterium* by a polyphasic approach. J Clin Microbiol 39:3104–3109
45. Fontana J, Stout A, Bolstorff B, Timperi R (2003) Automated Ribotyping and pulsed-field gel electrophoresis for rapid identification of multidrug-resistant *Salmonella* serotype Newport. Emerg Infect Dis 9:496–499
46. Navarro F, Llovet T, Echeita MA, Coll P, Aladuena A, Usera MA, Prats G (1996) Molecular typing of *Salmonella enterica* serovar *typhi*. J Clin Microbiol 34:2831–2834
47. Liebana E, Garcia-Migura L, Breslin MF, Davies RH, Woodward MJ (2001) Diversity of strains of *Salmonella* enterica serotype *enteritidis* from English poultry farms assessed by multiple genetic fingerprinting. J Clin Microbiol 39:154–161
48. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV (2001) PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg Infect Dis 7:382–389
49. Riley LW (2004) Molecular epidemiology of infectious diseases: principles and practices. ASM Press, Washington, p 337
50. Maiden M, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998a) Multilocus sequence typing: a

portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140–3145
51. Maiden MCJ (2006) Multilocus sequence typing of bacteria. Annu Rev Microbiol 60:561–588
52. Call DR, Orfe L, Davis MA, Lafrentz S, Kang MS (2008) Impact of compounding error on strategies for subtyping pathogenic bacteria. Foodborne Pathog Dis 5:505–516
53. Sabat AJ, Chlebowicz MA, Grundmann H, Arends JP, Kampinga G, MEESSEN NE, Friedrich AW, Van Dijl JM (2012) Microfluidic-chip based multiple-locus variable-number tandem repeat fingerprinting with new primer sets for methicillin-resistant *Staphylococcus aureus*. J Clin Microbiol 50:2255–2262
54. Maiden MCJ, Ygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG (1998b) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140–3145
55. Yan SS, Pendrak ML, Abela-Ridder B, Punderson JW, Fedorko DP, Foley SL (2003) An overview of *Salmonella* typing public health concern. Clin Appl Immunol Rev 4:189–204
56. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA (2012) PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13:87
57. Kotetishvili M, Stine OC, Kreger A, Morris JG, Sulakvelidze A (2002) Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. J Clin Microbiol 40:1626–1635
58. Chen Y, Zhang W, Knabel SJ (2007) Multi-virulence locus sequence typing identifies single nucleotide polymorphisms which differentiate epidemic clones and outbreak strains of *Listeria monocytogenes*. J Clin Microbiol 45:835–846
59. Wang RF, Cao WW, Cerniglia CE (1994) A 16S rDNA-based PCR method for rapid and specific detection of *Clostridium perfringens* in food. Mol Cell Probes 8:131–137
60. Fox GE, Magrum LJ, Balch WE, Wolfe RS, Woese CR (1977) Classification of methanogenic bacteria by 16S ribosomal RNA characterization. Proc Natl Acad Sci U S A 74:4537–4541
61. Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271
62. Olsen GJ, Overbeek NL, Larsen N et al (1992) The ribosomal database project. Nucleic Acids Res 20(Suppl):2199–2200
63. Olsen GJ, Woese CR (1993) Ribosomal RNA: a key to phylogeny. FASEB J 7:113–123
64. Tang YW, Ellis NM, Hopkins MK et al (1998) Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli. J Clin Microbiol 36:3674–3679
65. Clarridge JE III (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 17:840–862
66. Relman DA, Falkow S (1992) Identification of uncultured microorganisms: expanding the spectrum of characterized microbial pathogens. Infect Agents Dis 1:245–253
67. Brouqui P, Raoult D (2001) Endocarditis due to rare and fastidious bacteria. Clin Microbiol Rev 14:177–207
68. Harris KA, Fidler KJ, Hartley JC, Vogt J, Klein NJ, Monsell F, Novelli VM (2002) Unique case of *Helicobacter* sp. osteomyelitis in an immunocompetent child diagnosed by broad-range 16S PCR. J Clin Microbiol 40:3100–3103
69. Ekblom R, Jochen BW (2014) A field guide to whole-genome sequencing, assembly and annotation. Evol Appl 7:1026–1042
70. Colwell RR (1970) Polyphasic taxonomy of the genus *Vibrio*: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus* and related *Vibrio* species. J Bacteriol 104:410–433
71. Platts-Mills JA, Liu J, Houpt ER (2003) New concepts in diagnostics for infectious diarrhea. Mucosal Immunol 5:876–885

Chapter 10
Application of DNA Fingerprinting: DNA and Human Trafficking

Maria Jesus Alvarez-Cubero, Maria Saiz, Luis Javier Martinez-Gonzalez, Juan Carlos Alvarez, and Jose Antonio Lorente

Abstract It has been estimated that every year an average 1.2 million children are victims of trafficking. This data is not only overwhelming but a disgrace. That is why, the efforts to fight against this crime have to be constant, not only when natural disasters affect countries where human trafficking is chronic problem. Countering trafficking should be an ongoing legal, but more so humanitarian, effort. The use and helpfulness of DNA typing for identity establishment is well known. Till today, DNA-PROKIDS, which operates with different levels of activity in 16 different countries, has generated relevant and promising data. Since April 2010, more than 11.200 samples have been processed, leading to more than 900 positive identifications. Two hundred fifty-seven illegal adoptions were detected with these investigations. In the nonexistence of this initiative, the children would have been in a precarious situation either would have been sold into illegal adoption, under exploitation, or would have died without proper identification. Families of the respective child would still be in distress for the loss of their kids.

Keywords Human trafficking · DNA fingerprinting · DNA-PROKIDS · Identification

---

Author contributed equally with all other contributors.María Jesús Álvarez Cubero and María Saiz Guinaldo

M. J. Alvarez-Cubero · L. J. Martinez-Gonzalez
GENYO, Centre for Genomics and Oncological Research: Pfizer/University of Granada/ Andalusian Regional Government – PTS, Granada, Spain

M. Saiz · J. C. Alvarez
DNA-PROKIDS Program, Department of Legal Medicine, Faculty of Medicine – PTS, University of Granada, Granada, Spain

J. A. Lorente (✉)
GENYO, Centre for Genomics and Oncological Research: Pfizer/University of Granada/ Andalusian Regional Government – PTS, Granada, Spain

DNA-PROKIDS Program, Department of Legal Medicine, Faculty of Medicine – PTS, University of Granada, Granada, Spain
e-mail: jose.lorente@genyo.es

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_10

10.1 Human Trafficking

One of the main social problems that has been increasing recently is human trafficking. Far from sharing culture, customs, and knowledge, globalization also promotes trades where human beings and sexual or exploitative services are sold, which are generally *bought* by the most developed countries from countries on process of development. Therefore, this problem affects the whole world and multiple projects act to minimize it [1].

According to United Nations Office on Drugs and Crime (UNODC), Article 3, paragraph (a) of the Protocol to Prevent, Suppress and Punish Trafficking in Persons, "trafficking in Persons is the recruitment, transportation, transfer, harboring 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" [2].

10.2 DNA Technology Applications

New DNA technologies can play a very important role in avoiding the worldwide human trafficking. They are used routinely in immigration to ensure the relationship through genetic comparison of those involved, determining their similarity. To confirm or deny assertions of biological kinship between two or more individuals, DNA reports are essencial in important cases such as whether to advance or stop requests for immigration and international adoptions.

In addition, biological tests can give information on missing individuals and mysterious human remains. On these two aspects, genetic technologies can be employed for identification of missing persons in cases of human trafficking [3].

10.2.1 What is DNA-PROKIDS?

Many factors are responsible to minimize strong prosecution in human trafficking cases throughout the globe such as social, economic, cultural, and political factors. In this regard, DNA-PROKIDS comes into play as an internationally collaborated organization which incorporates forensic DNA typing to fight against child trafficking globally. Dr. Jose Antonio Lorente, in the capacity of Director, Genetic Identification Laboratory of University of Granada (Department of Legal Medicine), University of Granada (UGR), created the initiative DNA-PROKIDS in 2004. After 2009, the organization got support from Prof. Dr. Arthur J. Eisenberg, Institute of Applied Genetics, Department of Forensic and Investigative Genetics, University of North Texas Health Science Center (UNT-HSC), USA, to spread its wings.

The main objective is to combat trafficking in persons through the identification of victims and their blood relatives through DNA typing mostly the children: an unidentified child is the ideal victim for human traffickers and exploiters, as far as to fight and prevent illegal adoptions. Furthermore, another purpose was to implement the initiative not only in the countries where the pilot study (Mexico and Guatemala) was conducted as a national control but also in all corners of the world to combat international human trafficking [4, 5].

DNA-PROKIDS has three basic purposes for the identification of victims and their returning to their respective families, prevention of human trafficking through victim identification and collection of information regarding their origins, as well as the police and judicial system protocols [5].

Therefore, DNA-PROKIDS aims to create a repository worldwide with the available genetic knowledge promoting international and systematic collaborations. This database includes the following points:

1. Reintegrating missing kids to their relatives employing DNA technologies
2. Preventing illegal adoptions by using DNA technology to compare the biological link between the adopted kids and the relative who put them up for adoption
3. Applications of inputs for improvements in police and judicial systems to deal with kids and women trafficking cases efficiently
4. Analyzing and suggesting the implementation of common legislative frameworks to fight child trafficking
5. Assessing and solving problems associated with societal and language barrier to overcome child trafficking
6. Providing collaborated coordination and training of DNA experts globally regarding DNA-based identification from different sources

DNA-PROKIDS promotes a common action, involving both governments and institutions, and seeks the establishment of DNA registries around the world. Goals and objectives are being increasingly unified, but there is still a long way to go.

10.2.2 DNA-PROKIDS Operating Procedures

Since May 2010, DNA-PROKIDS began to collaborate progressively with laboratories in 16 participating countries (Table 10.1). The project focuses its efforts on the distribution of DNA kits for collection of samples, on providing training and law enforcement resources in participating countries, on returning the processed samples to the country of origin that did not have the resources to carry out the analysis, on developing awareness campaigns, and on promoting the participation of the communities and in the definition of best performs to international DNA archive formation.

Depending on the capabilities of the country, the samples can be analyzed within the country or send to the UGR or UNT-HSC for the generation of DNA profiles and creation of databases. Once the samples arrive at UGR or UNT-HSC, they are processed

Table 10.1 DNA-PROKIDS: collaborations with the different participating countries

Country	Collaborating institution
Brazil	Department of Legal Medicine – Faculty of Medicine – USP
Panama	Institute of Legal Medicine and Forensic Sciences
Guatemala	National Institute of Forensic Sciences (INACIF)
Indonesia	Eijkman Institute for Molecular Biology
México	Attorney General's Office (PGR) and PGJ Baja California, PGJ Guanajuato, PGJ Chihuahua, PGJ Michoacán, PGJ Chiapas
Nepal	National Forensic Science Laboratory
Philippines	DNA Analysis Laboratory, NSRI – University of the Philippines
Spain	Laboratory of Genetic Identification – University of Granada
Sri Lanka	Molecular Medicine Unit. Faculty of Medicine, University of Kelaniya
Thailand	Royal Thai Police
USA	University of North Texas Health Sciences Center
El Salvador	Civil National Police
Bolivia	Public Ministry
Honduras	Institute of Legal Medicine – Public Ministry
Perú	National Police of Peru
Paraguay	Public Ministry

with the same methodology used for analysis of forensic evidence but separately preserved and managed. Samples are analyzed in a batch constituting of mostly 20 kits and analyzed as reference samples, i.e., collected under informed consent. Genetic comparison between alleged relatives and minors is carried out differently depending if there are prior suspicions about children's situation or they are unidentified minors. In the first case, the relatives provide information about the child, which facilitates direct genetic comparison between these alleged relatives and the minor. The second case involves unidentified minors, and the DNA-PROKIDS database is blindly searched; analysts do not try specific trios of child, mother, and alleged father. Once the study is carried out, UNT-HSC or UGR laboratory managers return this information as a statistical analysis, a likelihood ratio, or direct exclusion to the corresponding laboratory. Therefore, each participating country creates its own database to handle data that is generated in the institution itself or sent anonymously from the UGR or UNT-HSC.

10.2.3 Advances and Results of DNA-PROKIDS

DNA-PROKIDS collaborating institutions ensure that the kits received collect samples regularly. They inform the DNA-PROKIDS headquarters that the procedure is being carried out correctly. This allows determining the specific number of kits needed for the sampling based on the conditions and needs of each participating country.

A specific case that exemplifies the international action of DNA-PROKIDS and how it was possible to avoid the trafficking of minors in some specific cases was in

2010, after the earthquake of 7.3 scale that devastated the area of Prince Port (Haiti). Twenty-five Haitians minors were illegally trafficked to Santa Cruz (Bolivia) post-earthquake in early January. They were attended by two Haitians and Bolivian persons, and they were not able to confirm their familial relationship. Cops doubt about their kindship and avoid their traffic to Brazil or Argentina for illegal adoptions, child exploitation forced labor, etc. Thirteen of these children have been returned to their original families after DNA level identification. This was possible only due to the initiative led by DNA-PROKIDS and Bolivian General Fiscally. The rest 12 children that could not be identified at DNA level were kept in Bolivia, and the research is going on by police and judicial personnel for their proper identification and waiting to be delivered to the Haitian government.

10.2.4 DNA-PROKIDS Around the World

One of the aims with which DNA-PROKIDS began in 2004 was to expand beyond the pilot countries where the study was conducted: Guatemala and Mexico. Today, 16 countries are already involved throughout the world, and some more are interested in taking part in this project (Table 10.1).

In addition to these countries, the *United Nations Global Initiative to Fight Human Trafficking* (UN.GIFT) is also working to combat human trafficking, especially children and women. International Labor Organization (ILO), the Office of the United Nations High Commissioner for Human Rights (OHCHR), the United Nations International Children's Emergency Fund (UNICEF), the United Nations Office on Drugs and Crime (UNODC), the International Organization for Migration (IOM), and the Organization for Security and Cooperation in Europe (OSCE) jointly established UN.GIFT in March 2007 [2].

10.2.5 Resources Allocation for Each Country

DNA-PROKIDS seeks to encourage the countries with which it has an agreement to actively participate in order to eliminate as many cases of child trafficking as possible. It attempts to raise awareness among the different governments and the population of the current situation and the problems involved in the trafficking of minors.

As long as countries actively demonstrate efforts to address this problem, DNA-PROKIDS will send the necessary resources required by the institutions of each country.

After running a pilot program in Guatemala between 2006 and 2009, since 2010, more than 11,200 samples have been analyzed, generating more than 900 and 218 inclusion and exclusion results, respectively. Two hundred fifty-seven illegal adoptions were detected with these investigations.

10.2.6 Challenges and Implementation Difficulties

There are many challenges when genetic technologies are introduced to prevent and prosecute child trafficking. As this is an international program, DNA-PROKIDS has to adapt to the social and cultural attitudes toward genetic technologies of different countries; respective government, boarder security forces, and law enforcement agencies should be part of the initiative.

In addition, the beneficiaries of DNA-PROKIDS could not understand the pros and cons of sample collection, storage, and profiling completely. Many people believe to be obliged by providing samples or sense forced to contribute. In this regard, it is essential to ensure that everyone who wants to participate in DNA-PROKIDS gives willingly their consent.

In order to regularize the entire international project and to achieve these purposes, five key challenges are proposed, including the following:

1. *Adaptation of the memorandum of understanding and handling international collaborations*: The success of DNA-PROKIDS is based on strong and effective collaboration of the participating associations. The memorandum of understanding (MoU) aims to provide opportunities for all collaborating parties who want to help and organize as well as assign a specific, well-detailed role for each one and strengthen these partnerships.
2. *Guarantee a mission-based use of DNA-PROKIDS resources*: The resources allocated to each country, according to the development of the project they wish to cover, become difficult to apply since DNA-PROKIDS preserves autonomy of the participating countries. The main problem DNA-PROKIDS faces is the starting of sample collection. The problems of communication and the creation of agreements between the participating countries and DNA-PROKIDS are mainly due to the lack of understanding when dealing with the collection policy and its subsequent analysis. The implementation of programs in countries which have already signed MoUs has sometimes been suspended for weeks or months. Currently, police in participating countries are not obliged to collect contextual information related to samples or documents of DNA-PROKIDS, thereby minimizing privacy problems and administrative procedures. It is true that this collection of complementary background evidence could be handy for DNA-PROKIDS managers to control related cases. In addition, demanding background evidence can facilitate evaluation of programs and subsequently reduces the mismanagement of DNA-PROKIDS assets for other applications.
3. *Identification of methods to facilitate database security and management*: DNA-PROKIDS specifies that consignments should be accurately documented to minimize delays and interference. For example, collection kits should be documented as “relationship kits” and certain terminology such as “reagents,” “collection devices,” and “biological samples” should not be used. DNA-PROKIDS can assess the real-time status and position of the samples and sample collection kits. Since participating countries do not report as they send the samples to be analyzed, DNA-PROKIDS cannot track and control the flow of samples or know if

they have experienced difficulties in border barriers or other types of delivery problems.

4. *Identification of methods to facilitate security and privacy of the profiles*: DNA-PROKIDS currently limits who can access information about samples and prohibits the disclosure of confidential information on the basis of national and international laws. Similarly, databases, being individual and independent in each country, are managed according to their own laws and regulations. It is important to emphasize that there is no common database where all the data are pooled but independent, country-owned databases that can connect each other when necessary.
5. *For systematized and effective management of sample analysis*, the protocols could ensure a standard methodology for storage and the allocation of tasks to individuals and propose a schedule for sample processing. The protocol should be standardized, and a person should be authorized to be requested and receive information about the results of that search. For a good management, it is highly essential to determine whether profiles can be deleted and how an individual can be approached for deletion of his/her profile.
6. *Conduct education campaigns to encourage the prioritization of efforts against human trafficking, role of law enforcement agencies, and public participation*: Implementation of the program is highly influenced by the societal, economic, and political issues of the participating countries. In many countries where trafficking in persons is prevalent, there are other types of problems such as food and shelter provision and poverty alleviation, as they control drug crime and homicide, as well as management of political instability. All of them are often considered to be in priority for paying attention than to fight against human trafficking. This is why it is often challenging to involve other third parties in the collection of DNA samples in order to eliminate human trafficking. This situation can be aggravated in times of disasters (natural, wars, etc.), when an important event can cause multiple misfortunes to collapse into a "cascade effect." It also requires to spread consciousness regarding DNA-PROKIDS program and to encourage indigenous community to report suspected cases.

10.2.7 Financing

Nowadays, DNA-PROKIDS is possible, thanks to the support of national governments of participating countries, the University of Granada, the collaboration of the government of Spain (MAEC-AECID), the Andalusian Regional Government, and our main private sponsor, ZOGBI Commercial Distributor, a Mexican company (www.dczogbi.com).

Former sponsors to thank include BBVA, BMN (Granada Savings Bank), Fundación Botín, and very specially Life Technologies Foundation; we thank them for their help and support.

10.2.8 A Very Special Social Support

One of the disadvantages of globalization is the increasing rate of human trafficking, including prostitution, exploitation, and degrading labor conditions. DNA-PROKIDS project was exposed to Vatican in a meeting of 2013. During this meeting the current problem of human trafficking and the innovative technologies applied in this field were exposed. José Antonio Lorente, founder of DNA-PROKIDS, was one of the invited speakers of this international meeting among other 19 expertise in this field. Pope Francis was really impressed by DNA-PROKIDS program and its achievements, and it was declared that scientific advance in DNA analysis and databases should be used now and in the future. At the end of this meeting, it was concluded that governments should have an obligation to include DNA data in those cases of unidentified children or in those children living in highly risky situations, as forced migrations, wars, natural disasters, etc. [6].

10.2.9 Relevant Cases

10.2.9.1 DNA-PROKIDS in Melilla, Spain

Melilla is a Spanish city situated in the northern coast of Morocco, by the Mediterranean Sea and close to the Strait of Gibraltar (Fig. 10.1).

DNA-PROKIDS in Melilla started in 2013 when the Spanish law enforcement agency Guardia Civil detected that international mafia used human trafficking. Mafia boarded women with children in small boats to cross to Spanish continent by a rate around 1500–2000 € [7]. After the initial researches, Spanish agents discovered that most of these children have no biological descendant of these women. All these events triggered new intervention actions by the collaboration of the Interpol, minor district attorney, judicial authorities, Minor General Direction Health and Social Office of Melilla, and University of Granada (for DNA testing).

This protocol activated DNA testing of 12 sub-Saharan and Algerian persons, proving that ten cases were negative, so there was no biological relation among progenitors. Minors were captured and bought and subsequently given to immigrant adults, who were also exploited in Europe.

Mafia embarked the adults with the minors (usually under 10 years old) in small boats, with the physical, health, and life risks of this trip. Once, a minor death drowned body was collected from the sea, but this body has not been reclaimed by any immigrant adult.

In other situations when the Guardia Civil intercepted small boats, adult immigrants threatened to throw a minor to the sea or even to fire the small boat with all occupants.

There are also cases of a minor to be registered twice in the same childcare center by two different supposed biological mothers, once totally abandoned in the center. Similarly, in other time, several supposed mothers claimed for the same minor to be registered as a family unit. At the end, it was known that the minor was sold by his bio-

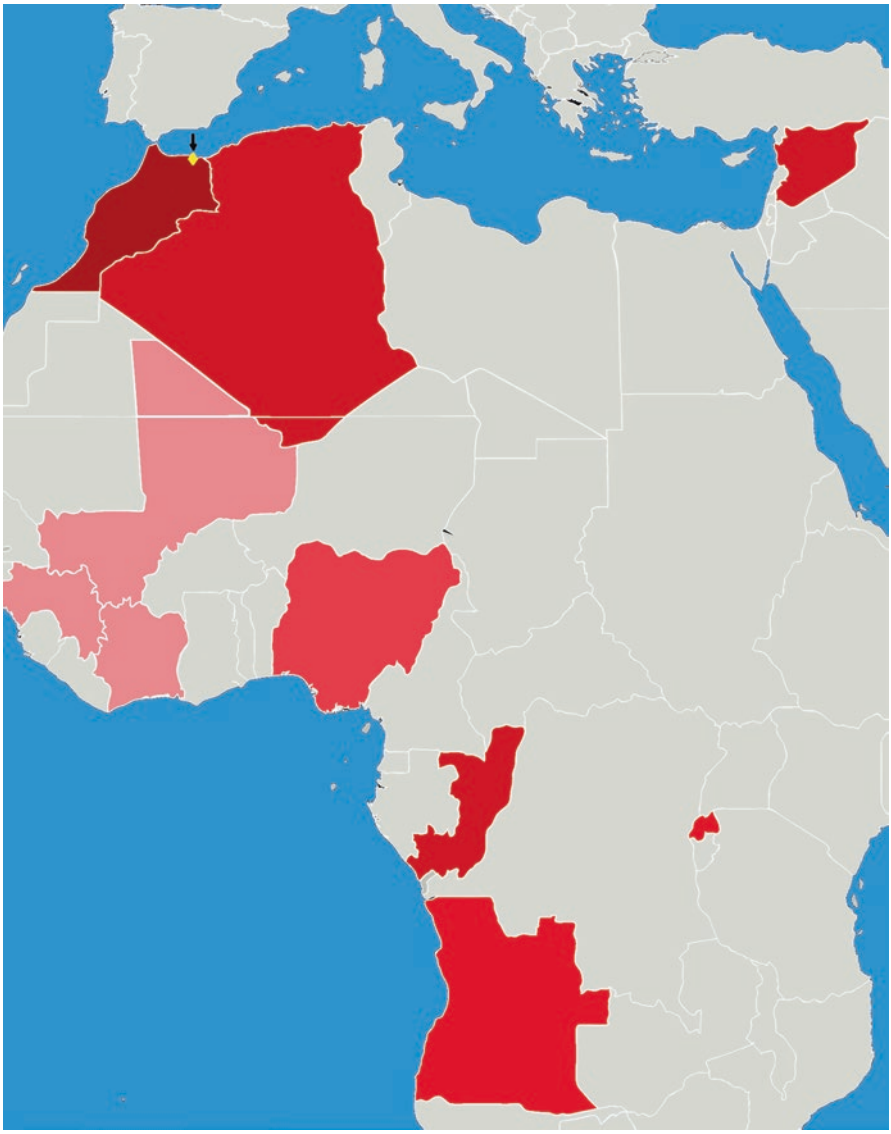

Fig. 10.1 City of Melilla is represented in yellow (*not to scale*); red colors represents the density of immigrants that arrived in that city in 2013, whose origins were Syria, Nigeria, Mali, Congo, Morocco, Rwanda, Algeria, Angola, Côte d'Ivoire, and Guinea

logical mother for 20,000 dinars, who also was in the same immigrant center. Rescued minors were under the tutelage of the social services of the Autonomous City of Melilla.

In addition, after conducting these investigations and the checking by different involved authorities, it has been possible to stop the mafia for using minors with this aim. So, one of the main DNA-PROKIDS objectives has been achieved; with these investigations people are more aware and human minor trafficking has declined.

10.2.9.2 Law Alba-Keneth (Guatemala)

The *Law of Alba-Keneth Alert System* was accepted and uncovered after implementation of DNA-PROKIDS in Guatemala. This name of law has been derived from two children, Alba Michelle España and Kenneth López Agustín. This law is an amalgamation of synchronized activities between public organizations allowing the children or adolescent acceleration, localization, and protection. In the case of murder or missing of a child, the convalescence and protection of the family should go side by side. All the participating organizations should abide to the *Law of Alert System Alba-Keneth*. Additionally, this law also anticipates the establishment of a database of missing and/or trafficked minors to support their families. Creation of a DNA database of the trafficked children and their blood relatives will enhance the analysis of biological relationship between individuals [8]. Alba Michelle España Díaz, an 8-year-old girl, living in the suburb of La Barrera, Camotán, Chiquimula, disappeared on June 14, 2007. Finally, her cadaver was recovered in a deplorable manner. Another 4-year-old boy from Japan, Keneth Alexis López Agustín, was missing since December 16, 2009. His cadaver was later recovered with signs of cut throat and was buried beneath a house's patio on the 23rd of December. In both cases, the accused admitted the involvement of huge amount of money for trafficking the children.

10.3 Genetic Identification

10.3.1 Samples' Collection

DNA samples from minors can be taken either directly with buccal swabs or with lancets to extract saliva or blood or, in special cases, using material containing DNA from these minors. However, to take relative's samples (mother, father, siblings, etc.), it must be taken in situ, with a prior consent. For this aim, DNA-PROKIDS has designed a collection kit in which both the sample extraction protocol and the data collection methodology are detailed. It also contains a sticker with a bar code that will identify the samples taken during processing (Fig. 10.2).

10.3.2 Information Collection

To make a good data collection and to obtain a good identification, the kit also presents/displays two tickets in which the personal data of each member has to be filled to whom the DNA test is going to be taken place. One of them is used to collect data from minors and the other one for the relative's data. These cards detail the names, surname, identification number, bar code (Fig. 10.3), date of birth, sex, photograph, and the code related to the relationship between relatives and minors. The various

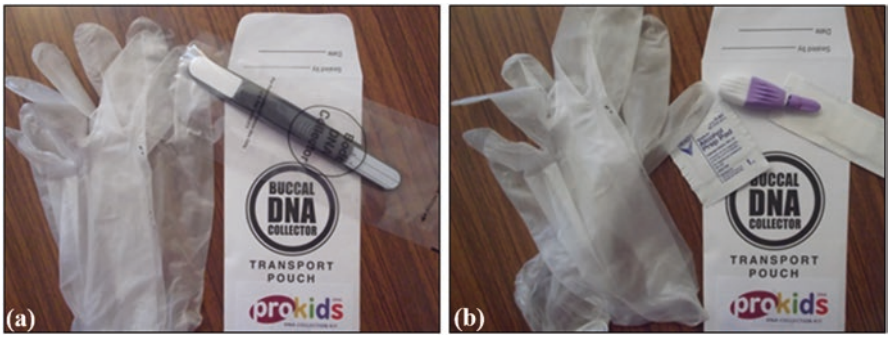

Fig. 10.2 Saliva extraction kit (left) and blood (right). The kit contains the necessary material for taking both types of samples: gloves; collector, which serves both to deposit the blood sample and the saliva; lancet; band aid; alcohol wipe; and envelope to store the biological sample

sections to be filled in are detailed both in Spanish and English with the idea of creating a universal understandable kit.

10.3.3 Databases and Software

DNA-PROKIDS uses genetic profiling software called M-FISys, from the English Mass Fatality Identification System (Fig. 10.4). This software was created to analyze DNA of unidentified remains of September 11, 2001, and to compare them with their relatives. It is used as a database since it allows storing case histories without relation between them. This software is a system of information management and uses collaboration and data exchange tools. It uses graphs with an intuitive interface, and, despite handling complex genealogies, in set it is an easier system to use than other software with similar features.

It presents different STR alleles with their corresponding complex formulas taking into account the allelic frequencies of each of them. In addition, this program allows the integration of other biometric data and the automatic report generation. One of the great advantages of this system is the exchange of confidential encrypted data with other DNA-PROKIDS databases. This means that M-FISys uses and requires a special and unique format for genetic files, allowing to export a genetic profile from one database to another in a different city or country. The sent file is encrypted to avoid that the receiving laboratory can see it or can import it into its database. It can only be used to search for direct identification or kinship matches. This feature has the advantage that data from different regions can be collated if one has the intuition that the missing child or the family of the missing child is in a different place. In addition, in order to avoid problems of confidentiality, it will not be known which profile is being checked or who is being collated, which is very useful in the Latin American countries participating in DNA-PROKIDS project.

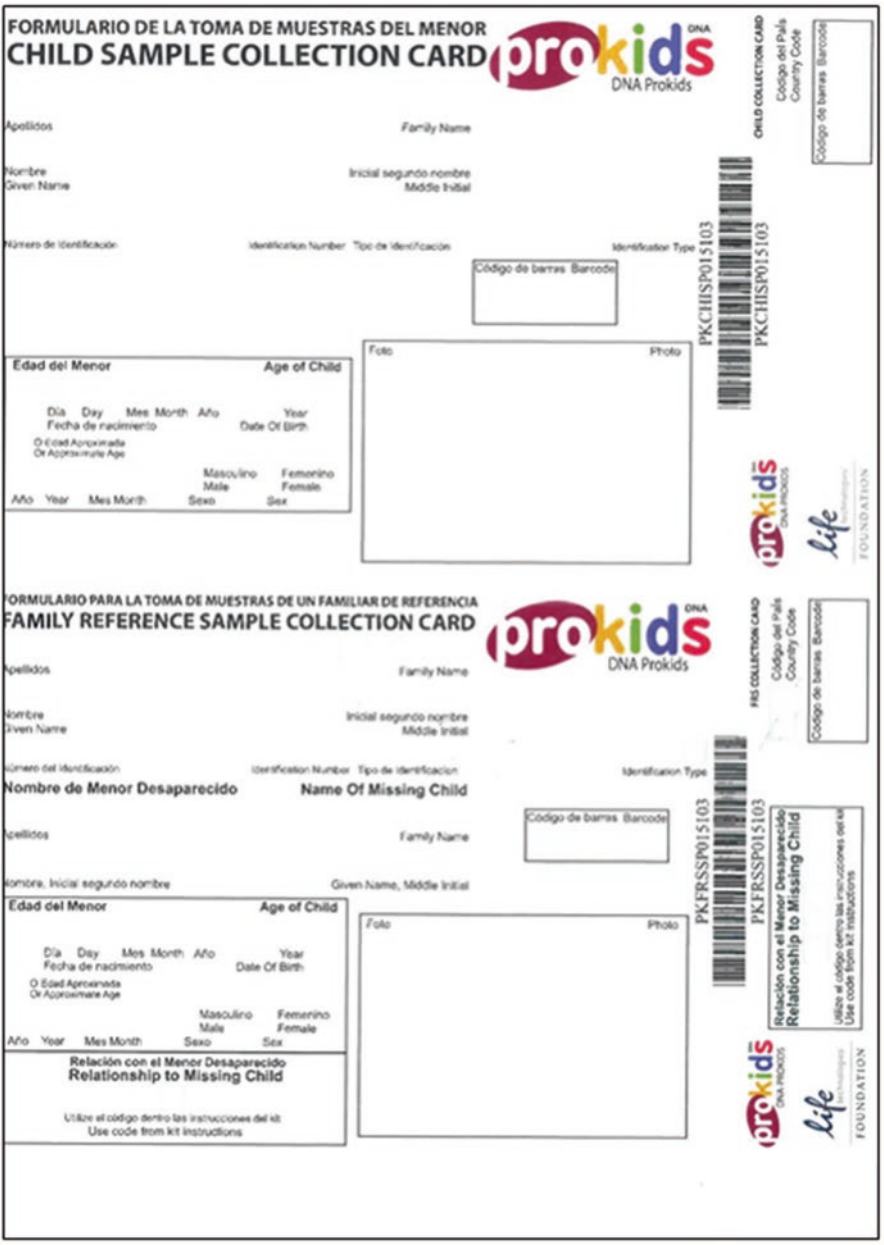

Fig. 10.3 Sample form for the child and the relative

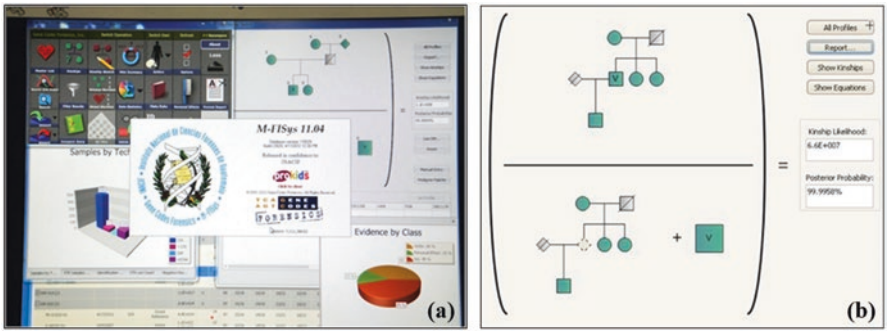

Fig. 10.4 **(a)** Option overview in M-FISys software. **(b)** By a simple creation of a family tree, you can enter the child data and his family. The program itself, based on complex calculations, returns the paternity index data and the resulting kinship probability

10.3.4 Main Technologies in Genetic Identification

DNA-PROKIDS initiative has developed standard kits for sample collection. The kit contains cotton swab, blood puncture device and informed consent form, documents related to chain of custody, and personal data of the donors. Additional details regarding sample collection and contact details also have to be mentioned (Fig. 10.2).

DNA-PROKIDS uses the same DNA typing methodologies used for forensic analysis and missing person identification. Extensively validated molecular markers such as STRs and single nucleotide polymorphisms (SNPs) are extensively used for DNA-PROKIDS analysis [9–12]. SNPs appear due to variations from single nucleotide substitutions, insertions, or deletions at one or more positions along the target genome. It occurs around 1 in 1000 bp in any human genome which accounts for 85% of human genomic variation. Nowadays, there are data about millions of SNPs that are available, and a subgroup of these is useful for human identification. Analysis of SNPs is helpful in increasing the likelihood ratio in cases with an inclusion result and also in dealing with problems associated with mutations which are easy to find when working with STRs and mtDNA. All samples are processed by standardized forensic commercial kits and subsequently analyzed in a Genetic Analyzer Equipment 3130.

Moreover, the analysis of mitochondrial DNA is also targeted (primers HV1 and HV2 regions and whole sequence) following Sanger methodology and finally analyzed in sequencer 3130 system.

New NGS (next-generation sequencing) technologies are offering new insights in forensic genetics. Parallel sequencing of appropriate genetic loci with the help of NGS plays an important role in criminal justice system (Fig. 10.5). The use of currently available NGS platforms, i.e., Illumina GA, SOLID, and 454, has indicated their usefulness in population studies with proper optimization [13]. The use of PCR technology is the biggest disadvantage of NGS as it allows the generation of many PCR artifacts such as introduction of alien base sequences or others. In this

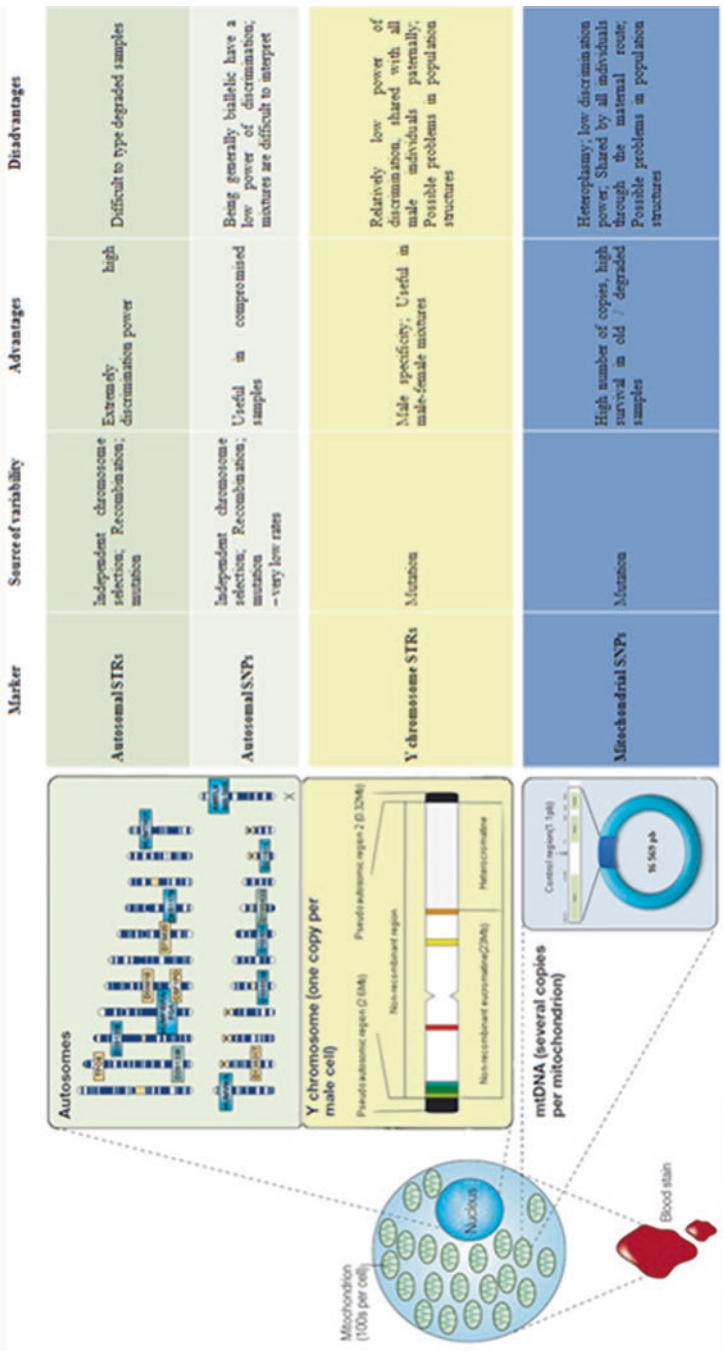

Fig. 10.5 Sources of human variability. Figure based and modified from Jobling et al. 2004 [15]

regard, the third-generation HT technology is proven to be useful by sequencing a single DNA molecule independent of a PCR amplification step mostly by synthesis-mediated sequencing. However, the usefulness of this advanced technology HT-NGS supplied by many vendors throughout the globe needs to be validated properly prior to its use [14]. The need of a very qualified and trained personal and the high cost of the data are also two of the limitations of these methodologies.

References

1. Huckerby J, Gu A (2010) Más Allá de las Fronteras: Explorando los Enlaces entre Trata. Globalización y Seguridad, Bangkok
2. UNODD (2009) UN.GIFT – global initiative to fight human trafficking. http://www.ungift.org/
3. Palmbach TM, Blom J, Hoynes E et al (2014) Utilizing DNA analysis to combat the world wide plague of present day slavery – trafficking in persons. Croat Med J 55:3–9. https://doi.org/10.3325/cmj.2014.55.3
4. Eisenberg A, Schade BS (2010) DNA-Prokids: using DNA technology to Help Fight the Trafficking of Children. Forensic Mag 7:15–19
5. Lorente JA, Saiz M, Álvarez-Cubero MJ, Álvarez JC (2012) La genética contra la trata de menores: La identificación genética de menores desaparecidos. In: Nuevos restos en la lucha contra la trata de personas con fines de explotación sexual. Un enfoque multidisciplinar. Thomson Reuters, Pamplona
6. Scarpa S (2008) Trafficking in human beings: modern slavery. Oxford University Press, Oxford, p 230. ISBN 9780199541904
7. Rodrigo S (2013) Rescatados diez niños usados como pasaportes humanos para entrar en Melilla
8. Tanto POR (2010) Ley del Sistema de Alerta Alba-Keneth. pp 1–6
9. Bulbul O, Phillips C, Argac D et al (2009) Internal validation of 29 autosomal SNP multiplex using a ABI 310 genetic analyser. Forensic Sci Int Genet Suppl Ser 2(1):129–130
10. Daniel R, Sanchez JJ, Nassif NT et al (2009) Partial forensic validation of a 16plex SNP assay for the inference of biogeographical ancestry. Forensic Sci Int Genet Suppl Ser 2(1):477–478
11. Gross TE, Zaumsegel D, Rothschild MA, Schneider PM (2013) Combined analysis of two different ancestry informative assays using SNPs and Indels in Eurasian populations. Forensic Sci Int Genet Suppl Ser 4:e25–e26. https://doi.org/10.1016/j.fsigss.2013.10.012
12. Walsh S, Chaitanya L, Clarisse 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
13. Harismendy O, Ng PC, Strausberg RL et al (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 10:R32
14. Ziętkiewicz E, Witt M, Daca P et al (2012) Current genetic methodologies in the identification of disaster victims and in forensic analysis. J Appl Genet 53:41–60. https://doi.org/10.1007/s13353-011-0068-7
15. Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. Nat Rev Genet 5:739–751

Further Readings

Palmbach TM, Blom J, Hoynes E, Primorac D, Gaboury M (2014) Utilizing DNA analysis to combat the world wide plague of present day slavery – trafficking in persons. Croat Med J 55:3–9

Eisenberg A, Schade BS (2010) DNA-PROKIDS: Using DNA technology to Help Fight the Trafficking of Children. Forensic Mag:7, 15–19

Lorente JA, Saiz M, Álvarez-Cubero MJ, Álvarez JC (2012) La genética contra la trata de menores: la identificación genética de menores desaparecidos. In: Nuevos retos en la lucha contra la trata pers. con fines explot. sex. un enfoque Multidiscip. Thomson Reuters, Pamplona

Lorente JA, Entrala C, Alvarez JC, Lorente M, Arce B, Heinrich B et al (2002) Social benefits of non-criminal genetic databases: missing persons and human remains identification. Int J Legal Med 116:187–190

Ge J, Sun H, Li H, Liu C, Yan J, Budowle B (2014) Future directions of forensic DNA databases. Croat Med J 55:163–166

Lorente JA, Entrala C, Alvarez JC, Arce B, Heinrichs B, Lorente M et al (2001) Identification of missing persons: the Spanish "Phoenix" program. Croat Med J 42:267–270

Borovko SR, Korban VV, Kritskaya SV, Yeumenenka SA (2009) Missing people: problems of identification of unknown bodies using DNA database. Forensic Sci Int Genet Suppl Ser 2:260–262

Lorente JA, Entrala C, Alvarez JC, Lorente M, Villanueva E, Carrasco F et al (2000) Missing persons identification: genetics at work for society. Science 290:2257–2258

da Silva LAF, Vilaça W,Azevedo D,Majella G,Silva IF,Silva BF (2009) Missing and unidentified persons database. Prog Forensic Genet 13, Proceedings of the 23rd International ISFG Congress, 2:255–257

Alvarez-Cubero MJ, Saiz M, Martinez-Gonzalez LJ, Alvarez JC, Eisenberg AJ, Budowle B et al (2012) Genetic identification of missing persons: DNA analysis of human remains and compromised samples. Pathobiology 79:228–238

Saiz M, Alvarez-Cubero MJ, Alvarez JC, Lorente JA Forensic genetics against children trafficking: missing children genetic identification. In: Handbook of missing persons. Springer, Suiza, p 2016

Chapter 11
Three Decades of DNA Evidence: Judicial Perspective and Future Challenges in India

G. K. Goswami and Siddhartha Goswami

Abstract During last 30 years, DNA has emerged as a potent forensic tool in advancing justice in India as well as throughout the globe. DNA profiling assists in human identification with great precision and is used for various purposes including adjudication of civil and criminal matters. In criminal domain, DNA helps in stitching crime with criminal and in identification of victim. In civil courts, DNA has increasingly been used in resolving paternity disputes by identifying putative father despite not being recognized under Indian legal lexicon. Presumption of legitimacy under Section 112 of Indian Evidence Act, 1872, recognizes socio-legal father provided a child is born within lawful wedlock. Earlier 'presumptive' father, a legal fiction, and 'putative' father, a genetic reality, were assumed to be the one and the same person, but DNA has exposed the 'genetic truth' of childbirth by lifting the veil from 'twin fatherhood' and has opened a Pandora's box in Indian legal panorama by heralding coexistence of both socio-legal and putative father especially under laws of inheritance. This article attempts to explore the legal trends for paternity determination by using DNA profiling through examining various judicial pronouncements of Indian courts.

Keywords DNA profiling · Presumption of legitimacy · Paternity · Admissibility · Forensic evidence · Genetic truth · Twin fatherhood

G. K. Goswami (✉)
Gujarat Forensic Sciences University, Gandhinagar, India

Academy of Central Bureau of Investigation, Gaziabad, India

S. Goswami
Jindal Global Law School, Sonipat, Haryana, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_11

11.1 Introduction

DNA profiling, as a genetic eyewitness, has globally attained an aura of invincibility in the last three decades, being an impartial and unbiased scientific evidence to ensure accuracy, transparency and fairness in administration of justice. In 1986, DNA entered in criminal justice system in a spectacular fashion when Prof. Alec Jeffreys of Leicester University in England assisted the United Kingdom police to identify the rapist of two minor girls[1] [1]. Interestingly in this case, DNA as forensic tool not only helped in identifying the real offender Colin Pitchfork but also exonerated Richard Buckland despite his admission of committing rape and murder of one teenager girl near Leicester [2]. Since DNA analysis provides a unique and specific profile for an individual like thumb impression, Prof. Jeffreys coined the term DNA fingerprint [3] which subsequently came to be known as DNA profiling, DNA typing or DNA testing.

Justice system is best described as a quest for truth, and forensic tools including DNA technology became an important ally in that exploration. In 1988, DNA fingerprint was first admitted in the court as evidence in Florida v. Tommy Lee Andrews [4]. In fact the judiciary was so overwhelmed with DNA evidence that Justice Joseph Harris had observed, 'DNA technology could be the greatest single advance in the search for truth, conviction of the guilty, and acquittal of the innocent since the advent of cross-examination' [5]. Justice Kennedy of the US Supreme Court in Maryland v. King [6] observed the usage of DNA technology:

> The advent of DNA technology is one of the most significant scientific advancements of our era. The full potential for use of genetic markers in medicine and science is still being explored, but the utility of DNA identification in the criminal justice system is already undisputed. Since the first use of forensic DNA analysis to catch a rapist and murderer in England in 1986, law enforcement, the defence bar, and the courts have acknowledged DNA testing "unparalleled ability both to exonerate the wrongly convicted and to identify the guilty. It has the potential to significantly improve both the criminal justice system and police investigation practices.

DNA has emerged as a reliable but reactive forensic tool (for detection of crime) rather than proactive tool (for prevention of crime). As a potent forensic tool, DNA not only assists in connecting crime with the perpetrator but also in proving innocence of a wrongly convicted innocent person [7]. DNA has proved its worthiness in enhancing justice to sexually exploited victims and in strengthening a child's right to know parentage with more efficacy and precision. It further helps in preserving dignity of the victim during medical examination (recording sexual history of victim, conducting 'two-finger penetration test') and cross-examination by defence causing embarrassment during trial proceedings. However, DNA test entails

[1] In the United Kingdom, two minor girls were raped and murdered in two different incidents, first in Narborough, Leicestershire, in November 1983, and the second in Enderby, also in Leicestershire, in July 1986. Interestingly, the real accused *Colin Pitchfork*, who later confessed the felony and was convicted for life imprisonment, bribed a man named Ian Kelly for giving blood sample in his name to conceal his identity during DNA dragnet.

a wide range of genetic surveillance resulting in both use and misuse of technology especially in relation to the privacy rights of an individual.

Since 1987, DNA evidence has been assisting Indian courts in adjudicating on both criminal and civil matters. The first case solved with the help of DNA technology was a parentage dispute case [8]. In 1991, a disputed paternity was settled using DNA fingerprinting in Kunhiraman v. Manoj [9]. Over the last 30 years, DNA has enabled right-based perspective for vulnerable sections of society by ensuring fast track justice [10].

This article delves upon the appreciation of DNA evidence by the Indian judiciary during last three decades and succinctly sketches the progressive judicial approach for strengthening various procedural protocols involved in forensics. The structure of the article is designed to discuss the role of DNA in criminal adjudication especially in sexual offences, manslaughter, terrorist attacks, etc. Three phases of judicial approach for determining paternity of a child using DNA have been identified ranging from presumption of legitimacy to unveiling the truth of genetic parentage. This new approach of the 'quest for genetic truth' may potentially conflict with the existing law on presumption of legitimacy which shall be deliberated in greater detail. Further various issues like Innocence Project and challenges before investigating agencies and judiciary in using DNA evidence shall also be discussed in the subsequent sections.

11.2 Forensics of Human Identification: Uniqueness of DNA Fingerprinting

In the society, establishing the identity of an individual is necessary for several purposes. Since long, photograph, sketch, biometrics and fingerprints figured are among the most commonly used scientific means for identifying a person, but these evidentiary tools normally have several limitations especially when the body undergoes purification and mutilation. Serological analysis, to some extent, rescue from such adverse situations. The conventional blood tests for human identification include (i) the blood cell grouping (the ABO blood cell antigen test), (ii) human leukocyte antigen (HLA) testing which is a tissue-typing test for determining whether an organ transplant will be acceptable to recipient's system, (iii) serum groups and (iv) haemoglobin variants [11]. However, these serological tests also suffer from various limitations [12].

The advent of genetic forensics marked the beginning of human identification with greater precision [13]. DNA resides in the nucleus of each cell, dead or alive, and furnishes a personal genetic blue print of an individual which has gained appreciation in forensic biology [14]. DNA does not change during lifetime and remains preserved for centuries together. The uniqueness of DNA lies in sequencing of building blocks called nucleotide, which is constituted by base pairing of adenine (A), thymine (T), guanine (G) and cytosine (C) existing in the twisted rope ladder

structure (double helix). The genetic make-up of an individual consists of a linear arrangement of the nucleotides constituting the genetic code which reads three nucleotides at a time like a recipe. The sequence of base pairing is the biological fingerprint and key to the uniqueness for personal identification except for Siamese twins. Microgram sample for DNA testing may be extracted from the blood, saliva, hairs, bones, excretory content, semen, etc.

The DNA testing methodologies include RFLP, PCR, STR, RAPD and mDNA (mitochondrial) [15]. Touch DNA (t-DNA) technique has given a new impetus as few cells only can be amplified into several copies of DNA using polymerase chain reaction (PCR) technique [16]. Human beings have 99.9% of their DNA in common pool, and Jeffreys successfully discovered an ingenious way of pinpointing the differences. The key features of DNA profiling include sequences of code known as short tandem repeats (STRs), which differ widely among individuals and provide nearly infallible means of connecting a genetic sample to its donor source as claimed by Jeffreys [17]. In fact Jeffreys original technology in 30 years underwent advancement from Southern blot to PCR, from radioactive to fluorescence and slab gels to capillary electrophoresis [18]. The technological upgradation has further enhanced credibility and reliability of DNA in administration of justice.

DNA forensic provides invaluable assistance in analysing, evaluating and solving issues like (i) linkage to crime, (ii) identification of the culprit or the victim and (iii) correct parentage in paternity disputes [19]. The matrix of DNA mainly addresses human identification based on analysis of biological sample and has established credibility in both civil and criminal legal fraternity. In criminal justice realm, DNA has successfully been used in investigating sexual offences and other violent body offences especially in cracking cold cases where conventional methodologies of investigation were found lacking. Non-human DNA also plays a vital role in forensic analysis of evidence [20].

11.3 DNA in the Courtroom

DNA profiling emerged as a panacea in justice system because human identification has a vital role both in criminal and civil adjudication. In 1987, DNA entered into legal dominion and emerged as a reliable forensic tool for ‘policing the human identity’ [21]. The first reported use of DNA profiling was in a civil dispute to prove a familial relationship when a Ghanaian boy was refused entry into the United Kingdom for want of proof that he was the son of a woman having right of settlement in the United Kingdom [22]. The immigration authorities contended that the boy could be the nephew of the woman but not her son. DNA test proved a strong probability of mother-son relationship and the UK government permitted him immigration [23].

In 1987, Robert Melias became the first accused convicted for charges of rape based on DNA evidence in England [24]. The successful conviction of Gary Dobson and David Norris for the racist murder of teenager Stephen Lawrence was made

possible by the retrieval of miniscule quantity of victim's DNA on the killer's clothing. This conviction established DNA evidence as 'magic bullet', 'genetic eyewitness' and eureka in detection of suspects in blind criminal cases [25]. In Andrew v. Florida [26], the Florida Appeal Court first accepted DNA-led identification in sexual assault case by convicting Tommie Lee Andrews [27]. Among civil matters, use of DNA in resolving paternity disputes has greater academic significance since law of legitimacy of childbirth has conflicting issues with the scientific findings revealed by DNA testing. The issue of parentage shall be dealt with, in the subsequent sections of the discourse.

11.3.1 Admissibility of DNA Evidence

In the global arena, despite legal system having considerable experience in dealing with scientific opinion, DNA in courtrooms posed more challenges because of its procedural complexities. In the United States, Canada and other jurisdictions, the courts applied two major standards, namely, the Frye test (general acceptance test) [28] and the Daubert test (relevance-reliability test) [29]. Justice Blackburn in the *Daubert* case casted duty upon the trial judge to act as 'scientific gatekeepers' for assessing the validity of scientific evidence as empirical scientist. The Daubert test was first applied in the United States v. Martinez [30] for determining admissibility of DNA test. As additional standard, Australian courts used 'prejudicial effect test' for DNA evidence by evaluating the prejudicial effect of the scientific evidence with its probative value [31]. Justice McInerney observed, 'If scientific testing in the particular case is unreliable or if it has a tendency to produce a misleading or confusing impression for the jury, or if the weight to be afforded in the results is so minimal as to preclude the jury being satisfied beyond reasonable doubt that the Crown has established the fact which it seeks to prove, then clearly I have a duty to exclude it from the jury - whether it is a result of ruling that the evidence is inadmissible or whether it is excluded in the exercise of my discretion' [32]. The courts in England used 'usefulness standard' for appraisal of the scientific evidence [33].

In fact DNA fingerprinting successfully transformed from diagnostic tool to forensic evidence being approved by the courts with utmost caution and vigilance [34]. In People v. George Wesley, the court placed enormous reliance on scientific evidence and established DNA as 'gold standard test' in the 'search for truth' [35]. Subsequently the courts recognized certain caveats on admissibility of DNA as reported through a series of commentaries and articles [36]. In the landmark case, People v. Castro [37], the New York Court formulated a three-prong test for admissibility of DNA evidence.[2] The court held that in the Castro case, theory and technique

[2] Firstly, is there a theory, which is generally accepted in the scientific community, which supports that DNA test can produce reliable results? Secondly, are there currently existing techniques or experiments which are generally accepted in scientific community, capable of producing reliable results in DNA identification?, and, thirdly, did the testing laboratory perform the accepted scientific techniques in analysing the forensic sample in the particular case?

of DNA test had well established, but particular DNA protocol observed by FBI laboratory was unable to suffice the inculpatory identification of the accused. In United States v. Kozminski, the Federal Court of Appeal held that four factors must be assured to uphold the admission of expert testimony: (1) a qualified expert, (2) testifying on a particular subject, (3) in conformity to a generally accepted explanatory theory and (4) the probative value overweighing any prejudicial effect [38].

The State of Tennessee v. Paul Ware [39] introduced mitochondrial DNA (mtDNA) into the criminal justice realm by analysing red hairs recovered from the autopsy of a 4-year-old victim of rape. The courts upheld the admissibility of mtDNA since it qualifies the criterion of relevance-reliability test. In a bank robbery case, detail appraisal of mtDNA was conducted by the Apex Court in the Connecticut v. Pappas [40]. Considering heteroplasmy and contamination as major weaknesses of mtDNA, the court observed that probative value of mtDNA may not be outweighed merely on prejudices. In R v. Murrin [41], the British Columbia Supreme Court held that probability of two individuals having similarities in mtDNA profiles was higher than the nuclear DNA profile and mtDNA may not prove identity conclusively but may be used for corroboration. On the issue of heteroplasmy, the court has observed:

> In our view, neither the discovery of heteroplasmy within the control region nor the alleged danger of "false inclusions" provides a reason for rejecting evidence of this novel scientific theory or technique. Like the ever-present danger of contamination, the existence of heteroplasmy is a complicating factor, which requires substantial exercise of judgment by mtDNA analysts. The effect of heteroplasmy and its treatment in the databases will have to be explained carefully to the jury to enable them to understand the true frequency of the questioned mtDNA sequence in the general population. (para 121)

11.3.2 Precautions for DNA Profiling

It may be a dangerous fallacy to believe that DNA evidence is infallible. Year 1997 was a watershed year when the reliability of DNA evidence was questioned and cases of 'failed DNA' were noticed. DNA per se may be 'infallible', but translating biological sample into DNA profile involves multifaceted anthropogenic interventions making it far from infallible.[3] Identification, lifting, preservation, labelling and transportation of biological samples with intact chain of custody (CoC) are some

[3] People v. Castro 545 N.Y.S.2d 985 (Sup. Ct.1989). It was the first case where DNA profile was seriously challenged for DNA admissibility. Castro was accused of double murder of his neighbour and her baby daughter. A blood stain was lifted from Castro's watch for a DNA match with the victims. The court held that: (i) DNA identification theory and practice and techniques applied are generally accepted among the scientific community. (ii) The accreditation of the DNA laboratory was also needed to its reliability to observe the established scientific standards. The Castro ruling supports the proposition that DNA identification evidence of exclusion is more presumptively admissible than DNA identification evidence of inclusion. In Castro's case, the court ruled that DNA tests could be used to show that blood on Castro's watch was not his, but tests could not be used to show that the blood was that of his victims. The court observation in this case was based on Frye test established in Frye v. the United States 293 F. 1013 (D.C. Cir.), 1923. However, Castro's case was never tried. He pleaded guilty to the murders in late 1989.

basic requirements for scientific evidence to be admissible in the court. During trial proceedings, 'tampering with the sample' remains the most common allegation labelled by the defence lawyers against the prosecution [42].

For credible expert opinion, the forensic laboratory and the scientific procedure used for forensic analysis must be accredited. The scientist conducting the test must be qualified and experienced fulfilling the requirement for an expert. The Supreme Court in Rajiv Singh v. State of Bihar [43] has commented upon the shoddy investigation based on poor quality of forensic analysis conducted by inexperienced person of private laboratory in the guise of State Forensic Science Laboratory. The High Court of Andhra Pradesh, in Patangi Balrama Venkata Ganesh v. State of A. P [44], has observed that DNA test needs quality control. In M. V. Mahesh v. State of Karnataka [45], the High Court of Karnataka held that benefit of doubt arising from malpractice or irregularities in scientific processing of evidence ought to go to the accused.

11.4 DNA Evidence in Indian Courts

In last three decades, ever since DNA emerged as forensic tool for unfolding the truth, Indian judiciary has proactively relied upon DNA as 'champion' evidence during administration of justice both in civil and criminal matters as represented under Fig. 11.1.

The issue of scientific standardization of DNA technology has broadly been settled, but the legality of DNA evidence is questioned in the courts since DNA expert is yet to be legally recognized under Indian lexicon mainly Section 45 of the Indian Evidence Act, 1972, and Section 293 of Cr. PC. In order to bridge this legal void, several attempts have been made to enact law on DNA testing, but none of the Human DNA Profiling Bill could see the light of the day in the Indian Parliament. Till date there is no regulatory legal framework in place except five guidelines culled by the Apex Court in Gautam Kundu v. State of West Bengal to regulate DNA testing. These guidelines have grossly been overlooked several times by the constitutional courts, and N. D. Tiwari v. Rohit Shekhar is a glaring case for forcing the respondent against his consent to give blood sample. DNA profile also involves several right-based issues like privacy and protection against self-incrimination, and judicial appreciation of circumstances on case-to-case basis is required before ordering to give DNA sample.

11.4.1 Self-Incrimination, Consent and Court Permission for Taking Bodily Samples

Right against self-incrimination, a well-celebrated basic human right, emanates from the Latin maxim *Nemo debet prodere se ipsum* (no man is bound to betray himself). The Fifth Amendment of the US Constitution and international human

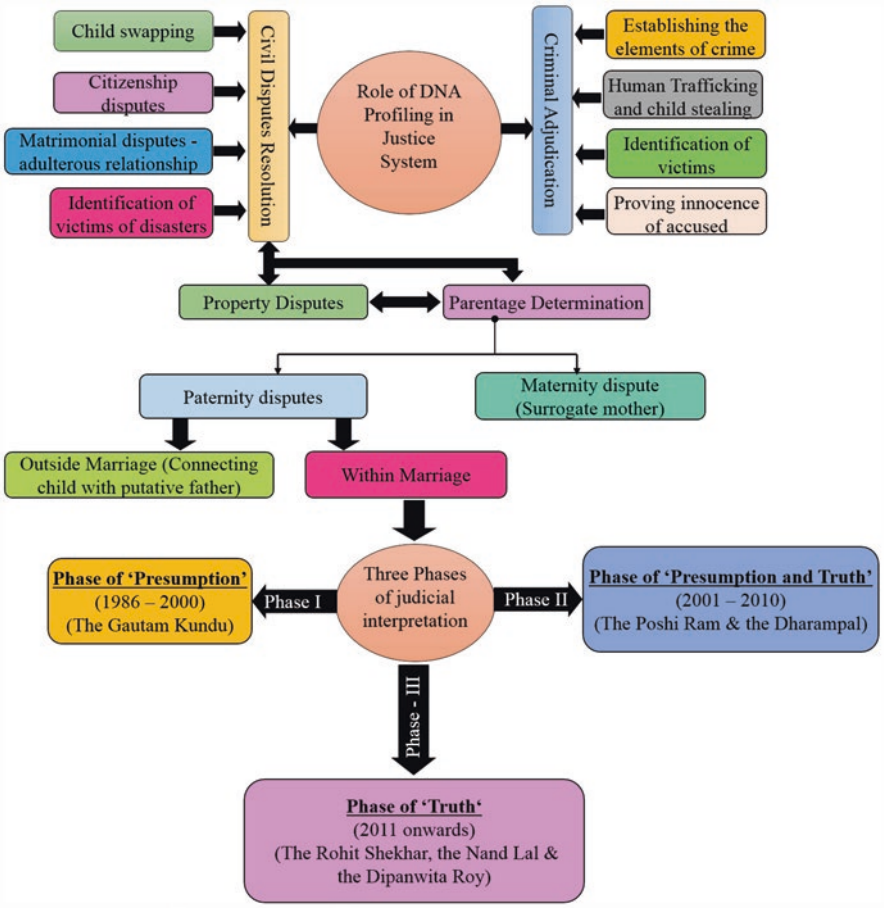

Fig. 11.1 DNA profiling in administration of justice

right instruments also protect against testimonial compulsion.[4] The right to silence of the accused is best described under Saunders v. the United Kingdom [46]. Ethics and reliability are the bastion for the right against self-incrimination ensuring protection of life and limb of the subject. Article 20(3) of the Indian Constitution enables a subject against testimonial compulsion as non-derogable right. Worldwide, the judiciary has faced the challenge of deciding whether taking forensic samples from a subject attracts testimonial compulsion or not. Historically, under the common law, blood test results were admissible only where parties voluntarily submitted to the tests [47]. The courts had no inherent power to compel a witness or a party to submit to the blood tests [48].

[4]The Fifth Amendment of US Constitution reads that 'No person… shall be compelled in any criminal case, to be a witness against himself'.

The apparent imbalance between Article 20(3) and provisions of the Indian Evidence Act for collection of evidence has appeared since inception of the Indian Constitution [49]. Several judgements of High Courts were referred to the Apex Court, which were clubbed and examined by the 11 judges constitutional bench resulting into landmark judgement of the State of Bombay v. Kathi Kalu Oghad [50]. The bench held that compelling an accused to submit his specimen of handwriting or signature or impression of his thumb, fingers, palm or foot for purpose of comparison during investigation does not attract infringement of Article 20(3). In Selvi v. State of Karnataka [51], the Apex Court tested the constitutional validity of extracting evidence by the use of advanced technology especially narcoanalysis, brain mapping (P-300 test) and polygraph (lie detector). The three-judge bench, heavily relied on the Oghad, held that forced intrusion into the mind of the subject not only violates Article 20(3) but also intrudes on the right to life with dignity under Article 21 of the Indian Constitution. The consent becomes necessary, if the scientific technique is invasive and affects bodily integrity.

11.4.1.1 Consent for DNA Sampling: Criminal Cases

Reflecting on an argument that taking samples of blood or semen inflicts torture on the subject and no provision in Cr. PC empowers the court to compel a person to give biological samples, the High Court of Andhra Pradesh has reflected by observing that '… Section 53 and 54 of the new Code there is such a power. In fact Section 53 provides that while making such an examination such force as is reasonably necessary for that purpose may be used. Therefore, whatever discomforts that may be caused, when samples of blood and semen are taken from an arrested person, it is justified by the provisions of section 53 and 54 Cr. PC [52]'. The Allahabad High Court in Jamshed v. State of UP [53] has observed that in the interest of justice, an examination of person can be made under Section 53(1) Cr. PC, which includes blood sample and an examination of organs inside the body. The High Court of Orissa held that in the interest of justice, courts under Section 53 Cr. PC. may issue directions to IO for collection of blood samples from the accused in criminal cases, but discretion must be used by the court with due precaution only when strong prima facie case is made out [54].

In 2005, the Criminal Procedure Code was amended introducing Section 53-A which enables a medical practitioner to take biological sample of the rape accused for conducting scientific tests including DNA profiling. Thus, for the accused of rape, consent is not required for taking his blood sample. In other criminal cases, the court decides on merit for taking blood sample against the consent. In answer to whether forced DNA testing against consent, in criminal cases, violates right against self-incrimination protected under Article 20(3) of the Indian Constitution, 1950, the judiciary has preferred harmonious construction of competing interests of the individual and the society in ordering DNA testing [55]. The courts have wider powers to order for taking DNA sample in criminal cases compared to civil matters.

11.4.1.2 Consent for DNA Sampling: Civil Cases

Civil courts have no explicit power to direct the subject to give blood sample for forensic tests. The Madras High Court, indicating a gap in the law, has observed that 'there is no procedure either in the Civil Procedure Code or the Evidence Act which empowers the Courts to enforce the party to give blood sample' [56]. In 1975, Kerala High Court held that taking blood sample exerts constraint on personal liberty and cannot be carried out without explicit consent [57]. In Dikhtar Jahan v. Mohmmed Farooq [58], the Supreme Court held that 'No person can be compelled to give sample of blood for analysis and no adverse inference can be drawn against a person on account of such refusal.' Similar view was observed earlier in Bipinchandra Shantilal Bhatt v. Madhuriben Bhatt [59] by the Gujarat High Court. In Hargavind Soni v. Ramdulari [60], Madhya Pradesh High Court observed, 'The blood grouping test is a perfect test to determine questions of disputed paternity of a child and can be relied upon by Courts as a circumstantial evidence. But no person can be compelled to give a sample of blood for blood grouping test against his will and no adverse inference can be drawn against him for this refusal'.

Subsequently, in several cases judiciary opined that the court may order for DNA test but cannot compel giving blood sample and the court may draw adverse view against the party in case of refusal [61]. Indeed Section 114 of the Indian Evidence Act, 1872, enables the court to draw an adverse inference, if the party does not produce relevant evidence in his power and possession. However, privacy right of the subject underpins his stand for refusal to undergo blood test, and the role of the court becomes crucial in balancing between the juxtaposed legal precepts.

The Apex Court in the Goutam Kundu v. State of West Bengal (hereafter called the Goutam Kundu) has emphasized upon consent and culled out five guidelines for the courts considering order of DNA test to determine paternity[5] [62]. The Supreme Court, in Banarsi Das v. Mrs. Teeku Dutta [63], upheld the High Court judgement for not compelling the defendant by the trial court to give blood sample for DNA testing for issuing the succession certificate. The Apex Court held that DNA test must be ordered in deserving cases and not as a matter of routine. The Court further observed that:

> We may remember that Section 112 of Indian Evidence Act was enacted at a time when the modern scientific advancement with DNA and RNA test were not even in contemplation of Legislature. The result of the genuine DNA test is said to be scientifically accurate. But that is not enough to escape for the conclusiveness of the section 112 of the Act, i.e. if a husband and wife were living together during the time of conception, but the DNA test reveals that the child was not born to the husband, the conclusiveness in law would remain irrebuttable.

[5] Five guidelines are (1) that courts in India cannot order blood test as a matter of course; (2) wherever applications are made for such prayers in order to have roving inquiry, the prayer for blood test cannot be entertained; (3) there must be a strong prima facie case in that the husband must establish non-access in order to dispel the presumption arising under Section 112 of the Evidence Act; (4) the court must carefully examine as to what would be the consequence of ordering the blood test; whether it will have the effect of branding a child as a bastard and the mother as an unchaste woman; and (5) no one can be compelled to give sample of blood for analysis.

However, in the recent past, courts are ordering DNA test against the consent of the subject which shall subsequently be deliberated in detail.

11.4.2 DNA Evidence in Criminal Cases

DNA in criminal adjudication has emerged as eureka especially for solving blind cases where conventional methods of evidence collection would have failed to disclose the truth. Investigating officers (IOs) play vital role in collection of evidence as emphasized by the Supreme Court, 'The investigating officers are the kingpins in the criminal justice system. Their reliable investigation is the leading step towards affirming complete justice to the victims of the case. Hence, they are bestowed with dual duties i.e. to investigate the matter exhaustively and subsequently collect reliable evidences to establish the same' [64]. In India, IOs are accused of unbridled discretion in collection of scientific evidence and even compromising with the quality of evidence collection. The procedural protocols for evidence collection may help to set the discretion right for meeting the end of justice.

11.4.2.1 Conventional Cases of Rape, Murder and Terror Attack

Rape is a crime against femininity normally committed at an isolated place with little scope of ocular witness, and hence any scientific evidence will be of immense significance for corroboration. Many a times, killing of the rape victim to eliminate the evidence aggravates the challenge for the investigator. Sexual abuse and brutal murder of *Ms. Priyadarshini Mattoo* [65], *Naina Sahani* (Tandoor case) [66], poetess *Madhumita Shukla* [67] and *Nirbhaya* [68] are a few examples of conviction by using DNA profiling. Similarly murder of victims after rape was established by DNA scrutiny in cases of self-proclaimed God-man, namely, *Swami Premananda* [69] and *Swami Shraddhananda* [70].

Many a times, the body disintegrates defacing the individual's identity in terror incidents, thus posing challenge for law enforcement agencies of victim identification. DNA assisted significantly to identify body parts of Rajiv Gandhi [71], Beant Singh [72] and several others who lost their lives in such attacks. Over 800 victims were identified by DNA analysis of biological samples recovered from the scene of crime pertaining to the 9/11 attack on World Trade Centre in United States [73]. In the recent past, DNA has become a reliable tool for corroborating with greater precision the cold cases of body offences and terror attack.

Identification of putrefied dead bodies and human skeletal remains has credibly been established by matching the genetic profile of the deceased with that of the surviving blood relatives. Recently in Sheena Bora murder case, DNA collected

from the exhumed skeletal remains (molar teeth and femur) matched with her mother and sibling, establishing her identity even after 3 years of her murder.[6]

11.4.2.2 Human Trafficking and Child Theft

Missing children and human trafficking have interlinkages, and exploitation of trafficked persons is well known. Over the years, identification of trafficked children has become a challenge for rescue operations and rehabilitation strategies. DNA can play a significant role in uniting such children with their families by matching DNA profile of parents and close blood relatives [74]. It is suggested to maintain DNA data of orphan children and the parents of trafficked persons to help management of rescued trafficked victims.

11.4.3 DNA in Civil Dispute Resolution

In civil matter the preponderance of probability constitutes the baseline for standard of proof, while in criminal matters, prosecution has to prove the guilt beyond reasonable doubt before the court of law. Since 1986, DNA has been used for resolving various riddles pertaining to civil matters, mainly parentage determination, child swipe, inheritance disputes, immigration and identification of putrefied dead bodies for settling various claims and inheritance-related issues. The parentage determination remains a vital issue in the society and hence has been discussed in detail.

11.4.3.1 Parentage Determination in India

Parentage determination continues to be a vital sociolegal issue for individual identity, lineage and inheritance. Before DNA profiling, the Indian courts had relied upon Section 112 of the Indian Evidence Act, 1872 (here after called Section 112), while deciding paternity dispute and the Supreme Court of India, in Dukhtar Jahan v. Mohammed Farooq [75], held that presumption of legitimacy is a conclusive proof except in case of non-access to each other at any time when the child could have been begotten. The Court observed, 'This rule of law based on the dictates of justice has always made the courts incline towards upholding the legitimacy of a child unless the facts are so compulsive and clinching as to necessarily warrant a finding that the child could not at all have been begotten to the father and as such a

[6] The skeletal remains (a skull, pelvis, teeth and limb bones) were recovered from the forest area of Gagode Khurd village in Pen Tehsil of Raigad district in the state of Maharashtra (India).

Sheena murder case: DNA from Raigad bones matches Indrani Mukerjea's blood sample" DNA Daily News Analysis, 08 September, 2015 (http://www.dnaindia.com/india/report-sheena-murder-case-dna-from-raigad-bones-matches-indrani-mukerjea-s-blood-samples-2122955).

legitimization of the child would result in rank injustice to the father. Courts have always desisted from lightly or hastily rendering a verdict and that too, on the basis of slender materials, which will have the effect of branding a child a bastard and its mother an unchaste woman'. However, in case the child is allegedly begotten outside lawful wedlock, the law classifies the child as illegitimate. In such situations, forensic analysis assists the court in determining putative fatherhood.

Traditionally, special protection is given to the status of legitimacy of a child under Section 112 which can neither be lightly repelled nor will be allowed to be broken or shaken merely by a balance of probability. Proof against the mother for committing adultery will not by itself suffice for proving the illegitimacy of the child [76]. The evidence of non-access, for the purpose of repelling, must be strong, distinct, satisfactory and conclusive; and the standard of proof against presumption of legitimacy must be similar to the standard of proof of guilt in a criminal case [77].

The rigours in presumption of legitimacy are justified under public policy to protect the best interest of the child and the mother. The sociolegal and biological father are presumed to be one and the same person under the fictional notion of legitimacy, enshrined under Section 112, although in real life both may be different persons. In fact 'legitimacy' and 'paternity' are distinct concepts; the earlier deals with legal presumption of paternity, while the latter establishes genetic composition of a child established by the genetic tests [78]. Traditionally the concept of legitimacy of a child only resolves paternity disputes since maternity was considered a biological reality before advent of surrogacy. Accordingly, under Section 112 legitimacy of a child denotes only paternity; and maternity determination under the existing Indian legal framework is completely missing. In the era of surrogacy, maternity may equally be questioned. Status of parentage determination under Indian lexicon is represented in Fig. 11.2.

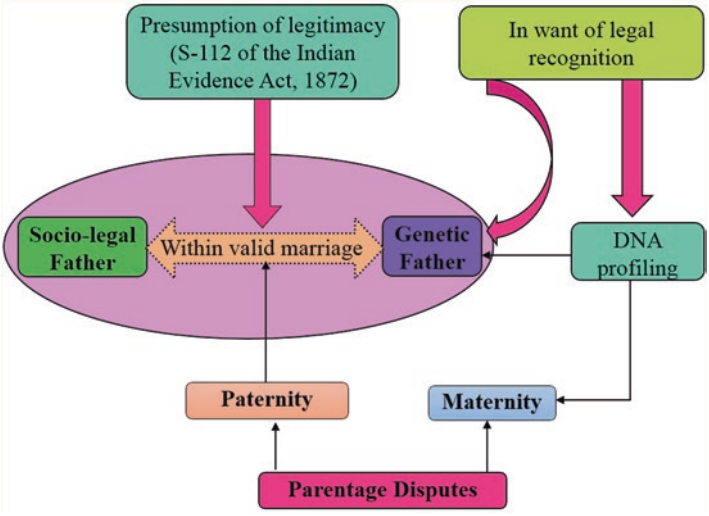

Fig. 11.2 Status of parentage determination under Indian Law

The pioneer case of Kunhiraman vs Manoj [79] set the tone of paternity determination in India by DNA test. Manoj was born to an unmarried young employed girl as a result of sexual liaison with neighbour who promised to marry her but later denied. In the father's name column of the birth register, Kuniraman's name appeared as declared by the mother, but he objected to it and denied paternity. On behalf of the child, the mother filed a petition under Section 125 Cr. PC for child's maintenance. Chief Judicial Magistrate found him to be the putative father and awarded monthly maintenance of Rs. 500. In the appeal, DNA test was conducted which proved him to be the pater of the child. This case marked the beginning of DNA as credible evidence in the Indian courts. Over the last three decades, Indian courts have increasingly appreciated DNA results for paternity determination despite the potential challenge of the legislative intent of Section 112.

11.4.3.2 Three Phases of Judicial Appreciation of DNA in Parentage Determination

In DNA-led paternity determination cases, various trends of judicial approach in India have been observed during this study. Initially the higher judiciary placed reliance on the law of evidence, and 'presumption of legitimacy' received preference over 'truth of paternity' derived from DNA analysis. Subsequently, judiciary started realizing the importance of 'truth' in paternity disputes, and judicial pronouncements of late clearly indicate a mixed response towards 'legitimacy' and 'truth'. The Rohit Shekhar v Narayan Dutta Tiwari [80] (hereafter called the Rohit Shekhar) has marked the beginning of judicial standpoint emphasizing on the 'truth of biological paternity' although the 'presumption of legitimacy' was capable of determining the sociolegal father of the petitioner. The judicial trends of DNA-led paternity determination in India, based on the analysis of various judicial pronouncements, have broadly been categorized into three phases as discussed below.

Phase of 'Presumption' (1987–2000)

In the first phase, the Indian courts preferred traditional approach where presumption of legitimacy remained sacrosanct for adjudication in accordance with the Section 112. The Indian judiciary faced two distinct situations for determining paternity, firstly paternity when challenged by the husband based on infidelity of the wife, but child was born within lawful wedlock, and secondly where victim mother gave consensual sexual access for various reasons including promise (deceitful) to marry, and in such cases question of paternity was raised before the court for deciding issues like culpability of putative father, claim for maintenance, inheritance and the right to know one's lineage [81]. In second situation, competent consent of female for access to sexual liaison constitutes a vital ground for distinguishing between civil and criminal liability.

The Goutam Kundu is the pioneer case, where the Supreme Court had placed reliance on Literal Interpretation of the Statute (Section 112) in advising the subordinate courts not to order blood test a as matter of routine but to carefully examine the consequences of ordering blood test with respect of not branding a child as bastard and the mother an unchaste woman. The court observed, 'This section requires the party disputing the paternity to prove non-access in order to dispel the presumption, 'access' and 'non-access' mean the existence or non-existence of opportunities for sexual intercourse; it does not mean actual cohabitation'. The court further held that the 'wife cannot be forced to give blood sample and no adverse inference against her for this refusal'. The Goutam Kundu intended to protect the right of privacy by emphasizing on free and informed consent of the concerned parties prior to undergoing DNA scrutiny.

During this phase of deciding paternity dispute for the child born within lawful wedlock, courts preferred 'presumption of legitimacy' rather than ordering a DNA test to know the 'genetic truth' of childbirth. This approach supports public policy by providing access to justice by providing maintenance and other rights to poor and marginalized female victims and their children born out of sexual exploitation. During this phase, court preferred literal interpretation of law and remained well within the boundaries of legislative intent [82].

Mixed Phase of 'Presumption and Truth' (2001–2010)

The beginning of the twenty-first century witnessed burgeoning demand from the petitioners to conduct DNA test for determining paternity. During this phase, initially Indian judiciary continued to prefer a balanced approach between competing interests of an individual and the community. In Kamti Devi v. Poshi Ram [83], the Supreme Court observed, 'The result of a genuine DNA test is said to be scientifically accurate. But even that is not enough to escape from the conclusiveness of section 112 of the Act e.g. if a husband and wife were living together during the time of conception but the DNA test revealed that the child was not born to the husband, the conclusiveness in law would remain irrebuttable [84]'.

Gradually the judiciary begins to place more reliance to know the genetic truth of paternity revealed by DNA testing. The Kerala High Court in Sajera v. P. K. Salim [85] ordered for blood test considering strong prima facie case of non-access of wife to husband. The footprints of judicial intent to prefer DNA test for paternity determination appear in Kanchan Bedi v. Gurpreet Singh [86]. Hon'ble Vikramjit Sen J. of Delhi High Court has observed, '… it appears to me to be difficult to resist that the law, as it presently stands, does not contemplate any impediment or violation of rights in directing persons to submit themselves for DNA test, especially where the parentage of a child is in controversy for the grant of maintenance' (para-8). Delhi High Court in Alika Khosla v. Thomas Mathew held that refusal to undergo

DNA test would bar a party from challenging the paternity of the child [87]. The Supreme Court, in Banarasi Dass v. Teeku Dutta [88], resolved the issue of contradiction between presumption and truth as under:

> Section 112 of the Evidence Act was enacted at a time when the modern scientific advancement and DNA test were not even in contemplation of the Legislature. The result of DNA test is said to be scientifically accurate. Although section 112 raises a presumption of conclusive proof on satisfaction of the conditions enumerated therein but the same is rebuttable. The presumption may afford legitimate means of arriving at an affirmative legal conclusion. While the truth or fact is known, in our opinion, there is no need or room for any presumption. Where there is evidence to the contrary, the presumption is rebuttable and must yield to proof. Interest of justice is best served by ascertaining the truth and the court should be furnished with the best available science and should not be left to bank upon presumptions, unless science has no answer to the facts in issue. In our opinion, when there is a conflict between a conclusive proof envisaged under law and a proof based on scientific advancement accepted by the world community to be correct, the latter must prevail over the former. We must understand the distinction between a legal fiction and the presumption of a fact. Legal fiction assumes existence of a fact which may not really exist. However presumption of a fact depends on satisfaction of certain circumstances. Those circumstances logically would lead to the fact sought to be presumed. Section 112 of the Evidence Act does not create a legal fiction but provides for presumption.

Ordering DNA test in suits of paternity dispute within lawful marriage, being extremely delicate and sensitive issue, poses a dilemma before the courts to opt between Section 112 and DNA testing. However, courts were more inclined towards DNA results. Consequently, by reversing the *ratio decidendi* in the Goutam Kundu, a three-judge bench of the Supreme Court in Sharda v. Dharam *Pal* [89] (hereafter called the Sharda) has culled out guidelines for the subordinate courts for deciding on DNA testing: (i) a matrimonial court has the power to order a person to undergo medical test; (ii) passing of such an order by the court would not be in violation of the right to personal liberty under Article 21 of the Indian Constitution; and (iii) the Court should exercise such a power if the applicant has a strong prima facie case and there is sufficient material before the Court. Despite the order of the court, if a respondent refuses to submit himself to medical examination, the court will be entitled to draw an adverse inference against him under Section 114 of the Indian Evidence Act, 1872.[7]

An order of the State Women Commission of West Bengal for conducting DNA test to determine paternity was challenged in Bhawani Prasad Jena v. Orissa State Commission for Women [90] (hereafter called the Bhawani Prasad Jena). The pregnant wife demanded maintenance both for herself and the child in womb, but the husband denied it based on suit filed for nullity of the marriage alleging non-consummation

[7] The Courts hold similar *ratio* in Dwarika Prasad Satpathy v. Bidyut Prava Dixit (1997) 7 SCC 675, Amarjit Kaur v. Harbhajan Singh (2003) 10 SCC 228, and Kamalanatha v. State of Tamil Nadu Appeal (crim.) 611–612 of 2003 in the Supreme Court of India, decided on April 05, 2005. The legal question arises, what shall be the impact of adverse inference or DNA test result when Section 112 does not recognize any other person as legitimate father than one who is the legitimate husband of the mother of child. Merely determining the putative father without there being accepted in society and recognized under the law may not yield desired result for the petitioner.

of the marriage. In an appeal preferred by the husband against the order for conducting DNA test, considering the best interest of the child, the Apex Court held that between the Goutam Kundu and the Sharda, the judge must prefer harmonious construction, and DNA test may be ordered if a strong prima facie case is made out. The Court held, 'The court has to consider diverse aspects including presumption under section 112 of the Evidence Act; pros and cons of such order and the test of 'eminent need' whether it is not possible for the court to reach the truth without use of such test [91]'. In this phase, on the pretext of no prima facie ground in some cases, court orders for DNA test were quashed in appeals [92]. On the similar facts, courts opted either for Section 112 or DNA test, resulting into legal ambiguity. During this phase judiciary appeared to be relying more upon the biology of impregnation, thus deviating from the sociology of legal presumption of childbirth.

Phase of 'Truth' (2011 Onwards)

Indian judiciary, in the recent past, has increasingly placed higher reliance on DNA for revealing the biological truth of paternity. In Rohit Shekhar v. Narayan Dutt Tiwari [93] (hereafter called the Rohit Shekhar), the judiciary preferred to enquire about putative father of the petitioner; and the respondent was forced to involuntarily submit blood sample for conducting DNA test, deviating from the guidelines culled in the Goutam Kundu. The Court referred the judgement in Re G (Parentage: Blood Sample) as held therein, 'Justice is best served by truth. Justice is not served by impeding the establishment of truth. No injustice is done to him by conclusively establishing paternity. If he is the father, his position is put beyond doubt by the testing, and the justice of his position is entrenched by the destruction of the mother's doubts and aspersions. If he is not the father, no injustice is done by acknowledging him to be a devoted stepfather to a child of the family. Justice to the child, a factor not to be ignored, demands that the truth be known when truth can be established, as it undoubtedly can'. [94] Further the Court referred *Re* H and G (Children Paternity Blood Test). Over 30 years ago in his speech in S v McC, [95] Lord Hodson said, 'The only disadvantage to the child which is put forward as an argument against the use of a blood test, not for therapeutic purposes but to ascertain paternity, is that the child is exposed to the risk that he may lose the protection of the presumption of legitimacy' [96].

In the Rohit Shekhar, a paradox emerged as to whether the best interest of child[8] was protected by revealing the genetic truth of petitioner's birth at the cost of branding him a bastard and his mother an unchaste woman, since at the time of the petitioner's birth, his mother was under valid marital obligation to another person. The High Court observed, 'There is of course the vital interest of child to not be branded illegitimate; yet the conclusiveness of the presumption created by the law

[8] The phrase 'the best interest of child' denotes the well-being of a child and Article 3(1) of the UN Convention on the Right of the Child, 1989 (UNCRC): For the purposes of the present Convention, a child means every human being below the age of 18 years unless under the law applicable to the child, majority is attained earlier.

in this regard must not act as a detriment to the interests of the child. The protective cocoon of legitimacy should not entomb the child's aspiration to learn the truth of her or his paternity'. [97] The Rohit Shekhar opened a flood gate in the courtrooms and set the judicial tone in projecting DNA as preferred forensic option over 'fictional notion of legitimacy' under Section 112 while deciding civil disputes involving disputed paternity.

The judicial approach of preferring science over law was further reinforced in Nandlal Wasudeo Badwaik v. Lata Nandlal Badwaik [98] (hereafter called the Nandlal Badwaik), where a child was born during the continuance of the valid marriage and under the aegis of Section 112, the appellant would be the legitimate father of the girl child. DNA profile match excluded the appellant as biological father. The test was repeated on the request of wife, but the second report further confirmed exclusion of the husband as putative father. At this stage, the wife requested the court to determine legitimacy under Section 112. In an appeal preferred by the husband pertaining to maintenance of wife and the alleged daughter, the Apex Court preferred DNA-approved genetic truth over the presumption of fact and legal fiction by concluding, 'The husband's plea that he had no access to the wife when the child was begotten stands proved by the DNA test report and in the face of it, we cannot compel the appellant to bear the fatherhood of a child, when the scientific reports prove to the contrary. We are conscious that an innocent child may not be bastardized as the marriage between her mother and father was subsisting at the time of her birth, but in view of the DNA test reports and what we have observed above, we cannot forestall the consequence. It is denying the truth. Truth must triumph is the hallmark of justice' [99].

The Nandlal Badwaik became pathfinder in Dipanwita Roy v. Ronobroto Roy [100] to uphold the decision of family court and the high court allowing DNA test for determining not only the paternity of newborn child but also to conclusively testify the veracity of accusations of infidelity labelled by the husband in the divorce petition filed in the family court under Section 13 of the Hindu Marriage Act, 1955. The Apex Court was, therefore, firm in holding that proof based on DNA test would be sufficient to dislodge a presumption under Section 112. Thus, the Goutam Kundu guidelines are now a matter of the past, and the judiciary is increasingly relying upon DNA test for variety of cases.[9]

Matters related to disputed maternity have also been coming up in Indian courts. The order of trial court for conducting DNA probe to compare genetic code of parties to establish genetic affiliation through maternal line from a common ancestor

[9] In property disputes, DNA technology is frequently used by higher judiciary to ascertain claims of the petitioner for establishing genetic linkages with the respondent. Darshan Singh v. Amarjit Singh @ Surjit Singh @ Kaka Singh Criminal Petition No. 2614 of 2014 Swarna Kanta v. Amarjit Singh @ Surjit Singh @ Kaka Singh v. Darshan Singh, Criminal Petition No. 3104 of 2014, both petitions dealt together in the Punjab and Haryana High Court, decided on July 21, 2014; Smt. H. Susheela v. G. Hanumanthappa Regular First Appeal No. 1219 of 2013, In the High Court of Karnataka, decided on September 25, 2014; and Namdeo Babasaheb Korde v. Babasaheb @ Babrao Ramkrishna, W.P. No. 7402 of 2012, In the Bombay High Court, decided on December 21, 2013.

was challenged in Sube Singh v. Smt Shanti Devi [101], and the Punjab and Haryana High Court upheld the order citing the ratio of the Nandlal Badwaik for using principle of exclusion of legal presumption to truth determined by the DNA test, exclaiming aphoristically, '… "Truth must triumph" is the hallmark of justice [102]'. The Nandlal Badwaik and the Dipanwita Roy became path-breaking judgements representing an era of complete departure from the existing law and earlier judicial precedents of the Apex Court such as the Goutam Kundu, the Banarsi Dass and the Bhabani Prasad Jena.

11.4.3.3 Analysis of Judicial Approach

India represents a melting pot of social milieu. DNA has assisted courts to protect the right of a sexually exploited female duped under false promise of marriage and her child born out of such relationship. In the recent past, the courts have indeed shed off the 'cocoon of presumption of legitimacy' and passionately allowed the 'scientific truth' to prevail, but in the absence of legislative intent, it appears as judicial overreach. Hence, the conscious bastardization of an innocent child and branding the mother of easy virtue need judicial reconsideration. Further recent judicial trend of deciding inheritance disputes by DNA test may abridge privacy and sexual rights of the mother branding her of immoral character. In the absence of statutory standing for separation of putative and sociolegal father, judicial passion for searching the 'genetic truth' may lead to legal confusion for defining rights and duties of 'twin fathers' vis-a-vis the child.

11.5 Future Challenges

DNA evidence in court room may face several challenges in view of the changing legal framework and the technological advancement in medically assisted reproduction. Few challenges have been discussed below.

11.5.1 Parentage in Assisted Reproduction

Surrogacy of womb, sperm and ovum has potentially diffused parentage for a child from conventional two parents system [103]. In gestational surrogacy,[10] in the absence of genetic linkage, the womb donor may not be identified even by DNA which continued to be a forensic challenge in case one desires to know about the

[10] Surrogacy has two variants, viz. gestational and traditional. The gestational surrogate mother has no genetic contribution to the foetus and only carries artificially fertilized embryo to the term. In traditional surrogacy the ovum belongs to the surrogate.

surrogate mother. Further genetic modifications like gene repair or substitution, etc. of embryo may complicate the parentage riddle. In fact assisted reproduction may be misused by various kinds of traffickers to produce a surrogacy-led child for committing heinous crimes like terrorism, contract killings, sex trading, forced labour, forced organ donation, forced marriage, begging, surrogacy and other transnational organized crimes. If anonymous gametes and womb are used for making a baby who is subsequently used to commit crimes, then even DNA may not be able to establish the genetic antecedents. Thus, unbridled surrogacy is likely to be exploited for harbouring 'criminal robots' without having identity, parentage, pedigree and nationality. In this field research is solicited.

11.5.2 Newly Emerging Sexual Offences

The Criminal (Amendment) Act, 2013, has widened the scope of defining rape beyond peno-vaginal penetration.[11] Traditionally medicolegal reports are based on examining injuries caused during sexual assault. The DNA analysis also rests upon transfer of biological content like semen, blood, hair, saliva, etc., from aggressor to the victim. In the absence of physical injuries and transfer of biological contents, the forensic tools may not be able to help in finding the truth necessary for administering justice. Further several newly defined sexual offences such as voyeurism [104] and stalking [105] may have limited scope for forensic tools to corroborate the allegations. These offences may be vulnerable for misuse by way of false accusation and secondary victimization of innocent persons.

11.5.3 Innocence Project: An Exit to Freedom

The secondary victimization of people by false implication in various bodily offences poses a challenge before the judiciary. The famous project 'Convicted by Juries, Exonerated by Science' [106] resulted in exonerating many innocent convicts in the United States by proving DNA profiling as finger of God to rectify wrongful conviction [107]. In Indian scenario, where cases of false allegations and tainted oral testimony-based wrongful conviction are alleged in abundance, Innocent Project may be a panacea to reverse the injustice. Further, the Cr. PC must be amended to facilitate preservation of case exhibits for a minimum period of four decades so that any subsequent forensic inventions may help to revisit conviction cases, if need be.

[11] Section 375 of Indian Penal Code: A man is said to commit rape if he penetrates his penis, or any part of his body or any object into the vagina, the urethra or anus of a woman or make her to do so with him or other person, against her will or without her consent.

11.5.4 Professionalism in CJS

Criminal justice system in India suffers from piecemeal approach and needs infusing professionalism at large. Justice has primarily been considered the key domain of judiciary under the aegis of 'fair trial' despite the fact that without ensuring transparency in evidence collection, fair trial remains a utopian preposition. Despite having legislation in place, investigative agencies are in habit of being reluctant in gathering scientific evidence due to several reasons. Section 166-A of the Indian Penal Code criminalizes the inefficient investigation and prosecution; the Apex Court has also expressed anguish on shoddy investigation and has emphasized for ensuring fair, competent and effective investigation [108]. Indeed, capacity building and competent supervision of various stakeholders of criminal justice system may be some of the correctional measures to accomplish fairness in administration of justice.

11.5.5 Legal Framework for Forensic Evidence in India

Section 45 of the Indian Evidence Act, 1872, applicable for both criminal and civil matters, deals with opinion of experts upon foreign law, science and art or as to identify handwriting or finger impressions [109]. However, for criminal proceedings, Section 293 of Cr. PC entails 'reports of certain government scientific experts' of chemical examination, explosive, fingerprint and serology. Interestingly despite being widely accepted in the courts, DNA expert lacks legal recognition. India has attempted several times to legislate the Human DNA Profiling Act, but efforts are yet to be fructified.

11.6 Conclusion

In the last three decades, DNA has widely been used in Indian courts for ensuring justice. In criminal matters, if chain of custody is maintained, DNA, without any legal controversy, greatly helps in corroborating the primary evidence [110]. In cold cases, the trace biological samples for DNA analysis collected from scene of crime assist in unfolding the truth. However, the cherry picking approach of investigator in evidence collection poses challenges. Procedural protocols on evidence collection and capacity building may bring the desired change. It is needed to introduce schemes for reassuring justice to wrongly convicted persons through DNA testing. The Supreme Court of India may consider to frame new guidelines for courts to issue orders for DNA test in the light of consent and right to privacy of the subject in consonance with the Article 20 (3) of the Indian Constitution.

Indian judiciary, during the journey of 30 years, has been confronted with the presumption of legitimacy and genetic truth of child's birth. A judicial trajectory from the Goutam Kundu to the Nandlal Badwaik via the Rohit Shekhar is a tale of 'journey from presumption to genetic truth' for paternity determination. However, genetic truth-based judicial approach has opened a Pandora's box for legal discourse regarding the coexistence of both putative father and sociolegal father of a child. The judicial adventurism by the trial courts for deciding property disputes using DNA in alleged cases of diffused parentage may further create complications rather than administering justice; it warrants corrective attention of both judiciary and legislature. In changing social milieu and technological advancement, over a century old, Section 112 needs overhaul to ascertain the parentage and not be limited to presumptive legitimacy of a child.

References

1. Sanders J (2000) Forensic case book of crime. True Crime Library/Forum Press, London, p 229 ISBN 1-874358-36-2
2. Tande CM (1989) DNA typing: a new investigatory tool. Duke Law J 1989(2):474–494
3. Jeffrey S, Alec J, Wilson V, Thein SL (1985) Hypervariable 'Minisatellite' regions in human nature. Nature 314:67; "Individual-specific 'Fingerprints' of human DNA," Nature 316(1985):76
4. Andrews v. State of Florida (1988) 533 So.2d 841
5. People v. Wesley 140 Misc. 2nd 306, 533 N. S. Y. 2nd 643 (Co. Ct. 1988)
6. 133 S.Ct. 1958 (2013).District Attorney's office for Third Judicial Dist. v. Osborne, 557 U.S.52, 55 (2009)
7. Edward C, Lundregan T, Miller N, McEwen T (1996) Convicted by juries, exonerated by science: case studies in the use of DNA evidence to establish innocence after trial. U. S. Department of Justice
8. Jayaraman KS (1989) Cut price fingerprints. Nature 340:175
9. II (1991) DMC 499)
10. In various cases like Gautam Kundu v. State of West Bengal 1993 (2) Scale 994. O. 6/95 and N. D. Tiwari v. Rohit Shekhar FAO(OS) No. 547/2011, the High Court of Delhi, decided April 27, 2012, the paternity was decided by DNA Profiling and justice was delivered to victims including maintenance to abandoned children and concubines. Lakshmi Kant Pandey v. Union of India [1984] AIR SC 469; Nandlal Wasudeo Badwaik v. Lata Nandlal Badwaik [2014] Criminal Appeal No. 24 of 2014 (Special Leave Petition Crim. No. 8852 of 2008, January 06, 2014); Rajiv Gandhi assassination (D.No.1151/1998, SC); Priyadarshini Mattoo (Santosh Kumar Singh v. CBI, 139 (2007) DLT 407, I (2007) DMC 654), and Poetess Madhumita Shukla (Amar Mani Tripathi v. State of Uttar Pradesh through CBI, Crim. Appeal No. 1248/2005, SC, Sept. 26, 2005). Verma SK, Goswamy GK (2014) DNA Evidence: The current perspective and future challenges in India, FSI. pp 183–9
11. Davis AR (1994) Are you my mother? The scientific and legal validity of conventional blood testing and DNA fingerprinting to establish proof of parentage in immigration cases. 1BYU L Rev 129–149. Available at: http://digitalcommons.law.byu.edu/lawreview/vol1994/iss1/1 (Last visited on January 12, 2018)
12. Virkler K, Lendev 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 188:1–17

13. Tande CM (1989) DNA typing: a new investigatory tool. Duke Law J 1989(2):474–494
14. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73
15. Romelka JM, Yan F (2013) Recent advances in forensic DNA analysis. J Forensic Res S12:001
16. Williamson AL (2012) Touch DNA: forensic collection and application to investigations. J Assoc Crime Scene Reconstr 18:1–5
17. Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA. Nature 316:76–79
18. Roewer L (2013) DNA fingerprinting in forensics: past, present, future. Investig Genet 4:22
19. Bhatt JN (2003) A profile of forensic science in juristic journey. SCC (Jour) 8:25
20. Iyengar A, Hadi S (2014) Use of non – human DNA analysis in forensic science: a mini review. Med Sci Law 54:41–50
21. Neufeld PJ, Oclman N (1990) When science takes the witness stand. Sci Am 18, 262
22. Jeffreys AJ, Brookfiled FY, Semeonoff R (1985) Positive identification of an immigration test case using human DNA fingerprint. Nature 317:818
23. Kelly KF, Rankin JJ, Wink RC (1987) Methods and applications of DNA fingerprinting: a guide for the non-scientist. Criminal Law Review 105–108 note, "Stemming the DNA tide; a case for quality control guidelines" Hamline Law Review, 16 (1992):211:213–214
24. Bureau of Justice Statistics (1991) Forensic DNA analysis: issues. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC, p 4 note 8
25. Dodd V, Laville S (2012) Stephen Lawrence verdict: Dobson and Norris guilty of racist murder. The Guardian, January 03, 2012. Available at: http://www.theguardian.com/uk/2012/jan/03/stephen-lawrence-verdict-guilty-murder. (Last visited on January 12, 2018)
26. 533 So.2d 841 (1988)
27. James R (2009) A brief history of DNA Testing. Time 19 June, 2009
28. The District Court Of Columbia introduced "general-acceptance" test in Fyre v. United States [293 F. 1093 (D.C. Cir. 1923]
29. Daubert v. Merrell Dow Pharmaceutical (1993) 125 L.Ed. 125 2d. 469
30. 3 F. 3d 1191 (1993)
31. R v. Lewis (1987) 88 F.L.R. 104
32. R v. Trans (1990) 50 A Crim R. 233
33. R v. Turner (1975) 1 Q.B. 834
34. Lander ES (1989) DNA fingerprinting on trial. Nature 339:501; Scheck BC (1994) DNA and Daubert. Cardozo L Rev. 15:1959
35. 611 N.Y.S.2d 97 (1994), 633 N.E.2d 451
36. Burk DL (1988) DNA Fingerprinting: possibilities and pitfalls of a new technique. Jurimet J 455; Lander ES (1989) DNA Fingerprinting on trial. Nature 339; Kaye, DH (1991) The admissibility of DNA testing. Cardozo L Rev 353; Lampert R (1991) Some caveats concerning DNA as criminal identification evidence with thanks to Reverend Bayes. Cardozo L Rev 303
37. 545 NY S 2d 985 (Sup Ct. 1989)
38. 821 F. 2d 1186 (6th Cir. 1987) para 79
39. 1999 WL 233592 at 13 (Tenn. Crim. App. Apr. 20. 1999)
40. 776 A.2d 1091 (Conn. 2001)
41. 181 DLR (4th) 320 (1999)
42. In CBI v. Santosh Singh AIR 1994 SC 786, the defense alleged of tampering with the DNA sample and the burden to rebut the allegation lies on the prosecution
43. 2015 SCC OnLine SC 1336
44. 2003 Cr.LJ. 4508 (A.P.)
45. AIR 1994 SC 786. In Appeal, the view of court was upheld by the Supreme Court [State of Karnataka v. M. V. Mahesh (2003) 3 SCC 353]
46. (1977) 23 EHRR 313

47. Liff v. Liff (Orse Rigby) 1948 WN 128
48. W. v. W. (1963) 2 All ER 841 (CA) P 67; S. v. McC [1972] A.C. 24
49. State of Uttar Pradesh v. Deomen Upadhyaya AIR 1960 SC 1125
50. AIR 1961 SC 1808: 1962 SCR (3) 10
51. 2010 7 SCC 263
52. Naik AKV v. The State of Andhra Pradesh. 1977 Cri LJ 1797 (Para 20)
53. 1976 Cri LJ 1680
54. Thogorani @ K. Damayanti v. State of Orissa 2004 Cr. LJ 4003 (Ori)
55. Kodi Satish Naidu v. The State of Andhra Pradesh 133 S.Ct. 1958 (2013)
56. Olavarupu Venkataswarlu v. Polavarappu Subbaya AIR 1951 Madras 910
57. Vasu v. Santha 1975 Ker LT 533
58. AIR 1987 SC 1049
59. AIR 1963 Guj 250
60. AIR 1986 MP 57
61. Satpathy DP v, Dixit BP (2000) Cri LJ 1: AIR 1999 SC 3348; Smt. Kaomti Devi v. Poshiram, 2001 (5) SCC 311); Dwarika Prasad Satpathy v. Bidyut Prava Dixit AIR 2001 SC 2266: 2001 Cri LJ 2615; Amarjit Kaur v. Harbhajan Singh 2003 (10) SCC 228; Bharti Raj v. Sumesh Sachdeo AIR 1986 All 259; Sadashiv Mallikarjun Kheradarkar v. Nandini Sadashiv Mallikarjun Khedarkar 1995 Cri LJ 4090 (Bom); K. Selvaraj @ Surendan v. P. Jayakumari 2000 Cri LJ 4748 (Kerala); Syed Mohammad Ghouse v. Noorunnissa Begum (2001) Cri LJ 2028 (Andhra Pradesh); Haribhai Chanabhai Vora v. Keshubhai Haribhai Vora AIR 2005 Guj 157
62. 1993 AIR SC 2295
63. 2005 (4) SCC 449
64. Amitbhai Anilchandra Shah v. Central Bureau of Investigation (2013) 6 SCC 348. Para 52 (j)
65. 2007 CriLJ 964
66. 2007 CriLJ SC 4008
67. (2005) 8 SCC 21
68. State through Reference v. Ram Singh, Crim App. No. 1398/2013, The High Court of Delhi, Decided on March 13, 2014
69. Kamalanantha v. State of Tamil Nadu, Appeal (crl.) 611–612 of 2013 SC decided on April 05, 2005 http://indiankanoon.org/doc/192852/ (last visited on January 12, 2018)
70. Swamy Shraddananda v. State of Karnataka (2008) 13 SCC 767
71. State of Tamil Nadu through CBI/SIT v. Nalini, (1999) 5 SCC 253
72. Beant Singh, The then Chief Minister of Punjab, was assassinated in a car bomb on 31 August 1995. DNA test helped in establishing identity of his defaced dead body
73. Missing Bodies. http://911research.wtc7.net/wtc/evidence/bodies.html (last visited on January 12, 2018)
74. Ge J, Sun H et al. (2014) Future directions of forensic DNA databases. Croat Med J 55:163–166
75. (1987) 1 SCC 624
76. Vasu v. Santha (1975) Kerala Law Times p 533
77. Morris v. Davies (1837) 5 Cl. and Fin. 163
78. Jayprakash v. Nisha 2013 (3) Ker LJ 85
79. II (1991) DMC 499: Kerala High Court; 1991
80. AIR 2012 (Delhi) 151
81. Goswamy GK (2015) The genetic truth of surrogate parentage. Med Leg J 83:188–193
82. Satpathy DP v. Dixit BP AIR 1999 SC 3348
83. (2001) 5 SCC 311
84. (2001) 5 SCC 311 para 10
85. 2000 Cri LJ 1208 (Ker)
86. 2003 (103) Delhi LT 165: AIR 2003 Del 446
87. 2002 (62) DRJ 851

88. (2005) 4 SCC 449
89. AIR 2003 SC 3450
90. (2010) 8 SCC 633
91. Bhawani Prasad Jena v. Orissa State Commission for Women (2005) 4 SCC 449 para 13
92. Kaur A v. Singh H (2003) 10 SCC 228; Kanchan Bedi v. Gurpreet Singh Bedi 2003 Rajdhani Law Reporter 229; Ramkanya Bai v. Bharatram (2009); and Ramkanya Bai v. Bharatram (2010) 1 SCC 85
93. AIR 2012 (Delhi) 151
94. (1997) 1 F.L.R. 360
95. (1972) AC 24
96. (2002) EWCA Civ 383
97. Rohit Shekhar v. Narayan Dutta Tiwari, IA No. 4720/2008 in CS (OS) 700/2008, In Delhi High Court, decided 23 December, 2010 (para 30) Available at: http://indiankanoon.org/doc/101695806/ (last visited on January 12, 2018)
98. (2014) 2 SCC 576
99. (2014) 2 SCC 576 Para 27
100. (2015) 1 SCC 365; 2014 (12) SCALE 126
101. 2014 SCC Online P & H 4982
102. (2014) 2 SCC 576 Para 27
103. Goswami GK (2016) Assisted reproduction and conflicts in rights. Satyam Law International, New Delhi
104. Section 354C of Indian Penal Code, 1860
105. Section 354D of Indian Penal Code, 1860
106. Connors Edward, Lundregan Thomas, Miller Neal and Tom McEwen (1996) "The DNA "Wars" Are over", excerpted from "Convicted by juries, exonerated by science" A case studies in the use of DNA Evidence to establish innocence after trial (National Institute of Justice)
107. http://www.pbs.org/wgbh/pages/frontline/shows/case/failure/ (Last visited on January 12, 2018)
108. Amitbhai Anilchandra Shah v. Central Bureau of Investigation (2013) 6 SCC 348; Rajeev Singh v. State of Bihar 2015 SCC OnLine 1336: (2016) CCR 70 (SC)
109. Setia H (2016) Evidentiary value of forensic reports in Indian courts. Res J Forensic Sci 4:1–7
110. Magalhaes T et al (2015) Biological evidence management for DNA analysis in cases of sexual assault. Sci World J Article Id 365674

Part III
DNA Fingerprinting: Case Studies

Chapter 12
Fundamentals of Autosomal STR Typing for Forensic Applications: Case Studies

Hirak R. Dash, Neha Rawat, Sonia Kakkar, and Arun Kumar Swain

Abstract Each human has 23 pairs of chromosomes, out of which 22 pairs are autosomes and 1 pair is sex chromosome. These 22 pairs of autosomes which determine somatic characters of an individual contribute to the unique genetic makeup of an individual which are inherited from its parents. In this regard, autosomal STRs present at the surrounding of the centromere of an autosomal chromosome are of high use due to their usability in generating a unique fingerprint in an individual. The usefulness of STRs as popular DNA markers relies on the fact that they can be amplified easily by polymerase chain reaction (PCR) and the single copy STR inheritance from each parent. Additionally, the highly variable nature of a number of STR repeats among individuals makes STRs effective for genetic identification of individuals. In this chapter, various autosomal STR markers and their usefulness in the criminal justice system have been discussed in detail.

Keywords Autosomes · STRs · Multiplex PCR · Criminal justice system

12.1 Introduction

Each individual has a unique set of genetic fingerprint. This unique genetic makeup is the hereditary blueprint imparted to the offspring from its parents. This genetic information is stored in the basic molecule of life, i.e. the deoxyribonucleic acid (DNA), which is present inside nucleated cells and is responsible for the phenotype

H. R. Dash (✉)
DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home (Police), Government of Madhya Pradesh, Sagar, Madhya Pradesh, India

N. Rawat
DNA Fingerprinting Unit, Forensic Science Laboratory, Sagar, Madhya Pradesh, India

Banaras Hindu University, Varanasi, Uttar Pradesh, India

S. Kakkar
Department of Forensic Medicine, PGIMER, Chandigarh, India

A. K. Swain
DNA Fingerprinting Unit, Forensic Science Laboratory, Bhubaneswar, Odisha, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_12

as well as the smooth functioning of an individual. Analysis of human DNA sequences reveals that the DNA sequences of individuals vary significantly except for homozygous twins, and a specific pattern of DNA or fingerprint can be considered to be unique for one individual. This individualistic discriminative characteristic of DNA sequences has revolutionized the criminal justice system and has become a useful tool for the identification of individuals as well as the criminal and civil disputed paternity cases [1].

Inside the nucleus of a human cell, DNA is present in the form of a threadlike structure which is tightly coiled many times around histone proteins that are called as chromosomes. A typical human chromosome consists of two sections called arms joined at a constriction point known as centromere (Fig. 12.1). The long and short arms of the chromosome are called 'q' and 'p' arms, respectively, and hence, the location of the centromere is useful to describe the location of specific genes. Chromosomes can be present in either of the three forms, i.e. threadlike chromatin during G1 and interphase of the cell cycle, sister chromatids during S-phase of interphase and condensed visible form during mitosis. In this context, at the beginning of each cell division during S-phase, the DNA is replicated generating two identical copies of DNA which are connected through the centromere. These replicated X-like structures are called as sister chromatid pairs. During mitosis, these sister chromatid pairs condense generating the visible karyotype [14].

A human cell harbours 22 pairs of autosomes and 1 pair of sex chromosomes. The major difference between the autosomes and sex chromosomes is that autosomes determine the somatic characteristics of an individual, whereas the sex chromosomes regulate the sex and related hormonal traits of an individual. Autosomal chromosomes in an individual exist as pairs in the human genome. In this context, both the autosomal chromosome pairs contain the same set of genes arranged in the same order [7]. However, the genome of one chromosome pair differs from the other chromosome pairs. A recent discovery suggests that autosomes can also participate in human sex determination. SOX9 gene found on chromosome no. 17 activates the function of TDF factor which is encoded by Y chromosome which is critical for

Fig. 12.1 (**a**) A typical human cell, nucleus, chromosomes and DNA; (**b**) Characteristic feature of a typical human chromosome

male sex determination. Thus, it is speculated that a mutation in the SOX9 region may affect the development of Y chromosome generating a female individual [12]. Autosomal chromosomes show Mendelian inheritance. In nature, there exist certain genetic disorders pertaining to improper gametogenesis or inheritance pattern, e.g. existence of three copies of chromosome 21 per cell in Down's syndrome [8].

Humans have two sex chromosomes popularly called as allosomes which are responsible for determining the sex of an individual. The two sex chromosomes in humans are called as 'X' and 'Y'. In this context, a female individual contains two homomorphic copies of 'X' chromosome arranged in the same order, whereas in males, the two sex chromosomes are heteromorphic containing a single copy of both 'X' and 'Y' chromosomes [11]. There exist certain common diseases linked to sex chromosomes. Gene defectiveness of X chromosome leads to red/green blindness. Additionally, certain developmental abnormalities in infants are also due to unusual combinations of sex chromosomes such as XXX or XXY [13]. A typical human autosomal STR profile has been shown in Fig. 12.2.

12.2 Autosomal STR Markers and Their Features

Short tandem repeats (STRs) are the short fragments of DNA of two to six base pairs in length found at a particular location of the chromosome. These short nucleotide sequences are repeated over and over, and the number of repeats is found to be unique for one individual [23]. STRs are mostly found surrounding the structural centre of the chromosome, the centromere. The usefulness of STRs as popular DNA markers relies on the fact that they can be amplified easily by polymerase chain reaction (PCR) and the single copy STR inheritance from each parent. Additionally, the highly variable nature of a number of STR repeats among individuals makes STRs effective for genetic identification of individuals [5]. The smaller size of STR sequences has made them a suitable candidate for forensic applications due to the prevalence of degraded DNA in most of the forensic samples. Thus, STR markers with low mutation rates and high power of discrimination are highly useful for human identification to identify the missing persons, perpetrator, victim and others [18]. Currently, whole genome sequencing has yielded millions of STR loci. Thus, an increase in the number of STR loci for identification purposes will effectively improve the discriminatory power to be used for individualization purpose [4].

12.2.1 Nomenclature of STR Markers

After the completion of human genome project in April 2003, the nucleotide sequence and length of all the 23 pairs of human chromosomes are known today. The presence of millions of nucleotides in specific sequence urges a uniform, simple nomenclature for the STR markers to be used for human identification. An STR

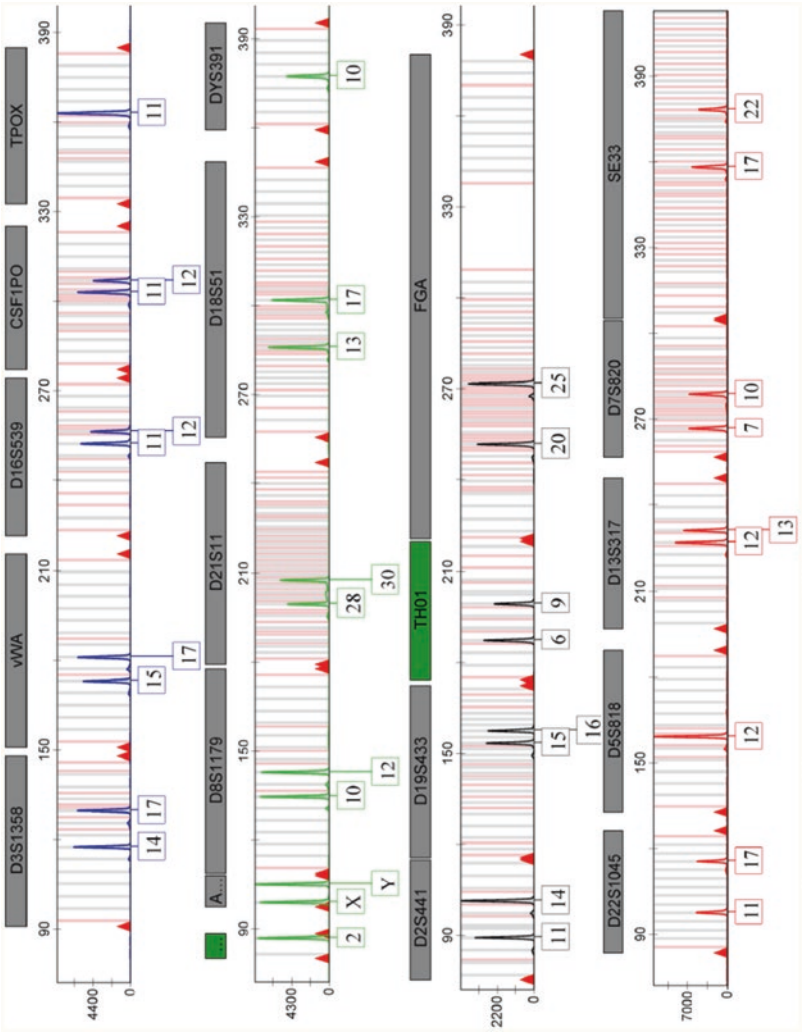

Fig. 12.2 A typical autosomal STR profile from a single-source sample

marker may present either as a part of a gene, and in this case, the gene name is used for STR designation, e.g. STR marker 'TH01' is present within human tyrosine hydroxylase gene found in chromosome 11. '01' in this case describes the presence of the repeat sequences within intron 1 of the gene. Additionally, the STR markers present outside the genes are designated by their positions on the chromosome. For example, the autosomal STR marker designated as 'D5S818' is self-explanatory from its nomenclature. Here, 'D' stands for DNA; the next character '5' refers to the chromosome number; 'S' signifies that the DNA marker is a single copy sequence; and the final designated number signifies the order of discovery and categorization of the marker within that specific chromosome. Thus, STR marker D21S11 can be described as D: DNA, 21: Chromosome 21, S: Single copy sequence and 11: 11th locus described on chromosome 21.

12.2.2 Types of STR Markers

Mostly, the 2–6 bp repeat sequences called as microsatellites or simple sequence repeats or short tandem repeats are used for human identification purposes [3]. However, some minisatellites are also used for this purpose such as 'D1S80', the 16 bp repeat unit with alleles varying within 16–41 repeat units. These microsatellites account for ~3% of the human genome and are scattered within the genome at the rate of every 10,000 nucleotides [20]. Bioinformatics tools can search through the human genome reference sequences to gather information on the available STR markers to be used for human identification purposes. Some commonly used autosomal STR markers have been given in Table 12.1. However, the nature of the STR markers varies as described further.

12.2.2.1 Simple, Compound and Complex

STR sequences vary in terms of their length, the number of repeats as well as the consistency of repeat pattern. Simple repeat STR markers contain units of identical length and sequences, whereas the compound markers consist of two or more adjacent simple repeats. Additionally, the complex STR markers contain several repeat sequences of unit length with variable intervening sequences [21]. In this regard, the desired STR markers should possess certain characteristic features to be included in the analysis in the criminal justice system. An ideal STR marker should possess the characteristic features of highest possible variation, ability to work with degraded DNA samples, less stuttering effect and high discriminatory power to resolve the alleles.

Table 12.1 Commonly used autosomal STR markers and their characteristics used for human identification [2, 6]

STR locus	Chromosome position	Nature	Nucleotide repeats	Repeat units	Allele range
D1S1656	1q42	Compound	Tetra	TAGA	8–20.3
TPOX	2p25.3	Simple	Tetra	AATG	4–16
D2S441	2p14	Compound	Tetra	TCTA/TCAA	8–17
D2S1338	2q35	Compound	Tetra	TGCC/TTCC	10–31
D3S1358	3p21.31	Compound	Tetra	TCTA/TCTG	6–26
FGA	4q31.3	Compound	Tetra	CTTT/TTCC	12.2–51.2
D5S818	5q23.2	Simple	Tetra	AGAT	4–29
CSF1PO	5q33.1	Simple	Tetra	AGCT	5–17
SE33	6q14	Complex	Tetra	AAAG	3–49
D6S1043	6q15	Compound	Tetra	AGAT/AGAC	8–25
D7S820	7q21.11	Simple	Tetra	GATA	5–16
D8S1179	8q24.13	Compound	Tetra	TCTA/TCTG	6–20
D10S1248	10q26.3	Simple	Tetra	GGAA	7–19
TH01	11p15.5	Simple	Tetra	TCAT	3–14
vWA	12p13.31	Compound	Tetra	TCTA/TCTG	10–25
D12S391	12p13.2	Compound	Tetra	AGAT/AGAC	13–27.2
D13S317	13q31.1	Simple	Tetra	TATC	5–17
Penta E	15q26.2	Simple	Penta	AAAGA	5–32
D16S539	16q24.1	Simple	Tetra	GATA	4–17
D18S51	18q21.33	Simple	Tetra	AGAA	5.3–40
D19S433	19q12	Compound	Tetra	AAGG/TAGG	5.2–20
D21S11	21q21.1	Complex	Tetra	TCTA/TCTG	12–43.2
Penta D	21q22.3	Simple	Penta	AAAGA	1.1–19
D22S1045	22q12.3	Simple	Tri	ATT	7–20

12.2.2.2 No. of Sequence Repeats and Microvariants

STR markers are also characterized based on their length of repeat units. They may contain di- (HMS3), tri- (D22S1045), tetra- (D18S51) or penta (PentaE/PentaD) nucleotide repeats. All alleles of an STR locus do not necessarily contain complete repeat units. Certain markers may contain non-consensus alleles that are liable to fall in between the alleles with full repeat units. They are called the microvariants and are common in some STR markers such as THO1 locus, where 9.3 is a common allele. In this case, nine complete tetra-nucleotide repeats are present and additionally one incomplete repeat of three nucleotides [16].

12.2.2.3 Homozygosity and Heterozygosity

Each person inherits one allele from each of his/her parents. When the person inherits the same allele from both of his/her parents, the person is identified to be homozygous at that marker. However, if the person inherits two different alleles from the parents, the person is defined as heterozygous at those markers. For example, if a person inherits eight from the father and eight from the mother at the same STR marker, he/she is said to be homozygous, whereas if the person inherits five from mother and eight from father, he/she is called as heterozygous at that particular marker. A homozygous marker generates a single peak in the electropherogram, whereas a heterozygous marker generates two peaks.

12.3 CODIS vs. Expanded CODIS

The Combined DNA Index System (CODIS) is a database created by the US Federal Bureau of Investigation (FBI). Originally there have been 13 STR loci that were recommended by the National DNA Index System (NDIS) in 1997. However, in order to (a) facilitate greater discrimination, (b) assist in missing person investigations and (c) encourage international data sharing, several new markers have been inducted to the core STR loci [15]. In early 2015, FBI announced the completion of validation project of additional core STR loci and hence added seven new loci (D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 and D22S1045) to the pre-existing 13 CODIS loci effectively from January 1, 2017. Thus the currently recommended new expanded core STR loci are described as follows (Table 12.2).

12.4 Application of Autosomal STR Typing in Criminal Justice System

The application of DNA technology to the biological evidence has revolutionized the criminal justice system to a greater extent. The ability of DNA technology to identify at a higher degree of certainty has led to the routine use of this technology in criminal investigations. Nowadays, this is one of the most sensitive and widely accepted scientific techniques [22]. Criminal investigations in the field of murder, sexual offences, kidnapping, robbery and identification have found DNA examination as a common tool today (Table 12.3). In this context, the primary aim of DNA profiling technique is to differentiate between individuals contributing to the biological materials of a criminal offence.

Table 12.2 Autosomal STR markers recommended in CODIS and the additional markers present in expanded CODIS list

Marker	Nature	Chromosome location	Date of induction
CSF1PO	Simple; tetra	5q33.1	Since 1997
D3S1358	Compound; tetra	3p21.31	Since 1997
D5S818	Simple; tetra	5q23.2	Since 1997
D7S820	Simple; tetra	7q21.11	Since 1997
D8S1179	Compound; tetra	8q24.13	Since 1997
D13S317	Simple; tetra	13q31.1	Since 1997
D16S539	Simple; tetra	16q24.1	Since 1997
D18S51	Simple; tetra	18q21.33	Since 1997
D21S11	Complex; tetra	21q21.1	Since 1997
FGA	Compound; tetra	4q31.3	Since 1997
TH01	Simple; tetra	11p15.5	Since 1997
TPOX	Simple; tetra	2p25.3	Since 1997
vWA	Compound; tetra	12p13.31	Since 1997
D1S1656	Compound; tetra	1q42	January 1, 2017
D2S441	Compound; tetra	2p14	January 1, 2017
D2S1338	Compound; tetra	2q35	January 1, 2017
D10S1248	Simple; tetra	10q26.3	January 1, 2017
D12S391	Compound; tetra	12p13.2	January 1, 2017
D19S433	Compound; tetra	19q12	January 1, 2017
D22S1045	Simple; tri	22q12.3	January 1, 2017

12.4.1 The Case of Identification from Mutilated Body

A mutilated body was found with missing frontal bone in 7.8 × 7.0 cm and intact teeth from a drain (Fig. 12.3). Primarily suspecting murder by a hard object, the dead body was in advanced stage of decomposition, and the investigating agency was clueless about its identification. The autopsy surgeon preserved the femur bone and skull of the deceased individual to be used for DNA-based identification. The DNA typing laboratory received bone and skull samples of the deceased along with the blood samples of the putative mother and father in EDTA vial with their written informed consent. Upon receiving, the DNA was isolated from the bone, skull and blood samples by the organic isolation method followed by the amplification of 15 autosomal STR markers by multiplex polymerase chain reaction. The PCR product was detected by Genetic Analyser 3100 and analysed by GeneMapper v3.2. Half of the alleles of each genetic marker analysed for putative mother's blood sample were found to be present in the profile generated from the bone and skull. Additionally, other obligate alleles were found to be present in the analysed genetic markers of the putative father (Table 12.4). In this process, the identification of the deceased individual was established to be the biological son of the putative father and mother.

Table 12.3 Applications of autosomal STR DNA fingerprinting

Applications	Features
Family connection examination	
Paternity	To determine the biological father of a child, or the biological child of a father
Maternity	To determine the biological mother of a child, or the biological child of a mother
Full sibling	To determine the full biological sibling(s)
Half sibling	To determine the half biological sibling(s) (for either the same mother or the same father)
Grandparentage	To determine the biological grandfather or grandmother of a child or vice versa
Aunt/uncle	To identify the uncle or aunt from either paternal or maternal lineage
Nephew/niece	To identify the nephew or niece from either brother's or sister's lineage
Identity examination	
Identification of alleged child	Generation and storage of DNA fingerprint of a child as a part of **child's safety**
Identification of individual	Generation and storage of DNA fingerprint of family members for individual identification
Analysis of the integrity of family trees	To record and preserve the uprightness of the family tree, generation of family relationships through autosomal STR testing is much stronger and definitive
Forensics examination	
Matching and identification	To interconnect the crime scene, accused and victim in cases of murder, identification, robbery and rape

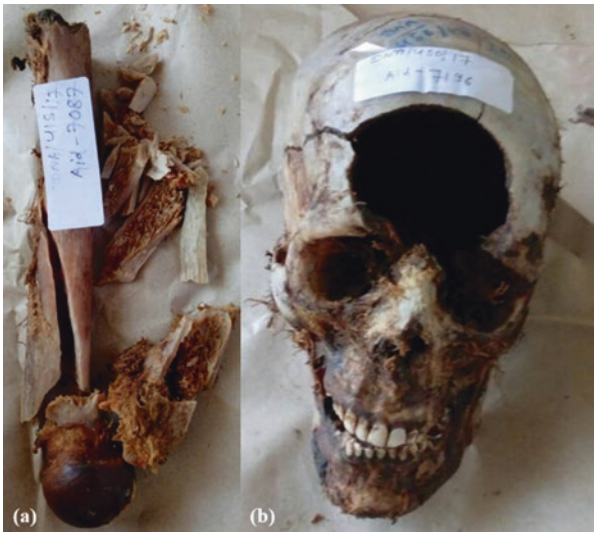

Fig. 12.3 Questioned (**a**) bone and (**b**) skull samples recovered from the mutilated dead body for identification through DNA fingerprinting technique

Table 12.4 Autosomal STR DNA profile obtained from the source of the skull, bone and blood samples of the tentative father and mother

Genetic markers	Blood sample (tentative mother)	Bone/skull (Decd.)	Blood sample (tentative father)
D3S1358	15,17	17,18	15,18
TH01	9,9	9,9	8,9
D21S11	31,32.2	31,31.2	31.2,32.2
D18S51	14,16	16,17	14,17
Penta E	15,18	11,18	11,20
D5S818	11,13	11,13	11,11
D13S317	8,11	11,11	11,11
D7S820	8,10	10,11	10,11
D16S539	11,11	11,11	11,11
CSF1PO	10,11	11,12	12,12
Penta D	9,10	9,10	10,10
vWA	17,18	14,17	14,15
D8S1179	13,15	15,15	14,15
TPOX	8,11	11,11	8,11
FGA	20,22	22,23	23,24
AMELOGENIN	X,X	X,Y	X,Y

12.4.2 The Case of Murder

A case of suspected murder was reported, and the investigating officer recovered blood-stained soil and a knife from the crime scene. During autopsy analysis, the surgeon recovered blood-stained clothes from the deceased individual. Further investigation leads to the recovery of clothes of the accused which also contained traces of blood stains. Laboratory examination of the articles also followed the same line as described previously for DNA isolation, multiplex PCR and genotyping. The obtained result was found to be consistent for all the samples received for examination, i.e. blood-stained soil, knife, clothes of the victim as well as the clothes of the suspect (Table 12.5). Thus DNA examination was fruitful in linking the deceased, suspect, murder weapon and scene of the crime which is the fundamental aspect of forensic science [10].

12.4.3 The Case of Sexual Assault

A minor complainant complains of sexual assault by a known relative of her. Based upon the complaint, the investigating agency started their investigation and arrested the suspect. Both the complainant and the suspect underwent a medical examination, and the medical officer preserved the vaginal swab slide, underwear of the complainant and blood samples of the complainant and suspect for DNA examination with their written informed consent. A mixed autosomal STR DNA profile was

Table 12.5 Autosomal STR profile of the articles received for DNA examination in a suspected murder case

Genetic markers	Blood-stained soil (spot)	Clothes (deceased)	Knife (suspect)	Clothes (suspect)
D3S1358	18,18	18,18	18,18	18,18
TH01	9,9	9,9	9,9	9,9
D21S11	29,29	29,29	29,29	29,29
D18S51	14,16	14,16	14,16	14,16
Penta E	12,21	12,21	12,21	12,21
D5S818	11,12	11,12	11,12	11,12
D13S317	8,10	8,10	8,10	8,10
D7S820	8,9	8,9	8,9	8,9
D16S539	11,12	11,12	11,12	11,12
CSF1PO	10,13	10,13	10,13	10,13
Penta D	11,13	11,13	11,13	11,13
vWA	17,18	17,18	17,18	17,18
D8S1179	15,15	15,15	15,15	15,15
TPOX	9,11	9,11	9,11	9,11
FGA	21,24	21,24	21,24	21,24
AMELOGENIN	X,Y	X,Y	X,Y	X,Y

Table 12.6 Mixed autosomal STR DNA profile detected in a case of suspected sexual assault

Genetic markers	Blood sample (complainant)	Vaginal slide (complainant)	Underwear (complainant)	Blood sample (suspect)
D3S1358	15,16	15,16,18	15,16,18	16,18
TH01	6,9.3	6,8,9.3	6,8,9.3	6,8
D21S11	28,29	28,29,32.2	28,29,32.2	29,32.2
D18S51	13,14	13,14,16	13,14,16	13,16
Penta E	14,20	11,14,18,20	11,14,18,20	11,18
D5S818	11,13	11,12,13	11,12,13	11,12
D13S317	11,12	10,11,12	10,11,12	10,11
D7S820	11,12	11,12	11,12	11,12
D16S539	10,12	9,10,11,12	9,10,11,12	9,11
CSF1PO	12,12	12,12	12,12	12,12
Penta D	11,12	9,10,11,12	9,10,11,12	9,10
vWA	18,18	17,18	17,18	17,17
D8S1179	12,13	10,12,13,14	10,12,13,14	10,14
TPOX	8,11	8,10,11	8,10,11	10,11
FGA	22,24	22,24	22,24	22,24
AMELOGENIN	X,X	X,Y	X,Y	XY

detected from the source of articles' vaginal swab slide and underwear (Table 12.6) confirming more than one contributors to the profile [17]. The autosomal STR DNA profile generated from the blood samples of the complainant and suspect confirmed these two individuals as the contributors to the generated mixed profile from the

complainant's clothing and slide [9]. This confirms the sexual intercourse; thus, transferring the male's body fluid to the female's clothing and body, contributing to the DNA profile of the suspect to the generated mixed profile.

12.5 Future Perspectives

Variation of repeat units among the short tandem repeats (STRs) has become the primary focus of the DNA typing for forensic application since the last few decades. There exist few disadvantages of STR testing as the individual biostatistical efficiency of commonly used STR markers is limited, i.e. the measurement of exclusion power in paternity trio cases [19]. Additionally, the existence of significant mutation rates of the STR markers up to 0.05 can cause huge problem mostly in the case where the mother is not available for matching. The major problem of autosomal STR profiling is the samples from multiple perpetrators. A typical single-source profile mostly generates two peaks (corresponding to the alleles) at a single genetic marker (STR locus) of heterozygous nature. Thus, the mixture analysis becomes more important, and reliable techniques are still not available for common forensic applications. Use of an increased number of markers and balanced use of simple/compound/complex and bi-/tri-/tetra-/pentanucleotide repeat STR markers may lead to the more reliable use of autosomal STR DNA profile. Additionally, a combination of autosomal STR markers, Y chromosome STR markers and SNP markers will be useful for fruitful forensic application of DNA fingerprinting technique.

Ethical Statement All the samples analysed during the case studies were taken from the routine casework performed by author 1 at the DNA Fingerprinting Facility, State Forensic Science Laboratory, Sagar, Madhya Pradesh, India, after due approval from the institutional ethical committee. The reference blood samples have been collected with informed written consent from the donor(s) of the respective cases.

References

1. Butler JM (2015) The future of forensic DNA analysis. Phil Trans R Soc B 370:20140252
2. Butler JM, Hill CR (2012) Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Forensic Sci Rev 24:15–26
3. Chambers GK, MacAvoy ES (2000) Microsatellites: consensus and controversy. Comp Biochem Physiol Part B 126:455–476
4. Cortellini V, Cerri N, Verzeletti A (2011) Genetic variation at 5 new autosomal short tandem repeat markers (D10S1248, D22S1045, D2S441, D1S1656, D12S391) in a population-based sample from Maghreb region. Croat Med J 52:368–371

5. El-Alfy SH, Ahmed F, El-Hafez A (2012) Paternity testing and forensic DNA typing by multiplex STR analysis using ABI PRISM 310 genetic analyzer. J Genet Eng Biotechnol 10:101–112
6. Gettings KB, Aponte RA, Vallone PM, Butler JM (2015) STR allele sequence variation: current knowledge and future issues. Forensic Sci Int Genet 18:118–130
7. Harper JC, Coonen E, Handyside AH, Winston RM, Hopman AH, Delhanty JD (1995) Mosaicism of autosomes and sex chromosomes in morphologically normal, monospermic preimplantation human embryos. Prenat Diagn 15:41–49
8. Hernandez D, Fisher EMC (1995) Down syndrome genetics: unravelling a multifactorial disorder. Hum Mol Genet 5:1411–1416
9. Hu N, Cong B, Li S, Ma C, Fu L, Zhang X (2014) Current developments in forensic interpretation of mixed DNA samples. Biomed Rep 2:309–316
10. Hunter P (2010) Anything you touch may be used against you. EMBO Rep 11:424–427
11. Johnson NA, Lachance J (2012) The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. Ann N Y Acad Sci 1256:E1–E22
12. Koopman P (2001) Sry, Sox9 and mammalian sex determination. EXS 91:25–56
13. Lenroot RK, Lee NR, Giedd JN (2009) Effects of sex chromosome aneuploidies on brain development: evidence from neuroimaging studies. Dev Disabil Res Rev 15:318–327
14. Lundsteen C, Lind AM, Granum E (1976) Visual classification of banded human chromosomes. I. Karyotyping compared with classification of isolated chromosomes. Ann Hum Genet 40:87–97
15. Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, Bright JN, Onorato AJ (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 25:175–181
16. Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–958
17. 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
18. Saad R (2005) Discovery, development, and current applications of DNA identity testing. BUMC Proc 18:130–133
19. Schneider PM (2012) Beyond STRs: the role of diallelic markers in forensic genetics. Transfus Med Hemother 39:176–180
20. Subramanian S, Mishra RK, Singh L (2003) Genome Biol 4:R13. Available online at: http://genomebiology.com/2003/4/2/R13
21. Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences-a survey of twelve microsatellite loci for use as forensic identification markers. Int J Leg Med 107:13–20
22. Walsh SJ, Ribaux O, Buckleton JS, Ross A, Roux C (2009) DNA profiling and criminal justice: a contribution to a changing debate. Aust J Forensic Sci 36:34–43
23. Ziętkiewicz E, Witt M, Daca P, Gala JZ, Goniewicz M, Jarząb B, Witt M (2012) Current genetic methodologies in the identification of disaster victims and in forensic analysis. J Appl Genet 53:41–60

Chapter 13
Y-Chromosomal STR Typing and Case Studies

Jahangir Imam, Ajay Kumar Rana, and Romana Reyaz

Abstract The basis of Y-STR DNA profiling is derived from Y chromosome, which is passed down to all the male child from the father almost intact, from generation to generation without changing (except mutations in Y-chromosome STRs). The beautiful aspect of Y-STRs is its consistent and allele amplification among close paternal male relatives as well as dispersed male relatives. Y chromosome has lower mutability rate than the mtDNA which is always transferred to all the children from the mother, thus making it more stable generation after generation. This chapter elaborately discussed various Y-STR kits developed, its application, and accuracy. We have also highlighted the various reasons and conditions for mutations in Y chromosomes. In this chapter, we tried to establish the application of Y-STRs profiling in forensic science and criminal justice system and its implication. Y-STR-related case studies have also documented to show its use and applicability in forensic sciences.

Keywords Y chromosome · STRs · DNA typing · Forensic science · Haplotype · Case studies

13.1 Introduction

Y-STR markers have long been used in forensics to differentiate between male DNA samples in the cases of rape, paternity, patrilineage testing, and human identification because of their male-specific inheritance and geographical specificity. Y-STRs are short tandem repeats (STRs) found on the male-specific Y chromosome. The short arm of the Y chromosome is very vital as most of the vital genes for male sex determination, spermatogenesis, and other male-related functions are found on it. The polymorphic character of Y chromosome helped to developed Y-STRs which is different among unrelated males but remains unchanged among the paternal lineage generations. There are many virtues which make Y

J. Imam (✉) · A. K. Rana · R. Reyaz
DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home, Jail and Disaster Management, Government of Jharkhand, Ranchi, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_13

chromosome important and unique in the field of forensic sciences, which are mainly its male specificity, effective for small population size, absence of recombination making it explicitly male lineage, low mutation rate, and effective for population specific allele distribution [49]. We know that the main job of forensic science is individualization of person or lineages of descent, and Y-STR is an important player in this in differentiating the majority of unrelated haplotypes [30, 49]. The commonly used Y-STRs are variable in all populations and are also validated for forensic purposes. The largest forensic Y-chromosome haplotype database is available on YHRD (www.yhrd.org), hosted at the Institute of Legal Medicine and Forensic Sciences in Berlin, Germany.

Y-STR genotyping proves to be an excellent means to identify the male DNA which generally gets masked with female DNA particularly in sexual assault crimes where even with differential DNA extraction process, male and female components cannot be separated completely [22, 30]. The disadvantage of this extraction process possesses a great problem to the forensic scientists where the samples are mixed with more than two contributors, which makes the analysis difficult. This also hinders in PCR amplification of minor male DNA mixed with major female DNA component. The currently available commercial Y-STR kits are more robust and able to amplify the minor male DNA in sexual assault cases, even with mixed samples and also with aged and poor-quality semen and samples mixed with other body fluids of victim and suspects [4]. Y-STR genotyping is highly suitable for excluding a male suspect who is falsely implicated in rape cases where autosomal STR profiling is little difficult. Scientists are also in agreement that Y-STR analysis is a precise mechanism to exclude persons as possible contributors to DNA evidence, but it is not useful in exact identification of an individual person. Even the Y-STR match does not carry the power of discrimination and weight into court as an autosomal STR match. Even the Y-STR allelic match does not carry the power of discrimination and weight into court as an autosomal STR match. But Y-STR also has its application in the study of demographic history of human population, anthropology, and population genetics study [65]. The chronological details regarding the Y-STR DNA fingerprinting have been given in Table 13.1.

13.2 Applications of Y-STRs in Forensic Science

The haploid nature, strong linkage, and nonrecombination condition and genetic diversity of the most part of the Y chromosome (nonrecombining part of the Y chromosome, NRY) are very useful in population history study, paternal/kinship determination, familial searching, exclusion of male individual, and importantly in sex identification when the samples are mixed with male and female DNA [13, 33, 34, 38–40, 50, 57]. With the nonrecombining part of the Y chromosome, NRY, it has been inherited through the genealogical tree that unites all human paternal lineages [62]. But for determining male relatedness, diversity is required for Y chromosome. For Y-based genetic genealogy, two types of genetic polymorphisms are mostly

Table 13.1 Timelines and milestones for forensic DNA analysis

Timeline	Major area	Activities and milestones
1992	Inception	1. Discovery of Y-STR DNA fingerprinting
		2. First Y-STR was introduced
		3. Seven core Y-STR loci employed in forensic
1992–2000	Development and exploration	1. Improvement in Y-STR profiling
		2. Development of new Y-STR markers
		3. Growth of Y-STR in forensic
		4. Application of Y-STR in biogeographic ancestry
		5. Widespread use of Y-STR kit
2000–2010	Stabilization and standardization	1. First commercially available Y-STR kit developed with 11 recommended Y-STR loci plus the amelogenin
		2. Named as Y-PLEX 12 (ReliaGene Technologies, New Orleans, Louisiana)
		3. Development of PowerPlex Y kit (Promega Corporation, USA) with 12 Y-STR loci
		4. Development of PowerPlexAmpFISTR® Yfiler ™ kit (Thermo Fisher Scientific, USA) with 17 Y-STR loci
		5. Development of Argus Y-12 QS kit with same 12 Y-STR loci
2011–2017	Sophistication	1. Huge population database formation
		2. Development of cost-effective forensic Y-STR DNA analysis
		4. More better multiplexing system and DNA typing
		5. More research in the field of paternal lineage identification, kinship analysis, and familial searching
		6. Development of cheaper and more robust STR kits
		7. Development of skills and ability to interpret Y-STR evidence results
		8. Higher sensitive methodologies applied to casework and probabilistic software approached to complex evidence
		9. Development PowerPlex® Y23 (Promega Corporation, USA), Yfiler® Plus (Thermo Fisher Scientific, USA), and RM Y-STR (noncommercial)
		10. Use of RM Y-STR

used: single-nucleotide polymorphisms (SNPs), slow-mutating markers, and short tandem repeats (STRs), fast-mutating markers. Y-SNPs can be used to establish the evolutionary lineage and the broad geographic origin of a particular Y chromosome [10]. Similarly human Y chromosome has become a powerful tool in solving forensic cases, especially sexual assault cases for the identification of sex of the male, identification of male lineage, and identification of geographical origin of male lineages [35]. As stated above the most common application of Y-STR is in sexual assault cases, where many times autosomal STR typing is difficult or impossible because of the masking of low male DNA with high female DNA. Moreover, the detection power of Y-STR for male-female mixture DNA is also very high, 1:2000,

which is highly relevant in sexual assault cases [46]. Another important relevance of Y-STR in forensic is the detecting number of unrelated male contributors in mixed DNA samples of two or more unrelated male individuals. Another use of Y-STR is paternity or familial testing where DNA samples of alleged father are not available (deficiency cases), for missing persons and identification of male victims in some disaster [37]. Currently Y-STR haplotype data is also use to get the information about a particular Y haplogroup, and, therefore, Y haplogroups are more informative for biogeographic ancestry inferences. In forensic, this is very helpful in tracing unknown suspects as it can be possible to infer individual's paternal biogeographic ancestry from Y-chromosome analysis [32]. Y-STR can have many advantages but possesses some limitation also: Y-STR haplotype represents a single locus rather than a set of independent markers, and also Y-STR haplotype is much more polymorphic than single STR loci; it requires larger databases for analysis [8, 66]. Thus in case of non-exclusion of paternal relatives of suspect, Y-STR is not useful, and therefore weight of Y-STR evidence is lower than autosomal STRs in forensic investigation. The inability of Y-STR profiling in differentiating males of the same patrilineage is another limitation and thus required autosomal STRs. This limitation arises as a result of low mutation rate of Yfiler STR and sometimes possesses problem in differentiating fathers and brothers and distant paternal relationships. This differentiation can be addressed with the use of panel of Y-STRs with high mutation rates which can increase the power of differentiation. To achieve this aim, a set of rapidly mutating (RM) Y-STRs has recently been described, and multiplex genotyping tools were developed [5, 6]. This RM Y-STR can achieve the high mutation above 10^{-2}, which may increase the differentiating power of father-son four times than for simple Yfiler [4].

13.3 Currently Available Y-STRs for Forensic Use

Initially after the discovery of DNA fingerprinting, autosomal STRs were the markers of choice, and a lot of development occurs in designing the near-to-perfect autosomal STR markers. Y-STR introduction was late in forensic science as autosomal STRs had other application like linkage mapping study other than its use in forensic. Also, Y-STRs are nonrecombining in nature, and its application is limited, but later on as more and more challenging cases appear, scientists started working on the development of Y-STR markers. The very first Y-STR was described in 1992 by Rower and Epplen which was later on employed in examining forensic cases [51]. The important requirement for Y-STR marker is that it must be Y-specific and male-specific. On the basis of the above requirement, a core set of 15 Y-STR markers was developed out of which seven core loci employed in forensic use. This set is called as "minimal haplotype (MH) Y-STR loci" which were able to differentiate 74–90% of unrelated males which later on increased to 91–97% in European population [4, 28]. Earlier many results were provided by using seven to nine core Y-STR loci but are always important to increase the number of markers. Therefore, a panel of 17

Table 13.2 Different commercial Y-STR kits presently available for forensic application

Sl. No.	Kit name	No. of markers	Application and advantages	Make
1	PowerPlex® CS7 System (nonstandard STR marker system)	Seven STR loci	It is used as a confirmatory kit in paternity applications	Promega Corporation, USA
2	PowerPlex® Y	12 Y-STR loci	First sex-chromosome STR kit developed to identify male lineages	Promega Corporation, USA
3	AmpFISTR® Yfiler ™	17 Y-STR loci	Most commonly sex-chromosome STR kit used nowadays to identify male lineages as it works well in most outbred populations	Applied Biosystem (now Thermo Fisher Scientific)
4	Argus Y-12 QS	12 Y-STR loci + internal control	Sex-chromosome STR kit developed to identify male lineages. The internal control system provides helpful information about PCR efficiency and about the presence of inhibitors in tested samples	Qiagen
5	Yfiler® Plus	27 Y-STR loci	It includes 7 core Y-STR loci and 7 rapidly mutating (RM) markers	Applied Biosystem (now Thermo Fisher Scientific)
6	PowerPlex® Y23	12 Y-STR loci of Yfiler kit plus 6 additional new informative loci for male lineage differentiation	It allows Y-STR analysis of both human forensic samples and database samples. It features a fast amplification time and better tolerance to inhibitors of the PCR when compared to previous generations Y-STR multiplexes	Promega Corporation, USA
7	RM Y-STR set	13 RM Y-STR loci	It is additional Y-STR loci. It is particularly useful in detection of distant male relatives from close ones	Noncommercial

Y-STR loci [2] was developed and validated which contain the recommended seven to nine core loci. Most of these loci are tetranucleotide (same as autosomal STR loci) to prevent stutter formation. Y-STR kit development also evolved and improved with time and need (Table 13.2). Various Y-STR kits listed in Table 13.2 are developed, but currently two Y-STR kits, Yfiler® Plus PCR amplification kit (Thermo Fisher Scientific, Waltham, MA, USA) and PowerPlex® Y23 (Promega Corporation), are commonly and routinely used in forensic science. The Yfiler® Plus PCR

amplification kit (Yfiler® Plus, Thermo Fisher Scientific, Waltham, MA, USA) allows the multiplex amplification of 27 Y-STRs, including 7 rapidly mutating markers (RM Y-STRs). The Yfiler® Plus kit showed a high power of discrimination which is useful for criminal investigations, principally due to the inclusion of RM Y-STRs [48]. PowerPlex® Y23 (Promega Corporation) (PP Y23) System is a 5-dye multiplex genotyping kit that analyzes 17 Y-STR loci, commonly available in other Y-STR multiplex kits along with that 6 new Y-STR loci (DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643). The addition of six new markers has made the Y-STR analysis more discriminating [14, 30, 60]. The PowerPlex® Y23 Y-STR kit is highly sensitive and is best for very low quantity of male DNA which is generally marked by higher female DNA proportions. This kit is also effective for challenged forensic samples which are mixed with inhibitors. It is fast and sensitive with more rapid cycling protocol and can be employed for both forensic application and database samples. This new Y-STR panel is validated for forensic use. Another milestone development took place when high mutation rate in two Y-STRs was identified which were later on designated as rapidly mutating Y-STRs (RM Y-STRs). Later on additional 11 RM Y-STRs were developed, and now we have 13 RM Y-STRs which are extremely useful for paternal lineage differentiation and identification between close and distant male relatives [6]. Thus by using RM Y-STRs, individual identification of a male can be achieved and that also with Y chromosome. This is remarkable achievement which is till date only possible with autosomal STRs. But one major drawback of RM Y-STR is that it is not suitable for paternity/familial testing because of higher rate of mutations in RM Y-STR which makes statistical interpretations difficult. Overall, Y-STR haplotyping is very useful in number of cases, like gang rape, mass disaster, and others where many males are involved. It helps both in exclusion of any suspects involved in any such crimes and establishment of paternal lineage of any male by demonstrating haplotype matches.

13.4 Mutation Ability of Y-Chromosome STR Markers

Nonrecombining Y-chromosomal microsatellites (NRY-STRs) are widely used to study and understand the male population histories and for the identification of genealogical relationship within males in criminal justice system [5]. The NRY is highly desirable as mentioned above because it provides the strongest genetic differentiation over geographic distance when compared with other parts of the genome [24, 5, 36]. But many times mutations have been observed with currently available Y-STRs in paternity testing or even with other studies, and the information about these mutations are limited, as these Y-STRs have developed with small population sets and data is only available for a small set of particular loci [63]. Since no accurate mutation rate data is available and also the knowledge about the molecular cause of Y-STR mutations, exact genealogical studies are sometimes difficult to establish the relationship between putatively closely and distantly related males

[40]. Currently available Y-STR markers are capable enough to distinguish male lineages practically between recent timescale male lineage population but fail to differentiate members of the same male lineage which is a major limitation of Y-STR markers developed from NRY region as conclusions cannot be made at individual level as is usually required in forensic investigations [5]. As Y-STR markers have low-to-mid-range mutation rates which is not sufficient for microevolutionary studies because of low diversity provided by these Y-STRs [18, 23]. The study of Ballantyne [5] provided the mutation rate estimates for the largest number of 186 Y-STRs available to date which can be very helpful for future studies. With this study, now a researcher can select a custom set of Y-STRs suitable for particular applications, slowly mutating Y-STRs (SM Y-STRs) for evolutionary studies, medium-mutating Y-STRs (MM Y-STRs) for population history and genealogy study, and rapidly mutating Y-STRs (RM Y-STRs) for microevolutionary, for population history investigation and individualization among close and distant male lineage populations [5].

13.4.1 Molecular Factors Influencing Y-STR Mutability

13.4.1.1 Number of Repeats

There are strong relationships between the total number of repeats and direction of mutations. Longer alleles have more tendency toward the loss of repeats, while shorter alleles gained repeats more frequently [5], which means higher number of repeats increases the mutation rate.

13.4.1.2 Complexity of Repetitive Sequences

The difference between simple and complex repeats is that simple repeats are consisted of only one uninterrupted, homogeneous repeat stretch, while complex repeats are consisted of more than one repeat stretch or a homogeneous repeat stretch interrupted by 1 bp. The complexity of the locus influences the mutation rates. The more complex the locus, the more accumulation of mutations [5].

13.4.1.3 Length of Repetitive Motifs

The length in base pairs of the repetitive motif somehow also affects the Y-STR mutability. It has been found that there is significant decrease in mutation rate as the repeat length increased, but most available STRs are tetranucleotide in nature, and presence of low number of penta or hexa or even dinucleotide repeats as Y-STR locus denotes that average length of repetitive motif is best suited for Y-STR markers.

13.4.1.4 Father's Age

Father's age at the time of son's birth also influences the mutation rates. The more the age of father at the time of son's birth is more likely to accumulate mutations in Y-STR locus. The average father's age without Y-STR mutations observed was 30.55 (±10.73) years [5].

13.5 Y-STR Database Status Around the World

The goal of the database is to expand continuously the number of individuals (N) for each ancestral group and geographical location. The databases allow DNA laboratory all over the world to search and estimate the Y-STR haplotype frequency across populations for forensic and paternity casework. Two important objectives are pursued: (1) the generation of reliable frequency estimates for Y-STR haplotypes and Y-SNP haplotypes to be used in the quantitative assessment of matches in forensic and kinship cases and (2) the characterization of male lineages to draw conclusions about the origins and history of human populations. Mainly two Y-STR databases are available today: one is the US Y-STR Database (http://usystrdatabase.org/) (Fig. 13.1) funded by the US National Institute of Justice and managed by the National Center for Forensic Science (NCFS) and the European Y-STR database Y Chromosome Haplotype Reference Database (YHRD) (http://www.yhrd.org) (Table 13.3) established by Lutz Roewer and Sascha Willuweit

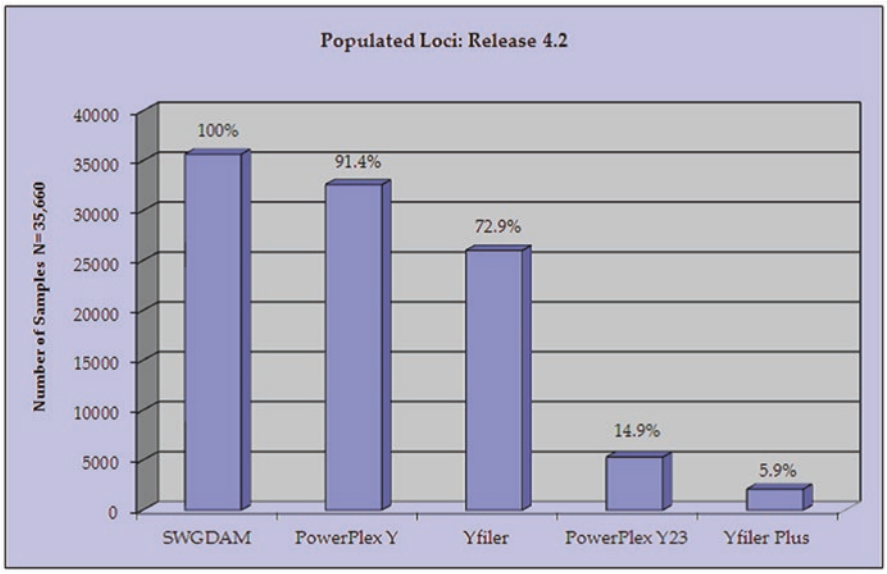

Fig. 13.1 US Y-STR Database last updated February 18, 2017

Table 13.3 Description of loci in YHRD

Dataset	Description	Y-STR loci	No. of markers	No. of haplotypes	No. of population samples	No. of national databases	No. of meta-populations
Minimal	YHRD core loci	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385	9	188,209	1099	132	33
PowerPlex Y	Promega PowerPlex®Y	DYS391 DYS389I DYS439 DYS389II DYS438 DYS437 DYS19 DYS392 DYS393 DYS390 DYS385	12	149,141	850	122	32
Yfiler	Applied Biosystem AmpFISTR® Yfiler ™	DYS456 DYS389I DYS390 DYS389II DYS458 DYS19 DYS385 DYS393 DYS391 DYS439 DYS635 DYS392 YGATAH4 DYS437 DYS438 DYS448	17	136,443	767	115	32
PowerPlex Y23	Promega PowerPlex® Y23	DYS576 DYS389I DYS448 DYS389II DYS19 DYS391 DYS481 DYS549 DYS533 DYS438 DYS437 DYS570 DYS635 DYS390 DYS439 DYS392 DYS643 DYS393 DYS458 DYS385 DYS456 YGATAH4	23	34,133	227	63	28
Yfiler Plus	Applied Biosystem AmpFISTR® Yfiler Plus ™	DYS576 DYS389I DYS635 DYS389II DYS627 DYS460 DYS458 DYS19 YGATAH4 DYS448 DYS391 DYS456 DYS390 DYS438 DYS392 DYS518 DYS570 DYS437 DYS385 DYS449 DYS393 DYS439 DYS481 DYF387S1 DYS533	27	18,921	77	30	26
Maximal	YHRD Max Loci	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392DYS393 DYS385 DYS438 DYS439 DYS437 DYS448 DYS456 DYS458 DYS635 YGATAH4 DYS481 DYS533 DYS549 DYS570 DYS576 DYS643 DYF387S1 DYS449 DYS460 DYS518 DYS627	27	1538	10	7	12

Source: https://yhrd.org/pages/resources/stats#haplotype_counts (Last accessed on 22.09.17)

(Charité-Universitätsmedizin Berlin). Both are proving to be very helpful in population studies and forensic analysis.

13.6 Application of Y-STRs in Criminal Justice System

The male sex chromosome (Y chromosome) is inherited paternally, i.e., from the father to only its male offspring(s) without the event of genetic recombination with any other chromosome during meiosis. Therefore, a paternal DNA sequence and its offspring(s) to a large number of generations (~1000) have identical/similar sequence [9]. This identity has been exploited in forensic science to determine the paternity/lineage propensity of an individual in question to its potential ancestries [31]; however it cannot be used to discriminate an individual from others even selecting several Y-STR markers or rapidly mutating Y-STRs (RM Y-STRs) [7, 55] as the variation is almost null in Y chromosome while passing from one generation to the next. Unlike the recommended 13 core loci (tetrameric short tandem repeats, STRs) by the Federal Bureau of Investigation (FBI), USA, for discrimination of an individual from rest of the people on earth as well as for generation of human identification STR databases [53], the number of Y-STR loci currently being used varies according to the need of study. For example, in forensic casework studies, 17 and 27 Y-STR loci developed by ABI, Life Technologies Inc. and 23 by PowerPlex, Promega Corporation [15, 52, 55] are currently being used, while for biogeographic ancestral lineage study, about 30–100 Y-STR analyses may be required to unequivocally settle on the following disputes in the court of law.

The Y-STR markers have been successfully used in the criminal justice system. Following are the areas or incidences where analysis of Y-STRs has helped to solve an array of cases:

13.6.1 Paternal Lineage

The Y-STR DNA markers which do not change during inheritance and are contributed to male offspring(s) only can be used to trace an individual's father, his father, his father's father, and so forth based on solely male member Y-STR analysis [44]. In case of male siblings whose father may or may not have related surnames, he can be identified with each other based on their Y-chromosome similarity/ identity as well as matching their Y-STR loci with their surviving father's brother/ grandfather or even their surviving son(s) [41].

13.6.2 Resolving Rapists

In case of rape, when semen has been contributed by more than one perpetrator, DNA analysis of vaginal swab using the autologous STR markers usually fails to resolve the number of perpetrators mainly due to a small quantity of male DNA relative to a huge quantity of female DNA as well as difference in quantity of DNA by male contributors and/or similarity in allelic repeats [17, 34]. In such cases, Y-STR analysis of vaginal swab will corroborate the autosomal STRs results by identifying the number of perpetrators and how paternaly they are related [43]. Such useful Y-STR analysis can be performed for the samples obtained from victim's fingernail scrapings, ligature marks on the neck, saliva on the skin, buccal swab, or any other bodily fluid left over victim's body or crime scene.

13.6.3 Meager Male DNA Analysis

For the mixed DNA samples (mixed blood/bodily fluid) where quantity of male DNA is very small in comparison to the large amount of female DNA present (e.g., touch DNA swabbed from female body, vaginal swab, nail scrapings, strangulation swab, etc.), the Y-STR profiling will specifically amplify and analyze the Y-chromosome DNA markers (even from the single sperm cell) [16, 29] with the help of highly specific primers designed against the flanking sequences of the Y-STR markers. Here also the number of perpetrators according to the number of haplotypes produced by Y-STR profiling can be pinpointed and delineated, and the honest justice is given to the victim/their relatives by imposing incisive penalty to the right perpetrators.

13.6.4 Kinship Analysis

The Y-STR test can be used for searching and identification of legitimate male family members (familial searching) among the doubtful pedigree lineages which is failed to be identified by autologous STR markers just after first generation [9, 64]. This kinship examination can be closely related/distantly related by combining Y-STR data with maternally inherited mitochondrial DNA (mtDNA) sequence analysis, viz., single-nucleotide polymorphism, SNPs [12]. For example, in case of dispute of inheritance of property among potential claimants within a territory or outside, the Y-STR plus mtDNA sequence analysis can be used to identify the genuine individual(s), their legitimate father and mother even if they have migrated far away from the original place based on similarity of Y-STR/mtDNA markers with the known related survivors.

13.6.5 Biogeographic Ancestry

Y-STR DNA profiling can be used to determine where their direct paternal ancestors came from, their locations in historic times, and how they have migrated throughout the world [61]. Sharing of Y-chromosome haplotype can be used to resolve the dispute of certain community or religious groups which they belong to (e.g., Ashkenazi and non-Ashkenazi Jews dwelling far away from each other share a common haplotype (J-P58) and identify themselves as Cohanim, i.e., Jewish priesthood members descendent from a single male ancestor) [25]. This test relies on large number of markers (30–100 Y-STR markers) in order to indisputably trace ancestry of the questioned individual/community with great confidence. The Y-STR analysis can be used to determine the immediate descendents of a person/group in question, their ethnic community where they actually belong to, as well as tracing back to their antique biogeographic ancestors who have migrated from eastern Africa, i.e., the cradle of *Homo sapiens* [11, 47]. There are persons who have been lost or have been evacuated from their home states in Great War times (intercountry, World War I and II), destined unknowingly to another region in search of food and forage, or have migrated to an unknown place and can be successfully localized to their ancestral region based on similarity of their Y-STR haplotype alleles with the Y-STR databases available for different biogeographic regions [20, 21, 45, 53, 67]. The human Y-STR study was one of the fundamental analyses along with mtDNA to understand how we have been traced back to have originated from eastern part African regions (the first *Homo sapiens*) [1, 3, 11, 42, 54].

13.7 Pros and Cons of Using Y-STR Markers

Y-STR analysis is extremely valuable tool to trace familial relationships among human males, to help identify misplaced persons, and can establish paternal relationships along a family tree when the alleged father is not available. Y-STR profiling is only male-specific DNA profiling test and is not influenced by the quantity of female DNA or any other autologous DNA present. In case of azoospermic or vasectomized male presence such as in gang rape cases or in a mixture solution of vaginal swab, the presence of such male can be resolved and identified by Y-STR analysis [56, 59]. Therefore, Y-STR analysis can be used to separate male DNA from other male or female DNA whose autologous STR markers coincide or match at certain loci. Thus the Y-STR typing systems provided useful results with autosomal STR in several cases and can be an important tool in forensic caseworks. The Y-STR data has been approved by the FBI, USA, for use by forensic laboratories generating DNA profiles for inclusion to the databases such as the US National DNA Index System (NDIS) and the Combined DNA Index System (CODIS) [19, 21].

Y-STR DNA profiling lacks the power of discrimination (cannot distinguish related or unrelated individuals) due to lack of recombination; however there are

significant variations observed in Y-STR typing in a random population of today's globalized world if the number of Y-STR markers is increased [26, 27]. The Y-chromosomal STR variations among population are accounted by point mutations in Y chromosome that is irreparable (due to lack of recombination repair) and is passed on as such to the next generation. It has been scientifically predicted that between 300 and 100 million years ago, the Y chromosome has evolved and separated from its homologous X chromosome due to paracentric and pericentric inversions, conversion of SOX3 to SRY gene, deletions of Y-chromosome segments, and several non-repairable point mutations till now [9]. The average rate of point mutation per generation per Y-STR locus is about in the order of 10^{-3} and is directly correlated with population diversity in today's demographic picture. There are mutational motifs in Y chromosome called rapidly mutating Y-STRs (RM Y-STRs) [7] which mutate ten times more faster than the rest of Y-chromosomal DNA, i.e., 10^{-2} base per generation (an average of one mutation per hundred generations) [9]. Such mutations may alter the confidence of conviction imposed on the accused/perpetrators.

In spite of such mutational changes occurring in Y-STRs over a century to thousands of generations, Y-STR DNA profiling statistically stands in a good confidence level as a supportive evidence in a court of law for accurately determining the paternity, the lineage, their involvement in gang rape cases, as well as delineating individuals in today's population diasporas to their antique ancestors [34].

13.8 Case Studies

13.8.1 Identification and Exoneration of Suspects in a Gang Rape

Y-chromosome short tandem repeat (Y-STRs) markers are widely used in forensic casework particularly in sexual assault cases that require the detection and identification of a male DNA source and also the exclusion of the so-called suspect. A gang rape case was registered in which five persons were alleged to rape. In this case, autosomal STR profile of male could not be generated from the female garments and vaginal swab. Therefore, Y-STR profiling is done with different Y-STR kits. After analysis, it showed that increasing the number of STR markers increases the power of discrimination. To identify the assailant, Y-STR analysis was performed on forensic samples of victim along with control blood samples of suspects. By using eight Y-STR markers (minimal Y-STR marker), two people out of five were excluded as being possible contributors. By using 16 Y-STR markers (Yfiler), four people out of five were excluded as being possible contributors. By using 22 Y-STR markers (PowerPlex Y-23), four people out of five were excluded as being possible contributors (Table 13.4). By using 27 Y-STR markers (Yfiler Plus), four people out of five were excluded as being possible contributors. The Y-STR profile did not

Table 13.4 PowerPlex Y-23® System allelic data analysis of the exhibits from victim's garments and vaginal swab with five suspects

Genetic markers	Victim's article		Blood samples of suspects				
	Underwear	Vaginal swab	Suspect I	Suspect II	Suspect III	Suspect IV	Suspect V
DYS576	16	16	20	16	16	16	20
DYS389I	13	13	13	13	13	13	13
DYS448	21	21	20	21	21	21	20
DYS389II	29	29	31	29	29	29	29
DYS19	15	15	16	15	15	15	15
DYS391	10	10	10	10	10	10	12
DYS481	22	22	22	22	22	22	23
DYS549	12	12	12	12	13	13	11
DYS533	12	12	12	12	12	12	12
DYS438	9	9	11	9	9	9	11
DYS437	15	15	14	15	15	15	15
DYS570	19	19	19	19	18	16	21
DYS635	21	21	23	21	22	22	23
DYS390	23	23	24	23	23	23	24
DYS439	11	11	10	11	12	11	10
DYS392	11	11	11	11	11	11	11
DYS643	11	11	11	11	11	11	10
DYS393	12	12	13	12	12	12	13
DYS458	16	16	17	16	16	16	16
DYS385a/b	12,16	12,16	11,14	12,16	12,16	12,16	11,14
DYS456	15	15	16	15	15	15	15
Y_GATA_ H4	12	12	13	12	12	12	12

match at four different loci, namely, DYS549, DYS570, DYS635, and DYS439. This case study showed the impact of using advanced Y-STR kits with more number of loci for better and accurate discrimination power. By using PowerPlex Y-23 and Yfiler Plus, we were able to exclude the four suspects out of five which is itself a big achievement as in criminal justice system.

13.8.2 Unusual Mutation Study with Y-STR

An unusual case was encountered where six brothers were accused of gagging and raping a woman. The Y-STR profiles of the accused brothers did not match with the male profile generated from the victim's garment and vaginal swab, but an unusual mutation was observed among the brother's Y-STR profile which was also analyzed further with the father's Y-STR profile. The Y-STR haplotype of all six brothers was found to be the same as that of their father except at locus DYS458 where unusual results were observed [58]. This allelic variation at single locus could have arisen

due to mutation during spermatogenesis. The mutation is interesting in the perspective also that there is gain as well as loss of one repeat at that particular locus in one generation, which is very interesting. As discussed above the father's age at the time of son's birth influences the mutation rates. The more the age of father at the time of son's birth is more likely to accumulate mutations in Y-STRs locus. In this case also the father was 47 years old when he fathered his first son and was 65 years old when his youngest son was born. The present study describes a family with the three mutational events in a single generation at Y-STR DYS458 [58].

Acknowledgment The authors are thankful to the director, State Forensic Science Laboratory, Ranchi, Jharkhand, for support.

References

1. Abu-Amero KK, Hellani A, Gonzalez AM, Larruga JM, Cabrera VM, Underhill PA (2009) Saudi Arabian Y-chromosome diversity and its relationship with nearby regions. BMC Genet 10:59
2. Applied Biosystems (2011)Amp FℓSTR Yfiler PCR Amplification Kit Users Manual
3. Badro DA, Douaihy B, Haber M et al (2013) Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern middle eastern populations with European and African populations. PLoS One 8:e54616
4. Ballantyne KN, Kayser M (2013) Additional Y-STRs in forensic: why, which and when. In: Jaiprakash S and Ray L (ed) Forensic DNA typing (current practices and emerging technologies). https://doi.org/10.12012F015361
5. Ballantyne KN, Goedbloed M, Fang R et al (2010) Mutability of Y-chromosome microsatellites: rates, characteristics, molecular basis, and forensic application. Americ J Hum Genet 82:341–353
6. Ballantyne KN, Keerl V, Wollstein A et al (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6:208–218
7. Ballantyne KN, Ralf A, Aboukhalid R et al (2014) Towards male individualization with rapidly mutating y-chromosomal short tandem repeats. Hum Mutat 35:1021–1032
8. Brenner CH (2010) Fundamental problem of forensic mathematics-the evidential value of a rare haplotype. Forensic Sci Int Genet 4:281–291
9. Burgarella C, Navascues M (2010) Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. Eur J Hum Genet 19:70–75
10. Calafell F, Larmuseau MHD (2016) The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research. Hum Genet 136:559. https://doi.org/10.1007/s00439-016-1740-0
11. Carvalho M, Brito P, Lopes V et al (2011) Analysis of paternal lineages in Brazilian and African populations. Genet Mol Biol 33:422–427
12. Cho S, Seo HJ, Lee J, Yu HJ, Lee SD (2016) Kinship testing based on SNPs using microarray system. Transfus Med Hemother 43:429–432
13. 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(3):e4838
14. Davis C, Ge J, Sprecher C, Chidambaram A, Thompson JM et al (2013) Prototype PowerPlex1 Y23 system: a concordance study. Forensic Sci Int 7:204–208

15. Dogan S, Primorac D, Marjanovic D (2014) Genetic analysis of haplotype data for 23 Y-chromosome short tandem repeat loci in the Turkish population recently settled in Sarajevo, Bosnia and Herzegovina. Croat Med J 55:530–536
16. Feng L, Xu C, Zeng X et al (2014) Y-chromosomal haplotyping of single sperm cells isolated from semen mixtures – a successful identification of three perpetrators in a multi-suspect sexual assault case. Croat Med J 55:537–541
17. Garvin AM, Fischer A, Schnee-Griese J et al (2012) Isolating DNA from sexual assault cases: a comparison of standard methods with a nuclease-based approach. Investig Genet 3:25
18. Ge J, Budowle B, Aranda XG, Planz JV, Eisenberg AJ, Chakraborty R (2009) Mutation rates at Y chromosome short tandem repeats in Texas populations. Forensic Sci Int Genet 3:179–184
19. 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
20. Ge JY, Yan JW, Xie Q, Sun HY, Zhou HG, Li B (2013) Development of Chinese forensic Y-STR DNA database. Fa Yi XueZaZhi 29:212–221
21. Ge J, Sun H, Li H, Liu C, Yan J, Budowle B (2014) Future directions of forensic DNA databases. Croat Med J 55:163–166
22. Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints'. Nature 318:577–579
23. Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, Lao O, Brauer S, Kru¨ger C, Roewer L et al (2009) Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTRYfiler PCR amplification kit. Int J Legal Med 123:471–482
24. Hammer MF, Zegurea SL (2002) The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. Annu Rev Anthropol 31:303–321
25. Hammer MF, Behar DM, Karafet TM et al (2009) Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. Hum Genet 126:707–717
26. Hanson EK, Ballantyne J (2004) A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. J Forensic Sci 49:40–51
27. Hanson EK, Ballantyne J (2007) An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. PLoS One 2:e688
28. Henry J, Simon C, Linacre A (2015) The benefits and limitations of expanded Y-chromosome short tandem repeat (Y-STR) loci. Forensic Sci Int Genet 5:e28–e30
29. Hu N, Cong B, Li S, Ma C, Fu L, Zhang X (2014) Current developments in forensic interpretation of mixed DNA samples (review). Biomed Rep 2:309–316
30. Jain T, Shrivastava P, Bansal DD, Dash HR, Trivedi VB (2016) PowerPlex Y23 system: a fast, sensitive and reliable Y-STR multiplex system for forensic and population genetic purpose. J Mol Biomark Diagon 7:3
31. Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124
32. 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 18:830–838
33. Kayser M (2003) The human Y-chromosome-introduction into genetics and applications. Forensic Sci Rev 15:78–90
34. Kayser M (2017) Forensic use of Y-chromosome DNA: a general overview. Hum Genet 136:621–635
35. Kayser M, Ballantyne KN (2014) Y chromosome in forensic science. In: Primorac D, Shanfield M (eds) Forensic DNA applications: an interdisciplinary perspective. CRC Press Taylor & Francis Group, New York
36. Kayser M, Brauer S, Weiss G, Underhill PA, Roewer L, Schiefenhovel W, Stoneking M (2000) Melanesian origin of Polynesian Y chromosomes. Curr Biol 10:1237–1246

37. Kayser M, Roewer L, Hedman M et al (2000b) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588
38. Kayser M, Krawczak M, Excoffier L et al (2001) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018
39. Kayser M, Lao O, Anslinger K et al (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum Genet 117:428–443
40. Kayser M, Vermeulen M, Knoblauch H, Schuster H, Krawczak M, Roewer L (2007) Relating two deep-rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping. Forensic Sci Int Genet 1:125–128
41. King TE, Jobling MA (2009) What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. Trends Genet 25:351–360
42. Lao O, Vallone PM, Coble MD et al (2010) Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and mitochondrial DNA. Hum Mutat 31:E1875–E1893
43. Magalhaes 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 365674
44. Martinez-Gonzalez LJ, Martinez-Espin E, Alvarez JC et al (2011) Surname and Y chromosome in southern Europe: a case study with Colom/Colombo. Eur J Hum Genet 20:211–216
45. Mrsic G, Grskovic B, Vrdoljak A et al (2012) Croatian national reference Y-STR haplotype database. Mol Biol Rep 39:7727–7741
46. Prinz M, Boll K, Baum HJ, Shaler B (1997) Multiplexing of Y chromosome specific STRs and performance for mixed samples. Forensic Sci Int 85:209–218
47. Purps J, Siegert S, Willuweit S et al (2014) A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet 12:12–23
48. Rapone C, D'Atanasio E, Agostino A, Mariano M, Papaluca MT, Cruciani F, Berti A (2016) Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler® plus kit) in an Italian population sample. Forensic Sci Int Genet 21:e1. https://doi.org/10.1016/j.fsigen.2015.11
49. Roewer L (2003) The use of the Y chromosome in forensic genetics-current practices and future perspectives. Int Congr Ser 1239:279–280
50. Roewer L (2009) Y chromosome STR typing in crime casework. Forensic Sci Med Pathol 5:77–84
51. Roewer L, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89:389–394
52. Rong Y, Gao J, Jiang X, Zheng F (2012) Multiplex PCR for 17 Y-chromosome specific short tandem repeats (STR) to enhance the reliability of fetal sex determination in maternal plasma. Int J Mol Sci 13:5972–5981
53. Ruitberg CM, Reeder DJ, Butler JM (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. Nucleic Acids Res 29:320–322
54. Semino O, Magri C, Benuzzi G et al (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034
55. Serin A, Canan H, Alper B, Sertdemir Y (2011) Haplotype frequencies of 17 Y-chromosomal short tandem repeat loci from the Cukurova region of Turkey. Croat Med J 52:703–708
56. Shewale JG, Sikka SC, Schneida E, Sinha SK (2003) DNA profiling of azoospermic semen samples from vasectomized males by using Y-PLEX 6 amplification kit. J Forensic Sci 48:127–129
57. Shi W, Ayub Q, Vermeulen M et al (2010) A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. Mol Biol Evol 27:385–393
58. Shrivastava P, Trivedi VB, Jain T, Ali M (2015) An unusual occurrence of repeated single allele variation on Y-STR locus DYS458. Egypt J Forensic Sci 6:289. https://doi.org/10.1016/j.ejfs.2015.05.003

59. Soares-Vieira JA, Billerbeck AE, Iwamura ES et al (2007) Y-STRs in forensic medicine: DNA analysis in semen samples of azoospermic individuals. J Forensic Sci 52:664–670
60. Thompson JM, Ewing M, Frank W, Pogemiller J, Nolde C et al (2013) Developmental validation of the PowerPlex Y23 system: a single multiplex YSTR analysis system for casework and database samples. Forensic Sci Int Genet 7(2):240–250
61. Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA 97:7360–7365
62. Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 41:539–564
63. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P et al (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361
64. Vanek D, Saskova L, Koch H (2009) Kinship and Y-chromosome analysis of 7th century human remains: novel DNA extraction and typing procedure for ancient material. Croat Med J 50:286–295
65. Wang CC, Li H (2013) Inferring human history in East Asia from Y chromosome. Investig Genet 4:11
66. Willuweit S, Caliebe A, Andersen MM, Roewer L (2011) Y-STR frequency surveying method: a critical reappraisal. Forensic Sci Int Genet 5:84–90
67. Zastera J, Roewer L, Willuweit S, Sekerka P, Benesova L, Minarik M (2010) Assembly of a large Y-STR haplotype database for the Czech population and investigation of its substructure. Forensic Sci Int Genet 4:e75–e78

Chapter 14
Applications of the Mitochondrion in Forensic DNA Typing

Ranyelle Reid

Abstract Prior to the coronary stent implantation (CSI) era, a typical biology course would introduce the mitochondrion simply as the energy-producing organelle of the cell. Little, if any, discussion was provided about the mitochondrial genome and its participation in human molecular inheritance and evolutionary biology. Now that human identification, via DNA typing, is the driving force behind several forms of television entertainment, the traditional role of the mitochondrion has taken a backseat. In this chapter you will learn how and why mitochondria have been targeted by scientists for use in forensic analysis, human molecular genetics, evolutionary biology, human migration studies, and recovery operations in identifying deceased persons, both ancient and modern. You will also learn how mitochondrial DNA (mtDNA) has provided forensic scientists with a valuable tool for determining the source of DNA recovered from damaged, degraded, or very small biological samples. This chapter explains how mtDNA analysis offers a unique maternal ancestral view of an individual's molecular pin code, through examination of a very specific region of the mitochondrial genome. This chapter will also evaluate both the pros and the cons of mtDNA utility in forensic analysis. Though data have proven increased utility of mtDNA in both historical and modern cases, it is still discounted by many and considered an unreliable forensic tool. An ongoing source of controversy in mtDNA analysis is centered on both data acquisition and data analysis (i.e., how differences in mtDNA sequences are reported). Before the forensic community can approve a DNA typing or classification technique, extensive research on its accuracy, reliability, and discriminatory power must be validated.

Keywords DNA typing · Mitochondria · Sequencing · Forensics · Matrilineal · Ancestor

R. Reid (✉)
BioResearch Molecular Devices, LLC, San Jose, CA, USA

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_14

14.1 Human Inheritance: Nuclear and Mitochondrial DNA

14.1.1 Nuclear DNA

Multicellular organisms develop by a slow process of progressive change which begins with the fusion of two specialized cells, i.e., the egg and sperm. This results in the formation of a zygote, which contains the genetic information carried out from both the contributing parents which regulates the development of the fetus [11]. Zygote divides mitotically and produces all cells of the body. In most of the cases, the nuclear genes are neither lost nor mutated; hence, the genetic nature of each cell is equivalent to that of the other cells except the lymphocytes. In case of lymphocytes, the cells tend to rearrange their DNA through differentiation for the formation of new immunoglobulin and antigen receptor genes [28]. Nuclear DNA is packaged into 23 pairs of chromosomes in normal cells. Each chromosome consists of one very long, linear DNA molecule concomitant with many protein and RNA molecules. A subunit of the DNA molecule which carries the organism's inherited traits is called as gene (Fig. 14.1). The DNA molecule carries hundreds to thousands of genes, the units of information that specify an organism's inherited traits.

Briefly, generalized eukaryotic cells have many different ways to partition off different functions to various locations of the cell. Thus, a typical eukaryotic cell is housed to different specialized compartments called as cell organelles to impart many different designated functions. These organelles play a pivotal role in cellular development as well as function of the organism.

Most of the knowledge regarding the inheritance of human genes have come from the careful observation of human traits over generations and the analysis of family history. Improved analysis of inheritance pattern of specific genes within the families has led to the discovery of major genetic traits within the families. However, the contribution of the great scientist Gregor Mendel cannot be ignored when the knowledge and research on genetics is being talked about. Mendel's early experiments with pea plants have led to the development of two thumb rules of genetics, i.e., equal segregation and independent assortment [43]. According to the first law of Mendel, the two members of a gene pair, the alleles, are separated into gametes, and half of the gametes carry one allele, and another half carries the second allele. Thus, half of

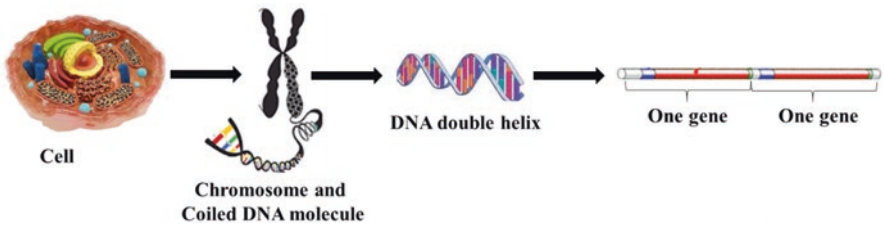

Fig. 14.1 Organization of nuclear DNA and genes in a generalized mammalian nucleated cell

our nucleated genes are inherited from our father and another half from our mother that have been reunited through fertilization in a randomly occurred event. Additionally, according to the second law of Mendel, the independent assortment, one gene pair, is segregated independent to the segregation of the other gene pairs. Thus, the segregation of one allele occurs independent to the other unlinked loci. Additionally, the law of independent assortment is considered to be the fundamental principle which allows the high discrimination power by applying the product rule to the unlinked core loci in forensic typing [17]. Thus, Mendel's law of independent assortment is the source of tremendous genetic variation that allows the exploitation in the field of medical, anthropological, and forensic applications.

Most of the studies so far have exploited the nuclear DNA markers for population genetic studies. The major cause of this is the polymorphicity of these nuclear markers with highest level of discrimination power among individuals except the homozygous twins [20]. In this regard, the population statistics is considered to be essential for statistical evaluation of a match in interpretation of a DNA profile. However, this chapter describes the inheritance of mitochondrial DNA and its suitability in forensic applications.

14.2 The Evolution of the Mitochondrion

Most of the studies on mitochondria deal with the energy-producing ability and metabolism of the cell involving mitochondrion [9]. However, little have been studied and explored regarding the mitochondrial genome and its participation in human genetic inheritance and phenotype determination. Currently human identification via DNA typing has become the driving force for many technological involvements; the traditional role of mitochondrion has taken a backseat. Still, it is ignorant to many people regarding the utility of mitochondrion at genetic level as most of them are aware on the pivotal role of this organelle in the energy house of the cell which is the location of the final phase of aerobic respiration (Fig. 14.2).

Mitochondrial genome (mtgenome) imparts a large forensic advantage due to its stability and high copy number which is the consequence of its function and evolutionary history. In this regard, it is highly essential to understand the necessary functions of the encoded genome and its appreciation to molecular genetics via mitochondrial gene variation [14]. Understanding the necessary functions encoded in mtgenome enhances our appreciation of the molecular genetics of mitochondrial gene variation. All living cells require energy from external sources. Many studies have deciphered the process of energy consumption, its conversion, and transformation by living cells through the process of cellular respiration (Fig. 14.3). Many stages of cellular respiration such as citric acid cycle and oxidative phosphorylation take place inside the matrix of mitochondria. Through the process of glycolysis, glucose is broken down into pyruvate followed by the complete breakdown of glucose by citric acid cycle and ATP synthesis by oxidative phosphorylation [27]. Following the generation of pyruvate during glycolysis, the oxidation of glucose is

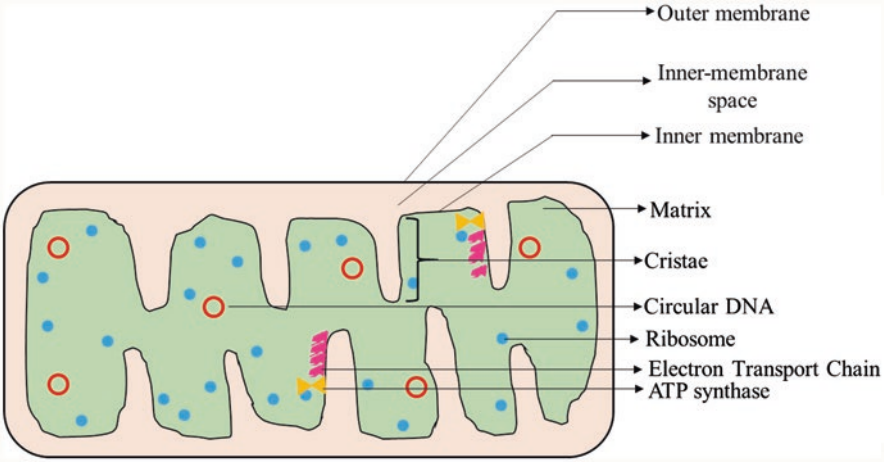

Fig. 14.2 A generalized ultrastructure of a typical mitochondrion

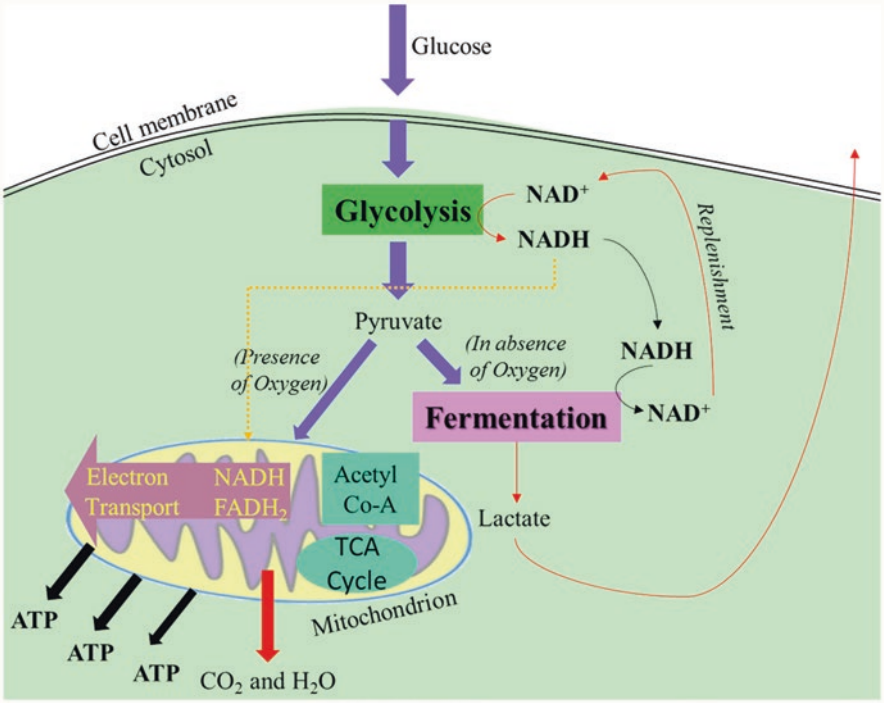

Fig. 14.3 Within mitochondria there are a series of metabolic processes involved in cellular respiration

completed in the mitochondrion in the presence of oxygen (O_2). Enzymes within the outer and inner membranes of the mitochondria assist in converting cellular materials into adenosine triphosphate (ATP), which fuels the metabolic activities of the cell [10].

The separation of biochemical processes such as the citric acid cycle, which occurs in the mitochondrial matrix, and oxidative phosphorylation, which traverses an electrochemical gradient of the tightly folded inner membrane, enables cells to use aerobic respiration to produce approximately 15 times more ATP than anaerobic respiration [10, 24] (Fig. 14.4). Five protein complexes in the mitochondrial inner membrane involved in the electron transport and oxidative phosphorylation pathways have been identified. Complexes I, II, III, and IV are part of the electron transport chain. Complex V is the enzyme complex that carries out oxidative phosphorylation reaction [5].

The dynamic function of a mitochondrion in energy production is reflected from its complex internal compartmental structure. A typical mitochondrion consists of two membranes separated into four distinct compartments, and each membrane-bound compartment can function cohesively for the efficient generation of ATP [3]. The double-membrane structure of a mitochondrion generates a narrow intermembrane space and a large internal matrix. The channel proteins and porins are present in the outer membrane of the mitochondrion which play a major role in selective filtration of cellular components. Additionally, the inner membrane of the mitochondrion harbors structures with extensive folding called as cristae [10]. Major catastrophes during the process of oxidative phosphorylation can lead to severe damage to the mitochondrion and to the organism possessing these organelles. Mitochondria produce reactive oxygen species (ROS) as by-products of inefficient electron transfer across the electron transport chain (ETC) [33]. These superoxide radicals further react to form other ROS, which may lead to mitochondrial trauma.

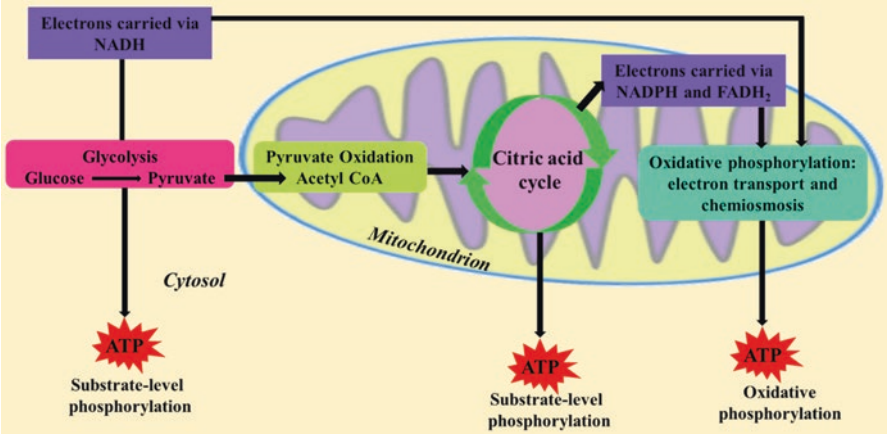

Fig. 14.4 Separation of biochemical processes within the mitochondrion enables cells to produce ATP more efficiently

Additionally, decrease in efficiency of mitochondrial ETC increases oxidative damage directly. Many studies have linked this accumulation of mitochondrial oxidative damage with age and have proposed that increased ROS production may shorten one's life span [44]. Several factors are believed to be responsible for elevated mtDNA mutation rates including inefficient DNA repair mechanisms, omission of DNA protective proteins, and continuous exposure to the mutagenic effects of reactive oxidative species generated by oxidative phosphorylation.

Mitochondria are found in nucleated cells of most eukaryotes which include plants, fungi, animals, and multicellular protists. The number of mitochondria per cell is largely dependent on the cell metabolic requirements. Typically the copy number of mitochondrial DNA (mtDNA) is 100–10,000 copies per cell [22] depending on cell type. Morphologically, mitochondria are diverse and fluctuate in size and shape regularly. They range in shape from long interconnected tubules to small separate spheres [29].

14.3 The Mitochondrial Genome

Evolution and the origin of eukaryotes are important topic in the field of population genetics. Decades of debates and studies have resulted in the rarely unquestioned bacterial ancestry of the mitochondrion. The endosymbiont theory is attributed to the origin and evolution of the mitochondria. The endosymbiont theory states that the mitochondrion evolved from a bacterial progenitor via symbiosis within the eukaryotic host cell. This theory provides an explanation for the mitochondrion having its own genome, separate from that of the nucleus it surrounds. The concept of symbiosis (Latin for *living together*) was first described by a Swiss botanist Simon Schwendener who discovered that lichens consist of a fungus and a photosynthesizer [23]. Endosymbiotic theories attest that cells unite, one inside the other, during evolution to give rise to novel lineages at the highest taxonomic levels via combination. This theory is in direct contrast with Darwin's description of evolution described as gradualism. Theorists still take issue with this explanation of endosymbiosis and rely more heavily on the origin of eukaryotes as the product of gene duplication, point mutation, and micromutational processes [38].

Most of the mitochondrial genome (mtgenome) codes for proteins and enzymes required for their function. The 13 proteins that are encoded by the compact, circular, double-stranded mtgenome are the subunits of the electron transport chain. The mtgenome also encodes two (02) ribosomal RNAs (rRNAs) and twenty-two (22) transfer RNAs (tRNAs) (Fig. 14.5). However, this coding region of mitochondrial genome makes up only about 3% of human genomic DNA.

Mitochondria are semiautonomously functioning organelles that contain an inherent genome that undergoes replication, translation, and transcription of their own DNA [13]. Mammalian mitochondrial DNAs (mtDNA) have two separate origins of replication. The origin of the heavy strand (guanine-rich) is located within a region termed the displacement loop (D-loop), and the light-strand (cytosine-rich)

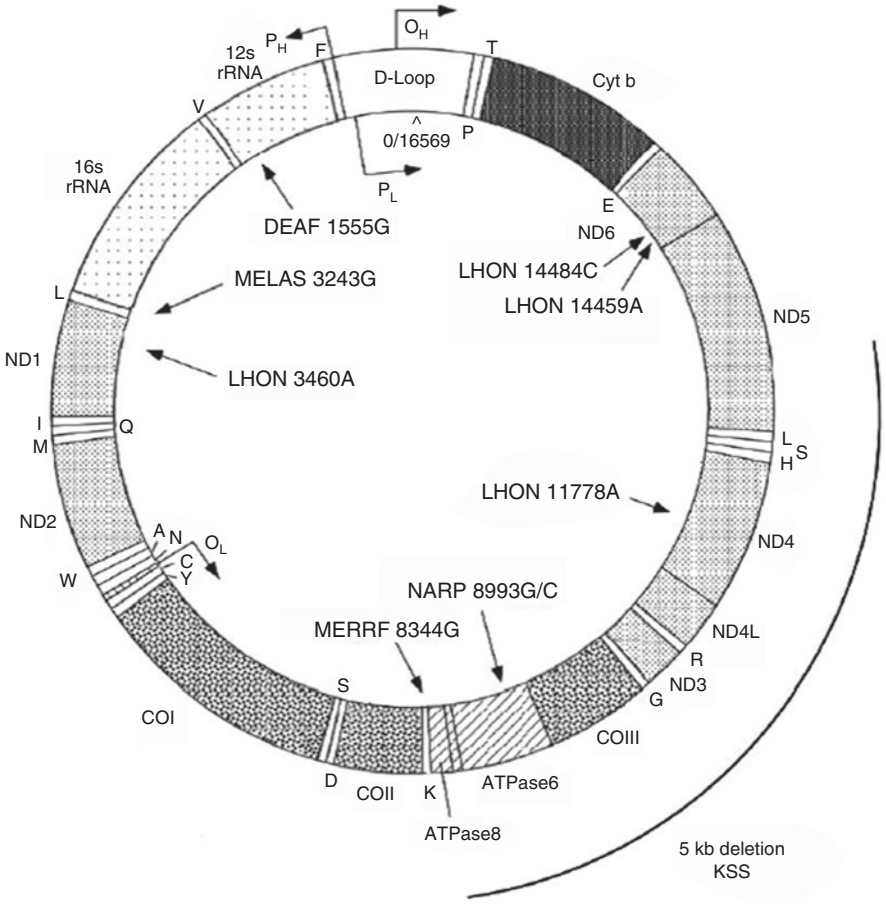

Fig. 14.5 Mitochondrial DNA map

synthesis originates within a cluster of five tRNA genes nearly opposite of the D-loop. The single focus of current forensic typing is the D-loop [30]. The D-loop consists of approximately 1100 base pairs of "noncoding" DNA and is commonly referred to as the hypervariable region due to an increased frequency of mutation as compared to the remaining portion of the mtgenome. The hypervariable region is further divided into three segments. Hypervariable region I (HV1) spans nucleotide positions 16,024–16,365; hypervariable region II (HV2) spans nucleotide positions 73–340; and hypervariable region III (HV3) spans nucleotide positions 438–574. HVI and HV2 are traditionally targeted, whereas HV3 is rarely examined in forensic settings. The hypervariable region has been reported to mutate at a rate of 10 to 17 times more frequently than marked areas of the nuclear genome, namely, single-copy nuclear polymorphic sequences (scnp's) [34].

Mtgenome population-specific variation is widely reported in human evolutionary studies. Many studies have found the usability of variation studies of mitochondrial genome to study the high-frequency haplotype variations among Northwest Africans [35]. Another study reveals the suitability of testing HV1 and HV2 regions in resolving the genetic relationship among Fulani nomads and neighboring sedentary populations in sub-Saharan Africa [7]. Additional studies on HV1 and HV2 regions have found that 70% of the 700 Hispanic individuals residing in United States belong to "Hispanic haplogroups" [1]. Though a typical mitochondrion contains 0.25% of a cell's total DNA, the presence of large numbers of mitochondria in the cytosol of physiologically active cells makes them a suitable source for the DNA to be used in forensic analysis [19].

Current studies suggest the utilization of mtDNA from hair and calcified tissues of forensic samples for utilization in ancient samples, mass disasters, and identification of missing persons with environmentally compromised samples [16]. The odds of mtDNA forensic markers surviving cellular damage are greater than that of the nuclear genome since there are hundreds to thousands of mtgenomes in each nucleated cell. MtDNA is also beneficial in cases with scanty extracted DNA samples [16]. The advantage of mtDNA analysis as an exclusionary tool is often overlooked due to its mode of inheritance. Unlike the nuclear genome, mtgenome does not undergo chromosomal recombination, Mendelian inheritance, or replication repair as only the mother passes clonal copies of her mtgenome to her progeny through the egg (Fig. 14.6). Thus, barring mutation, progeny inherits an identical mtDNA signature that is shared between maternally related individuals. MtDNA is passed through generations independent of male influence as the fertilizing sperm only contributes cellular components directly to the nucleus [39].

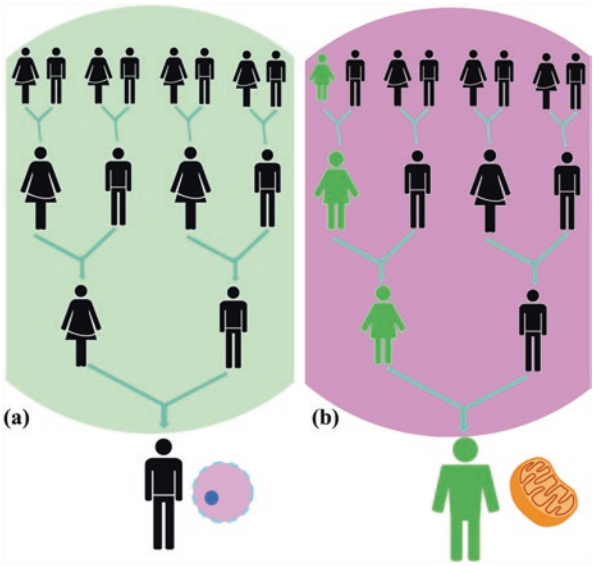

Fig. 14.6 Inheritance of (**a**) nuclear DNA from all ancestors and (**b**) mitochondrial DNA from a single, maternal lineage

14.3.1 Heteroplasmy

Specific genetic markers and procedures used for DNA testing in forensics depend largely upon the quality and quantity of DNA present. Additionally availability of control samples for comparison is also a prerequisite in choosing the methodology. Mitochondrial DNA (mtDNA) has long offered advantages for certain forensic genetic analyses because of its abundance. In comparison to nuclear DNA, each human cell contains hundreds to thousands of copies of mtgenome [26]. Thus, analysis of mtDNA becomes more relevant for ancient and forensic samples with limited sample size and contains highly fragmented and damaged DNA. In this regard, the high copy number of the mtgenome can produce a good profile even with several attempts to generate a profile from nuclear DNA. This is quite evident from the fact of generation of DNA profile from mtDNA in the case of charred remains collected after World Trade Center terrorist attacks in September 2001 in the United States [6]. MtDNA is mostly present in samples that may contain little or no intact nuclear DNA, such as hair shafts and aged fingernails [2]. As a result, mtDNA has been the historical marker of choice for these sample types [25]. MtDNA in forensic casework is of great use in identification cases. Maternal relatives can be used as references for unknown sample, in such cases. This is extremely valuable in a number of scenarios but particularly when the direct references or close relatives required for kinship analyses based on autosomal markers are unavailable.

Forensic comparisons using mtDNA are generally consistent. Evidence interpretation typically involves a direct comparison of the sequences of unknown origin to the sequences of known origin. When the mtDNA sequences of both unknown and known samples are consistent across all nucleotide positions considered for interpretation, the samples are designated as "cannot be excluded as originating from the same source or same lineage" [4]. Nonmatching mtDNA sequences between unknown and known samples are considered "exclusion." Routine mtDNA testing tends to be relatively uncomplicated and easy to interpret. However, there are scenarios that make mtDNA analysis interpretation challenging. A common challenge involves the analysis of heteroplasmy. Heteroplasmy refers to the presence of different mtDNA haplotypes within an individual or tissue [4]. Individuals may possess mtDNA molecules that differ in length, i.e., length heteroplasmy (LHP), or at single nucleotide positions, i.e., point heteroplasmy (PHP). Heteroplasmic variation of either type is not used for exclusionary purposes in forensic mtDNA analysis. This is because of the high mutation rate of mtDNA and the variation that has been observed between different tissue types. Shared PHP between maternal relatives can provide further support for non-exclusion and has proven increased utility of the mitochondrial genome in a plethora of historical cases. Though germ line and somatic mtDNA mutations occur with relative frequency and are often observed in mtDNA profiles as heteroplasmy, the phenomenon of heteroplasmy is believed to be rare by many [32].

Increased frequency of observed heteroplasmy in the general population has been demonstrated by Sanger technology and must be considered in the interpretation of forensic evidence. These Sanger-based studies have formed the basis of our understanding and led to the development of appropriate interpretation guidelines. However, the sensitivity and robustness of newer sequencing technologies have refined our knowledge of heteroplasmy.

Heteroplasmy generally generates mtDNA sequence data in the form of mixture. This distinguishes authentic mtDNA, heteroplasmy, and mixed data from other causes. Thus, the appearance of heteroplasmy adds an added level of complexity to data interpretation regardless of sequencing technology used [15]. Other factors contributing to the problems in analysis of mitochondrial DNA sequences include the mixtures of mtDNA from distinct individuals, contamination by nuclear mitochondrial pseudogene sequences, and chemistry-based sequencing errors [12].

14.3.2 Haplogroup Analysis

MtDNA hypervariable region-based haplogroups have recently garnered more interest in medical genetics, anthropology, and population genetics irrespective of the unique rationale of each field. However, forensic biology is yet to introduce the use of haplogroup classifications as an exclusionary tool for lead in investigation of cold cases. Though few studies have been carried out in the field of mitochondrial DNA and ethnicity [18], instances are there with the use of mtDNA haplogroup analysis in solving criminal investigations.

Mourad Topalian, an Armenian nationalist, was convicted in 1996 of conspiracy acts and possession of illegal weapons and explosives. In this case, "ancient" mtDNA analysis linked the suspect to the hair fragments found on an abandoned storage locker in the 1970s leading to the conviction process [42]. Additionally, molecular evolutionary studies involving human mitochondrial genome reveal the root of modern human origin to be Africa [37]. Most of these comprehensive studies have been carried out by the analysis of single nucleotide polymorphism (SNP) in mtDNA determined by RFLP technique followed by sequencing of HV1 region. It is believed that mtDNA is structured geographically which can be classified into many groups of related haplotypes. Formation of this haplotype is based on the migration of the people across the globe which has been accumulated in the form of genetic signature over time [41]. In this regard, a haplotype is a combination of alleles or genes that are located on the same mtgenome and are therefore inherited as a group. A haplogroup is a genetic signature comprised of a group of similar haplotypes that share a common ancestor (Fig. 14.7). As per current report, two major haplogroups (M and N) along with their derivatives are present in non-African region post migration of modern human out of Africa. Macrohaplogroup L is geographically limited to sub-Saharan Africa which has been subdivided into four major haplogroups, i.e., L0, L1, L2, and L3 [36, 41].

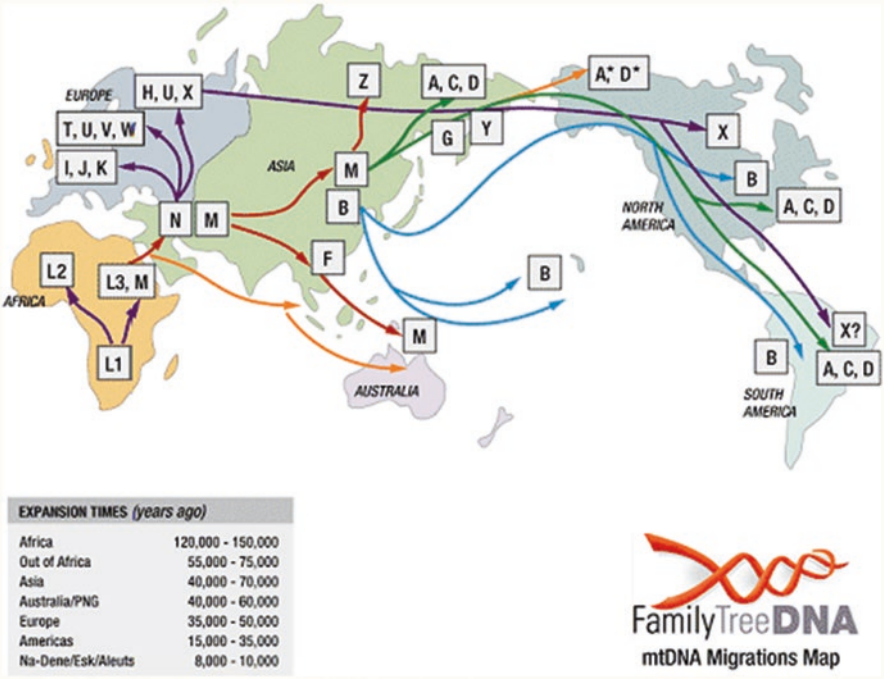

Fig. 14.7 Migrations of human populations and major haplogroups based on the analysis of human mitochondrial DNA. (Original author: https://www.FamilyTreeDNA.com)

14.4 Future Perspectives

There have been tremendous advances in the field of forensic science, and the technology has grown and perfected over centuries. Still there exist many rooms for growth. Forensic science has been criticized for a lack of standards and coordination: complaints about the unreliability of some scientific evidence used in courts are long-standing and also widespread [8]. A 2009 report on the facts of strengthening forensic science in the United States by the US National Research Council called for major reforms to the US forensic science system as well as throughout the globe. It advocates for better standardization of protocols and more research inputs for reliability of techniques used [21]. In this regard, the utility of Innocence Project cannot be ignored and should be conducted in other states as well [40].

Guidelines are published by the DNA Commission of the International Society of Forensic Genetics (ISFG) frequently concerning the application, analysis, and interpretation of mitochondrial DNA (mtDNA) in forensic casework. Though nuclear DNA guidelines for forensic casework have been widely established, the need for similar guidelines for mtDNA analysis is now required for the mtDNA reference population data used to assess the statistical weight of the evidence [31]. In this context, a total of 16 recommendations have been published by ISFG DNA Commission (Table 14.1).

Table 14.1 Recommendations for the use of mtDNA in forensic casework

Sl. No.	Guidelines
1	Good laboratory practice and standard operating protocols for mtDNA analysis must be followed
2	Controls (positive, negative, and extraction blanks) must be carried through the entire process
3	Minimum of 1 reverse sequence and 1 forward sequence (total of 2X coverage) for all reported consensus sequences
4	Avoid manual transcription of data, and independent confirmation of reported data by two scientists must be performed
5	Regular participation in proficiency testing programs for all laboratories using mtDNA typing in forensic casework
6	The full mitochondrial DNA control region should be sequenced for all population genetic studies
7	The revised Cambridge Reference Sequence (rCRS, NC001807) must be used for alignment
8	IUPAC conventions shall be used to describe differences to the rCRS and (point heteroplasmic) mixtures
9	The alignment and notation of mtDNA sequences should be performed in agreement with the mitochondrial phylogeny (established patterns of mutations)
10	Laboratories must establish their own interpretation and reporting guidelines for observed length and point heteroplasmy
11	For population database samples, length heteroplasmy in homopolymeric sequence stretches should be interpreted by calling the dominant variant, which can be determined by identifying the position with the highest representation of a non-repetitive peak downstream of the affected stretch
12	mtDNA population data should be subjected data quality control
13	The entire database of available sequences should be searched
14	Laboratories must justify database(s) and statistical approaches used in reporting
15	Laboratories must establish statistical guidelines for use in reporting an mtDNA match between two samples
16	Length variants in homopolymeric stretches should be omitted from calculations of frequency estimates

14.5 Conclusion

MtDNA's high diversity (the likelihood that two persons selected at random will be different) and low heterogeneity (signatures of the same type rarely occur in populations) yield an informative forensic marker. In many cases, at least 99% of the population will be excluded as contributors, and the pool of random individuals who could have contributed the same is less than 1%. Another advantage of mtDNA typing is in the instance of a "no-body homicide." Because mtDNA is maternally inherited, any maternal relative may donate the mtDNA reference sample to compare to suspected crime scene victim tissues. The utility of the mtgenome should not be discounted, as mtDNA typing is a reliable forensic tool.

References

1. Allard MW, Polanskey D, Wilson MR, Monson KL, Budowle B (2006) Evaluation of variation in control region sequences for Hispanic individuals in the SWGDAM mtDNA data set. J Forensic Sci 51(3):566–573
2. Anderson TD, Ross JP, Roby RK, Lee DA, Holland MM (1999) A validation study for the extraction and analysis of DNA from human nail material and its application to forensic casework. J Forensic Sci 44(5):1053–1056
3. Arora KK, Pedersen PL (1988) Functional significance of mitochondrial bound hexokinase in tumor cell metabolism. Evidence for preferential phosphorylation of glucose by intramitochondrially generated ATP. J Biol Chem 263:17422–17428
4. Bär W et al (2000) DNA Commission of the International Society for forensic genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 113(4):193–196
5. Brandt M (2003) Electron transport and oxidative phosphorylation. Rose-Hulman Institute of Technology. pp. 74–81. www.rose-hulman.edu/~brandt/Chem330/.
6. Brenner CH, Weir BS (2003) Issues and strategies in the DNA identification of world trade center victims. Theor Popul Biol 63(3):173–178
7. Černý V, Pereira L, Musilová E, Kujanová M, Vašíková A, Blasi P et al (2011) Genetic structure of pastoral and farmer populations in the African Sahel. Mol Biol Evol 28(9):2491–2500
8. Cressey D (2013) Tusk tracking will tackle illegal trade: forensic testing of seized ivory could track down poachers. Nature 494(7438):411–413
9. Daicho T, Yagi T, Abe Y, Ohara M, Marunouchi T, Takeo S, Tanonaka K (2009) Possible involvement of mitochondrial energy-producing ability in the development of right ventricular failure in monocrotaline-induced pulmonary hypertensive rats. J Pharmacol Sci 111:33–43
10. Montano E (2012) Improvements to the forensic analysis of mitochondrial DNA typing. In: Honors Scholar Theses, vol 228 http://digitalcommons.uconn.edu/srhonors_theses/228
11. Gilbert SF (2006) Developmental biology, 8th edn. Sinauer Associates, Inc, Sutherland, pp 25–41
12. Gürkan H, Özal SA, Esgin H (2012) Results of mitochondrial DNA sequence analysis in patients with clinically diagnosed Leber's hereditary optic neuropathy. Balkan Med J 29:306–309
13. Hameed IH, Jebor MA, Kareem MA (2015) Forensic analysis of mitochondrial DNA hypervariable region HVII (encompassing nucleotide positions 37 to 340) and HVIII (encompassing nucleotide positions 438-574) and evaluation of the importance of these variable positions for forensic genetic purposes. Afr J Biotechnol 14(5):365–374
14. Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713
15. 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
16. Kavlick MF, Lawrence HS, Merritt RT, Fisher C, Isenberg A, Robertson JM, Budowle B (2011) Quantification of human mitochondrial DNA using synthesized DNA standards. J Forensic Sci 56(6):1457–1463
17. Keim P, Pearson T, Okinaka R (2008) Microbial forensics: DNA fingerprinting of *Bacillus anthracis* (anthrax). Anal Chem 80:4791–4799
18. Kim W, Yoo TK, Shin DJ, Rho HW, Jin HJ, Kim ET, Bae YS (2008) Mitochondrial DNA haplogroup analysis reveals no association between the common genetic lineages and prostate cancer in the Korean population. PLoS One 3:e2211
19. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G et al (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 30(1):69–78
20. Kremer K, Arnold C, Cataldi A, Gutiérrez MC, Haas WH, Panaiotov S, Skuce RA, Supply P, van der Zanden AGM, van Soolingen D (2005) Discriminatory power and reproducibility of

novel DNA typing methods for *Mycobacterium tuberculosis* complex strains. J Clin Microbiol 43:5628–5638
21. Laurin JE (2012) Remapping the path forward: toward a systemic view of forensic science reform and oversight. Tex L Rev 91:1051
22. Malyarchuk BA, Rogozin IB (2004) Mutagenesis by transient misalignment in the human mitochondrial DNA control region. Ann Hum Genet 68(4):324–339
23. Martin WF, Garg S, Zimorski V (2015) Endosymbiotic theories for eukaryote origin. Phil Trans R Soc B 370:20140330. https://doi.org/10.1098/rstb.2014.0330
24. McBride H, Neuspiel M, Wasiak S (2006) Mitochondria: more than just a powerhouse. Curr Biol 16(15):551–560. https://doi.org/10.1016/j.cub.2006.06.054
25. Melton T, Dimick G, Higgins B, Lindstrom L, Nelson K (2005) Forensic mitochondrial DNA analysis of 691 casework hairs. J Forensic Sci 50(1):JFS2004230–JFS2004238
26. Miller FJ, Rosenfeldt FL, Zhang C, Linnane AW, Nagley P (2003) Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. Nucleic Acids Res 31:e61
27. Montemurro C, Vadrevu S, Gurlo T, Butler AE, Vongbunyong KE, Petcherski A, Shirihai OS, Satin LS, Braas D, Butler PC, Tudzarova S (2017) Cell cycle-related metabolism and mitochondrial dynamics in a replication-competent pancreatic beta-cell line. Cell Cycle 18:1–14
28. Nemazee D (2000) Receptor selection in B and T lymphocytes. Annu Rev Immunol 18:19. https://doi.org/10.1146/annurev.immunol.18.1.19
29. Okamoto K, Shaw J (2005) Mitochondrial morphology and dynamics in yeast and multicellular eukaryotes. Annu Rev 39:503–506
30. Pai CY, Chou SL, Tang TK, Wei YH, Yang CH (1997) Haplotyping of mitochondrial DNA in the D-loop region by PCR: forensic application. J Formosan Assoc 96:73–82
31. 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 13:134–142 File DOI Link ISSN 1872-4973
32. Payne BAI et al (2013) Universal heteroplasmy of human mitochondrial DNA. Hum Mol Genet 22(2):384–390
33. Quinlan CL, Perevoshchikova IV, Hey-Mogensen M, Orr AL, Brand MD (2013) Sites of reactive oxygen species generation by mitochondria oxidizing different substrates. Redox Biol 1:304–312
34. Quintáns B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic Sci Int 140(2):251–257
35. Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet 62(6):531–550
36. Reid RS (2012) Evaluation of Mitochondrial DNA Typing in a Forensically Relevant Population. Retrieved May 19, 2017, from http://digitalcommons.uconn.edu/dissertations/18/
37. Relethford JH (2001) Ancient DNA and the origin of modern humans. Proc Natl Acad Sci U S A 98:390–391
38. Schlacht A, Herman EK, Klute MJ, Field MC, Dacks JB (2014) Missing pieces of an ancient puzzle: evolution of the eukaryotic membrane-trafficking system. Cold Spring Harb Perspect Biol 6:a016048. https://doi.org/10.1101/cshperspect.a016048
39. Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. N Engl J Med 347(8):576–580
40. Soni N (2015) New science, old convictions - Texas senate bill 344: identifying further necessary reform in forensic science. J Law Biosci 2:149–157
41. Stewart JB, Chinnery PF (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat Rev Genet 16:530–542

42. Walton R (2006) Cold case squads. Pract Asp Crim Forensic Invest Cold Case Homicides. https://doi.org/10.1201/9781420003949.ch11
43. Westerlund JF, Fairbanks DJ (2010) Gregor Mendel's classic paper and the nature of science in genetics courses. Hereditas 147:293–303
44. Ziegler DV, Wiley CD, Velarde MC (2014) Mitochondrial effectors of cellular senescence: beyond the free radical theory of aging. Anat Soc 14(1)

Part IV
Future of DNA Fingerprinting

Chapter 15
Future of DNA Fingerprinting: Application of NGS in Forensic Science

Jahangir Imam, Pankaj Shrivastava, Shivani Dixit, and Amita Shrivastava

Abstract In the relatively short time frame since 2005, NGS has fundamentally altered genomics research and allowed investigators to conduct experiments that were previously not technically feasible or affordable. The various technologies that constitute this new paradigm continue to evolve, and further improvements in technology robustness and process streamlining will pave the path for translation into clinical diagnostics. NGS is no doubt one of most important and noteworthy technological advances in the biological sciences in the last two decades. NGS has also made its mark in the application in forensic sciences. It has overcome the limitations of capillary electrophoresis and also have the potential to provide multi-information like sequence variation detections, differentiating monozygotic twins, STR typing of degraded samples, etc. The best part of NGS is that we can parallel do the typing of CODIS STRs loci and sequencing study to detect the allelic variations simultaneously. Currently many NGS kits are being developed and available which have huge application in forensic field. This chapter reviews the discovery, advancement, applications, and development of new NGS-based forensic kits and highlighted the applications of NGS in the field of forensic science and criminal justice system.

Keywords NGS · CODIS · Genomics · Forensic science · Multi-informative · Monozygotic twins

J. Imam
DNA Fingerprinting Unit, State Forensic Science Laboratory, Department of Home, Jail and Disaster Management, Government of Jharkhand, Ranchi, India

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

A. Shrivastava
Department of P. G. Studies and Research in Biological Sciences, Rani Durgavati University, Jabalpur, Madhya Pradesh, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_15

15.1 Introduction to Forensic DNA Typing: Discovery

Over the past more than three decades, a huge amount of studies has helped to improve our understanding of unique differences among individuals and also to decipher human origin. The analysis of genetic differences among humans has long been used in courts of law to prove identity with certainty. Forensic DNA typing uses DNA analysis and comparison to resolve legal issues, such as paternity tests and inheritance matters, establish identity in criminal cases where biological evidence is found at crime scenes, and identify victims of mass disasters and missing persons from human remains. Alec Jeffreys at the University of Leicester, in the UK, discovered variable and heritable patterns from repetitive DNA using multi-locus probes, and he named the technique "DNA fingerprinting" [33, 34]. Alec Jeffreys was the first to identify common polymorphisms with the help of RFLP method, and this led to the discovery of different formats of DNA polymorphism study; which has the potential to identify a particular person (except for identical twins). This invention led the stone of a new area of science and further developed in due course of time as the most believed technology in criminal investigations. The technology earlier discovered by Jeffreys is obsolete now. It has gone to a variety of changes to make the technology simpler, user-friendly, and above all more convincing. Besides proving its utility in criminal investigation, this has also proved as multifaceted technology and is now being used in many biological disciplines, namely, in diversity, population genetics, conservation studies, and clinical and anthropological studies. Besides this DNA technology has become the technology which has crossed the boundary to academics, and this has become a socially acceptable technology as well. Forensic DNA typing is based on comparison of the nuclear DNA in a person with that identified in biological material found with the DNA of other person or at the scene of occurrence for the purpose of conclusive inclusion or exclusion. The first case which was solved in March 1985, using Jeffreys techniques, was not a forensic case, but it was a case of immigration [33, 34]. This first application of DNA technology not only saved a young boy from deportation but also established the technology in the public eye as the technology which can save the innocent. The technique identified individuals and helped in individualization of biological evidences, and in its first application, it was based on restriction enzyme digestion, followed by Southern blot analysis. Southern blot technique was laborious and requires good-quality DNA for further analysis. The original technology developed by Jeffreys is now obsolete for forensic use and underwent a huge and continual transformation in the basic technology (Fig. 15.1).

15.2 Capillary Electrophoresis-Based DNA Technology: The Present Technology

Capillary electrophoresis (CE) is one of most important advancements as a part of instrumentation in the field of forensic DNA typing. After PCR invention, scientists consider it as the second most needed development. Already presence of DNA

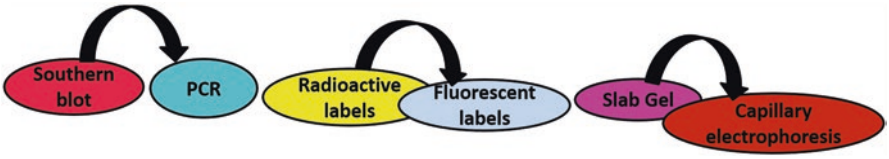

Fig. 15.1 Technological developmental transformation in forensic DNA typing technique since its inception

repeats and its application were known, and amalgam of capillary electrophoresis in forensic applications not only makes the work easier but also more accurate and authentic which is of paramount importance in forensic DNA analysis [59]. The application of capillary electrophoresis is not only limited for biological samples but has huge importance in the analysis of gunshot residues, explosive residues, and drugs. For forensic DNA analysis, STR profiling (highly polymorphic markers) which is based on fragment analysis is of great value for human identification (HID) due to the single-base resolution capability of CE [30, 31, 59]. Introduction of capillary electrophoresis in STR typing circumvents the tedious and expensive approach of DNA sequencing for STR typing. The approach of CE like precise sizing, its sensitivity for the detection of fluorescence emitted by different dyes, automatic electrophoresis, and data collection software are key factors in the worldwide adoption of CE as the preferred platform for forensic DNA analysis. The most common CE systems used in forensic DNA analysis include the ABI PRISM® 310, 3100, 3100 Avant and 3130, 3130xl, 3500, and 3500xL Genetic Analyzers (GAs). The advanced CE automated machines are developed with advanced features which is useful for forensic scientists [35, 36]. It has many advantages like normalization of peak height, accurate sizing of fragments, sample injection, single-base resolution, high run to run precision, good temperature control and automation, better sensitivity, high throughput, user-friendly, and easy software features to analyze the raw data to the level of precise accuracy [59]. Definitely the incorporation of CE in forensic application must be considered as a milestone for the mankind service.

15.3 NGS: The Upcoming Technology in Forensics

15.3.1 What Is NGS Technology?

Next-generation sequencing (NGS) technology which has overcome the limitations of conventional Sanger sequencing technology has grown rapidly in recent years in the field of genomics research because of its high-throughput capacity and low cost and ancient DNA analysis. So far in the advancement in technology, NGS is no doubt one of most important and noteworthy technological advances in the biological sciences in the last two decades. The evolution of sequencing technology from first to third generation is depicted as shown in Fig. 15.2 [45, 55]. NGS is a highly

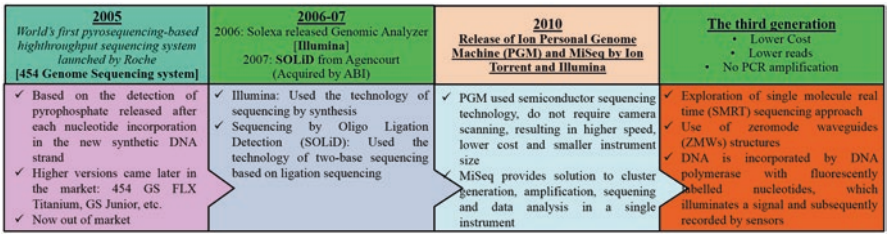

Fig. 15.2 Development in NGS technology: from inception to present

sensitive, high-throughput, low-cost, and much more faster technological advancement which makes it the choice for many biological research along with forensic as compared to other sequencing methods. NGS is now progressing toward maturity and expected to replace traditional sequencing method for DNA profiling which is considered gold standard till date.

15.3.2 Multi-informative Potential of NGS Technology

In recent years, as the technology is advancing, NGS is also not lagging behind with the introduction of STR analysis and parallel sequencing with MPS (massively parallel sequencing). This has proved exceptionally advantageous as the time, cost, speed, and improvement in sequence length and accuracy is dramatically improved [5, 17, 21, 22, 24, 50, 53, 57]. Several NGS platforms have become common in forensic research and commercial applications. The technology promises continued improvement and may become the next "gold standard" for forensic genetics [14]. The Illumina MiSeq FGx Forensic Genomics System became available in early 2015 and has been validated by the manufacturer. NGS analysis not only saves a significant amount of time and but can also provide a number of genetic information in a single run (Fig. 15.3). Furthermore, NGS-based methods provide the full sequence data along with the length-based genotypes [25, 61]. The availability of full sequence information from NGS makes it possible to investigate the mutation and/or the true variation (at the nucleotide level) within STR loci and identify previously unknown alleles and mutational events during paternity establishment [7].

15.3.3 NGS Application in STR Sequence Variation Detection

In today's forensic work, STR analysis has been mainly performed by size-based DNA separation using capillary electrophoresis (CE) [13]. This is mostly serving the purpose of forensic scientists in solving the various types of forensic cases. But as far as internal sequence variation is concerned in STR alleles, capillary

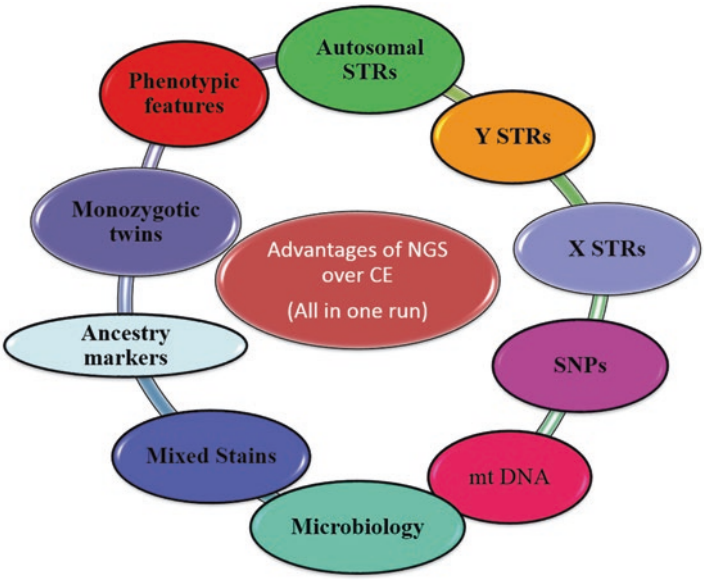

Fig. 15.3 Potential fields of application of next-generation sequencing technology in forensic science

electrophoresis fails which is only based on detection of the length of PCR product. This sequence variation in STR alleles is now considered very important in various cases where very low quantity of DNA is present or for mixed samples which makes the interpretation difficult. Few important techniques like mass spectroscopy and next-generation sequencing (NGS) have been now utilized in forensic also to study and identify the potential internal sequence variations in STR alleles [51–53]. STR sequence variation analysis is nowadays becoming more important as this will improve the discrimination between two individuals and also in mixed DNA samples. This also aids in the study of mutation rates in STR alleles and increases our knowledge about the STR mutability rate for that particular allele. NGS is the solution for the detection and exact identification of variations in STR alleles.

There are a lot of variations in STR repeats which always put the forensic scientists to think for better and correct interpretations. These STR loci are characterized on the basis of their STR repeat category like simple repeat, variant allele, compound repeat, and complex repeat [8]. The CE-based STR assay may identify the alleles according to their relative size compared to an allelic ladder containing sequenced alleles, even though internal sequence variation may be present [25]. Another demerit of CE-based STR analysis is allele peak variations with respect to allelic ladder sizing bean if there are insertions or deletions (indels) in the flanking regions of that allele which is not common but potentially present. The PCR-CE-based assay can be done in 1 day for a specific marker. But the big advantage of NGS is its combined STR analysis and NGS assay even if it takes relatively little longer time [8]. Another important advantage of NGS is its capability to analyze

degraded DNA samples which is difficult with CE-based analysis. Among the various NGS methods available, sequencing by ligations has the lowest error rate, and for STR sequencing which is short-read length, sequencing by ligations is the best platform for forensic genetic applications [8].

Many countries are building their STR profile DNA database of convicted and arrested persons concerned with offences, and this seems that STR profiling still valuable tool in forensic analysis work even though NGS will take over. Therefore, development of NGS with incorporated currently used STRs is the need of time as STRs alone contribute 15% of the human genome [4, 25]. Gettings et al. [25] presented the summary of each STR marker (24 autosomal forensic STR loci) which includes classification of STR marker, repeat unit length, the location of chromosome, repetitive STR region physical location in chromosome, and many more. The details about these markers have been presented in Table 15.1 (After Gettings et al. [25]).

Sequence variation study by NGS provides information about the sequence which can be very useful for the in-depth evaluation of STR alleles. The NGS result must be compared with the PCR-CE-length-based genotype. Three categories of discord were observed in comparing the NGS results to the CE data.

(a) *Presence of isoalleles*: If the allele sequence information is available, the loci which contain numerous isoalleles having internal sequence variation within the repeat region provide better resolution.
(b) *Flanking region indels*: In many STR loci, the 5′ and 3′ flanking regions showed variations which are within the range of PCR product. These are mainly SNPs, for example, at the D13S317 locus, where a four-base deletion in the 3′ flanking region resulted in a "9" allele by CE and a "10" allele by NGS [25].
(c) *Bioinformatic null allele*: This is of two types, Type 1, where the deletion of one base sometimes goes undetected as a result of which homozygous result appears (e.g., at Penta D locus), and Type 2, where the allelic ladder bin did not contain the matching allele (e.g., D12S391 locus, where a "17, 17.1 appears as homozygotes) [25].

Among the 24 autosomal STR loci, 9 loci showed the increase in alleles greater than 30% when sequenced with NGS platform as compared to PCR-CE-length-based genotype [25]. The remaining 15 STR loci showed less variation in the repeat region, for example, loci such as D5S818, D7S820, and D13S317. But less variation is also useful which are mainly in the flanking region which helps in the understanding of mutational events with evolution perspectives. The sequencing by NGS reveals the true variation of STR loci as new alleles have been detected by sequencing of simple STRs which is important as it will improve the statistical power of analysis [25]. If the variability in STR loci will be more, better will be the statistical power of investigation which helps in reducing the number of loci which is required in typing. Another advantage of NGS is to differentiate the STR typing homozygous genotype into heterozygous when the individual loci are sequenced.

Table 15.1 Autosomal STR variation study through NGS

Sl. No.	Classification of STR markers	Repeat unit length	Chromosome location	Physical location of the STR repeat region	Repeat motif sequence and size range of sequenced alleles	Primary source of variation (off ladder)	Presence of high-frequency polymorphisms in flanking region (within 150 bp)
1	D1S1656	Tetranucleotide repeat (TAGA)	Chromosome 1	230,769,616 to 230,769,683	8 to 19.3 repeats	[TAGA]n [TAGG] and [TAGA]n TGA [TAGA]n [TAGG]	Three SNPs with high frequencies close to repeat region from the 5′ end
2	TPOX	Tetranucleotide repeat (AATG)	Chromosome 2	1,489,653 to 1,489,684	4 to 14 repeats	Microvariant x.1, x.2, and x.3	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
3	D2S441	Tetranucleotide repeat (TCTA)	Chromosome 2	68,011,947 to 68,011,994	11 to 13 repeats	[TCTA]n [TNNN] [TCTA]n and TCTA]n TCA [TCTA]n	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
4	D2S1338	Tetranucleotide repeat (TGCC) and (TTCC)	Chromosome 2	218,014,859 to 218,014,950	(TGCC) 4 to 9 repeats and (TTCC) 6 to 19 repeats (total 10 to 26 repeats)	[TGCC]n [TTCC]n [GTCC] $[TTCC]_2$	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
5	D3S1358	Tetranucleotide repeat (TCTA)	Chromosome 3	45,540,739 to 45,540,802	7 to 19 repeats (total 11 to 23 repeats)	[TCTA] [TCTG]n [TCTA]n	No SNPs reported at greater than 5% frequency
6	FGA	Tetranucleotide repeat (CTTT)	Chromosome 4	154,587,736 to 154,587,823	13 to 29 repeats	$[TTTC]_3$ [TTTT] TT [CTTT]n [CTCC] $[TTCC]_2$	No SNPs reported at greater than 5% frequency
7	D5S818	Tetranucleotide repeat (AGAT)	Chromosome 5	123,775,556 to 123,775,599	4 to 20 repeats	[AGAT]n [ACAT] [AGAT]n and [AGAT] n [GAT] [AGAT]n	One SNPs with high frequencies close to repeat region from the 5′
8	CSF1PO	Tetranucleotide repeat (AGAT)	Chromosome 5	150,076,324 to 150,076,375	5 to 16 repeats	[AGGT]	No SNPs reported at greater than 5% frequency

(continued)

Table 15.1 (continued)

Sl. No.	Classification of STR markers	Repeat unit length	Chromosome location	Physical location of the STR repeat region	Repeat motif sequence and size range of sequenced alleles	Primary source of variation (off ladder)	Presence of high-frequency polymorphisms in flanking region (within 150 bp)
9	SE33 (most variable STR loci)	Tetranucleotide repeat (AAAG)	Chromosome 6	88,277,144 to 88,277,245	10 to 25 repeats	[AAAG]n AAAAG [AAAG]n and [AAAG]n AG [AAAG]n	No SNPs reported at greater than 5% frequency
10	D6S1043	Tetranucleotide repeat (AGAT)	Chromosome 6	91,740,225 to 91,740,272	9 to 23 repeats	[AGAT]n [ACAT] [AGAT]n	No SNPs reported at greater than 5% frequency
11	D7S820	Tetranucleotide repeat (GATA)	Chromosome 7	84,160,226 to 84,160,277	6 to 14 repeats	No sequence variation within the repeat region. Microvariant x.1 and x.3	Four SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
12	D8S1179	Tetranucleotide repeat (TCTA)	Chromosome 8	124,894,865 to 124,894,916	7 to 19 repeats	[TCTA]n [TCTG] [TCTA]n	No SNPs reported at greater than 5% frequency
13	D10S1248	Tetranucleotide repeat (GGAA)	Chromosome 10	129,294,244 to 129,294,295	7 to 19 repeats	[GGAA]n [AGAA] [GGAA]n	One SNPs with high frequencies close to repeat region from the 3′
14	THO1	Tetranucleotide repeat (AATG)	Chromosome 11	2,171,088 to 2,171,115	3 to 12 repeats	[TGAA]	No SNPs reported at greater than 5% frequency
15	vWA	Tetranucleotide repeat (TCTA)	Chromosome 12	5,983,977 to 5,984,044	10 to 22 repeats	[TCTA]n [TCTG] [TCTA]n	Four SNPs with high frequencies close to repeat region from the 3′
16	D12S391	Tetranucleotide repeat (AGAT)	Chromosome 12	12,297,020 to 12,297,095	------	[AGAT]n [AGAC]n [AGAT]	No SNPs reported at greater than 5% frequency
17	D13S317	Tetranucleotide repeat (TATC)	Chromosome 13	82,148,025 to 82,148,068	5 to 15 repeats	No sequence variation within the repeat region. Microvariant x.1 and x.3	One SNPs with high frequencies close to repeat region from the 3′

18	Penta E	Pentanucleotide repeat (AAAGA)	Chromosome 15	96,831,015 to 96,831,039	5 to 32 repeats	[AAAGA]n [AAATA] n and [AAAGA]n [AAGAA] [AAAGA] n	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
19	D16S539	Tetranucleotide repeat (GATA)	Chromosome 16	86,352,702 to 86,352,745	4 to 17 repeats	Microvariant x.1, x.2, and x.3	Two SNPs with high frequencies close to repeat region from the 5′ end and 3′ end
20	D18S51	Tetranucleotide repeat (AGAA)	Chromosome 18	63,281,667 to 63,281,738	8 to 40 repeats	Three variants [AGCA], [GGAA], between [AGAA]	No SNPs reported at greater than 5% frequency
21	D19S433	Tetranucleotide repeat (AAGG)	Chromosome 19	29,926,235 to 29,926,298	4 to 18 repeats	[TAGG]	No SNPs reported at greater than 5% frequency
22	D21S11	Tetranucleotide repeat (TCTA) and (TCTG)	Chromosome 21	19,181,973 to 19,182,099	24 to 38 repeats	[TA], [TCA], [TCCATA]	No SNPs reported at greater than 5% frequency
23	Penta D	Pentanucleotide repeat (AAAGA)	Chromosome 21	43,636,205 to 43,636,269	5 to 19 repeats	Flanking region deletions	No SNPs reported at greater than 5% frequency
24	D22S1045	Trinucleotide repeat (ATT)	Chromosome 22	37,140,287 to 37,140,337	8 to 19 repeats	[ATT]n [ACT], $[ATT]_2$	No SNPs reported at greater than 5% frequency

Modified from Gettings et al. [25]

15.3.4 Application of NGS in Differentiating Monozygotic Twins

Monozygotic twins, as the word suggests, arise from a single fertilized egg, and thus they are genetically identical as compared to fraternal twins who are not genetically identical. This puts the block in differentiating the monozygotic twins through autosomal STR profiling. Both individuals have exactly the same DNA sequence; conventional genotyping approaches such as STR, SNP, sex-chromosome STR, and mtDNA analyses cannot differentiate between them. In paternity cases and other cases where one of the monozygotic twins are involved, the other twin cannot be excluded through autosomal STR profiling. This scenario can put the legal justice system in a difficult position to exactly match the biological evidence with one of the identical twins conclusively. Monozygotic (MZ) twins are genetically identical, but they must have some genetic differences which have accumulated in their genome generally called as epigenetic changes [11, 39, 41]. These epigenetic or somatic changes are normal but not exception, and if it occurs during the embryonic development of the MZ twins, it will be more prevalent in the tissues of that particular individual [12]. Somatic mutations or epigenetic changes are random, and surely the MZ twins acquire different mutations [12]. May evidence supports that epigenetic markers can be used to distinguish monozygotic (MZ) twins [40], predict tissue type [23], and accurately determine the age of a DNA donor [6].

Epigenetic approaches based on NGS technology include whole-genome bisulfite sequencing [27], methylation beadchips, reduced representation of bisulfite sequencing [44], and methylated DNA immune precipitation sequencing [64]. All these NGS technologies basically work well with larger DNA fragments, but the challenge with forensic samples is having low and highly degraded DNA samples even mixed with other elements. So a highly accurate NGS technology is required for forensic DNA analysis for epigenetic changes. Genome-wide amplification of a bisulfite-modified DNA template, followed by quantitative methylation detection using pyrosequencing, is one of the best NGS methods for extremely low amounts of DNA [47]. Identification of extremely rare mutations to differentiate between the MZ twins using ultra-deep NGS technology is the landmark work for the solution to paternity and forensic cases where MZ twins are involved [65]. Illumina Human Methylation BeadChip NGS technology is also used to study the methylation pattern (CpG sites) as an epigenetic change for differentiating the MZ twins [40]. The high mutation rate of the mitochondrial DNA (mtDNA) has the potential to become a promising biomarker for the differentiation between MZ twins. With the advancement in various NGS technologies, it is now possible to characterize minor differences of mtDNA genomes (mtGenomes) between MZ twins. In the study conducted by Wang et al. [62], nucleotide differences and heteroplasmies of MZ twins' mtGenomes were mapped for from six pairs of adult MZ twins by NGS technology using the Illumina HiSeq 2000 sequencing system. Their experimental evidence suggests that variants of mtGenomes could be a perspective biomarker to distinguish MZ twins from each other [62].

15.4 NGS in Forensic Casework

NGS is a highly sought technology in the field of biological science, and it has been there since over a decade and now also has become an integral part in the field of forensic sciences [61]. Many attempts have been taken such as Y-chromosome sequencing among related male, metagenomic study of biological stains, mtDNA sequencing for forensic applications, and so on with the help of various NGS platforms [5, 10, 32, 42, 46, 48, 49]. Various studies have been undertaken for the sequencing of markers employed for forensic DNA analysis like STRs, SNPs, microRNA, and mtDNA [9, 16, 19, 20, 22, 24, 26, 28, 37, 38, 43, 58, 62, 63, 66–68]. Many of these studies proved the application of NGS in forensic science as it incorporates multiplexing, sequence variation study of STRs, and high throughput. With the promise which NGS holds as a potent technology in forensic, many commercial kits are released by different companies [16, 22, 28, 66, 67].

15.4.1 STR Typing of Degraded Samples Using NGS

CE-based system is commonly used for STR genotyping in forensic, but frequently it is observed that this system fails to generate the DNA profile data from degraded DNA samples and thus put the forensic scientists in a situation where it become impossible to interpret the results [1–3, 15, 29, 54, 65]. Figure 15.4 depicts the CE-based STR genotyping vs NGS which clearly shows the improved results with NGS. Recently few forensic analyses were done for STR sequencing via NGS in routine casework [18, 21, 53, 56, 57, 60].

15.4.2 Massively Parallel Sequencing of Forensic CODIS 13 Autosomal STRs

The limitations of CE-based STR genotyping are known, and many improvements have also been incorporated in this system, but using NGS for all the 13 CODIS STR markers for human DNA profiling is being tried recently, and it is believed in the forensic scientist community that this can be the parallel method with CE-based STR genotyping. Figure 15.5 depicts the NGS analysis of 18 markers (13 CODIS STRs, D2S1338, D19S433, Penta D, Penta E, and amelogenin) which is based on multiplex PCR system and NGS analysis, and this analysis produced consistent results with CE-based STR genotyping. Many additional applications like, detection of sequence variations at target region, generation of STR profiles with degraded DNA samples and even fomo mixed stains is a limitation with only CE-based DNA Profiling [37].

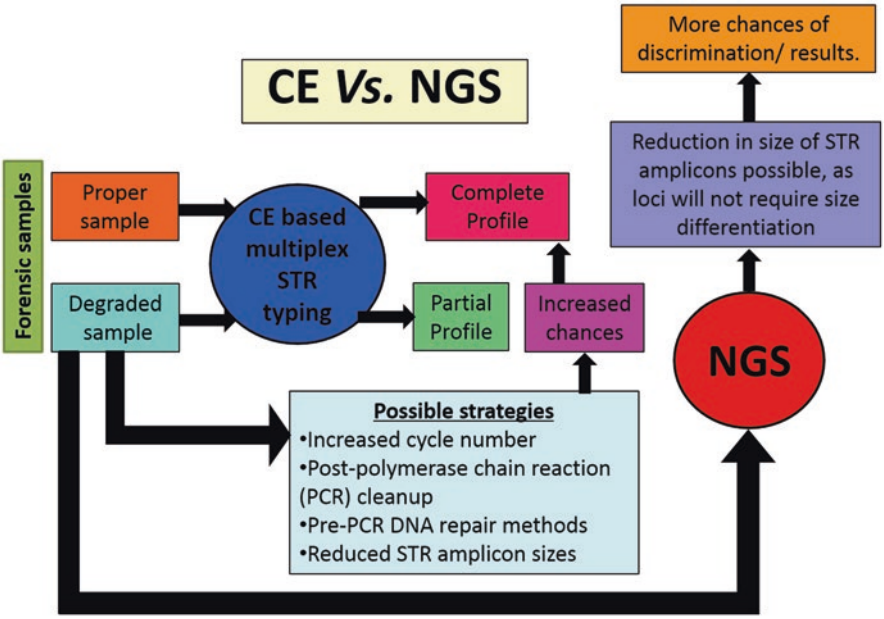

Fig. 15.4 Efforts for improved results in forensic DNA typing CE vs NGS for better results

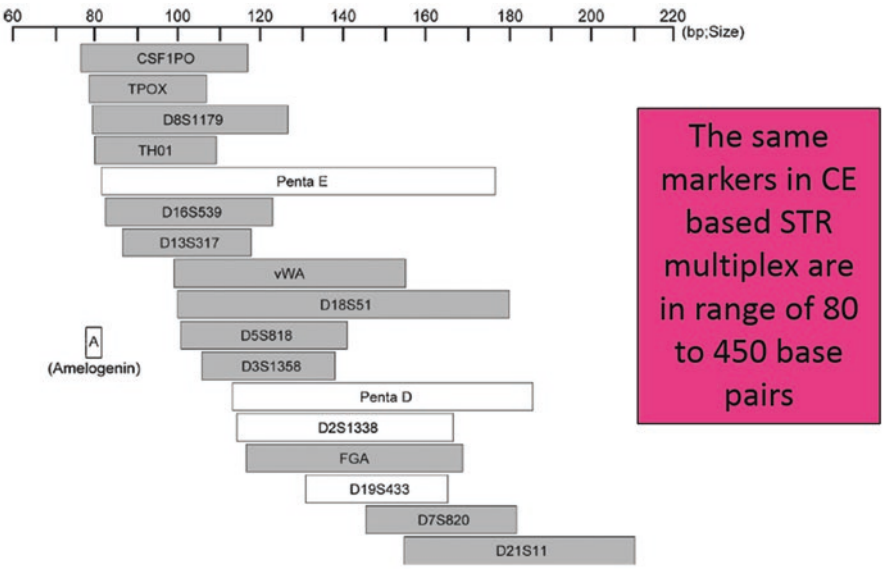

Fig. 15.5 Scheme of 18 forensic markers (13 CODIS shown in gray color) in the developed multiplex PCR system for NGS analysis. (Modified after Kim et al. [37])

Table 15.2 Available NGS-based kits for forensic use

Name of kit	Manufacturer/ supplier	STR loci included
HID-Ion AmpliSeq™ Identity Panel	Thermo Fisher Scientific	124 autosomal SNPs (most of the SNP*for*ID and individual identification SNPs (IISNs)) and 34 Y-chromosome SNPs
HID-Ion AmpliSeq™ Ancestry Panel	Thermo Fisher Scientific	Ancestry informative markers (AIMs)
Ion Torrent™ HID STR 10-Plex panel	Thermo Fisher Scientific	All-in-one solution from amplification of STRs and amelogenin and sequencing to data analysis
Ion Torrent PGM™ platform	Thermo Fisher Scientific	Includes 16 of 20 expanded combined DNA index system (CODIS) core loci and amelogenin
24-plex STR panel	Thermo Fisher Scientific	Cover all recommended combined DNA index system expansion markers (CODIS core loci, ESS markers, DYS391, and amelogenin)
Precision ID GlobalFiler NGS STR panels	Thermo Fisher Scientific	20 autosomal STR CODIS and expanded CODIS loci, 1 Y-chromosome STR locus, 1 autosomal NC02 locus, 3 autosomal low probability of identity (PI) (0.09) STR loci, 5 autosomal next-generation sequencing (NGS) STR loci, 1 indel polymorphic marker on the Y chromosome (Y indel), X and Y amelogenin, the sex determining marker
ForenSeq™ DNA Signature Prep Kit	Illumina	27 autosomal, 24 Y STRs, 7 X STRs and 94 identity, 56 ancestry, and 22 phenotypic SNPs in a single reaction
PowerSeq™ Systems (Auto, Y, and Mito)	Promega	22 autosomal STRs, 23 Y-STRs, and 10 amplicons covering the mitochondrial and amelogenin control region

15.4.3 NGS-Based Other Kits Available for Forensic Purposes

Since the inception and development of the technology, Thermo, Illumina, and Promega have made kits available for forensic purposes (Table 15.2).

15.5 Conclusion

NGS provides a possibility for constructing an all-in-one multiplex with relevant forensic markers that include STRs, SNPs, indels, and mtDNA markers and along with many other information. It is evident from the publications in the last few years that the use of NGS in forensic genetics is presently an important research area and will be in the near future as well. The technology has the enormous potential and could offer the first real alternative to PCR-CE analysis. Future research for development and improvement of NGS will make the technology be applied more smoothly, conveniently, and effectively for forensic DNA typing.

References

1. Abdelkrim J, Robertson B, Stanton JA, Gemmell N (2009) Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques 46:185–192
2. Allentoft M, Schuster SC, Holdaway R, Hale M, McLay E, Oskam C et al (2009) Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. BioTechniques 46:195–200
3. Allentoft ME, Oskam C, Houston J, Hale ML, Gilbert MT, Rasmussen M et al (2011) Profiling the dead: generating microsatellite data from fossil bones of extinct megafauna – protocols, problems, and prospects. PLoS One 6:e16670
4. Ayres KL, Chaseling J, Balding DJ (2002) Implications for DNA identification arising from an analysis of Australian forensic databases. Forensic Sci Int 129:90–98
5. Berglund EC, Kiialainen A, Syvänen AC (2011) Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet 2:23–37
6. Bocklandt S, Lin W, Sehl ME, Sanchez FJ, Sinsheimer JS, Horvath S et al (2011) Epigenetic predictor of age. PLoS One 6:e14821
7. 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 Dispatches:1–6
8. Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18:78–89
9. Børsting C, Fordyce SL, Olofsson J, Mogensen HS, Morling N (2013) Evaluation of the ion torrent HID SNP 169-plex: a SNP typing assay developed for human identification by second generation sequencing. Forensic Sci Int Genet 12:144–154
10. Brenig B, Beck J, Schütz E (2010) Shotgun metagenomics of biological stains using ultra-deep DNA sequencing. Forensic Sci Int Genet 4:228–231
11. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, Diaz de Ståhl T, Menzel U et al (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy number-variation profile. Amer J Hum Genet 82:763–771
12. Budowle B (2014) Editors pick: Molecular genetic investigative leads to differentiate monozygotic twins. Invest Genet 5:11
13. Butler JM (2012) Advanced topics in forensic DNA typing: methodology. Elsevier, Inc, Waltham
14. Caratti S, Turrina S, Ferrian M, Cosentino E, Leo DD (2015) MiSeqFGx sequencing system: a new platform for forensic genetics. Forensic Sci Int 5:e98. https://doi.org/10.1016/j.fsigss.2015.09.040
15. Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning AP, Fox SE et al (2011) Discovery of highly divergent repeat landscapes in snake genomes using high throughput sequencing. Genome Biol Evol 3:641–653
16. 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
17. 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. Forensic Sci Int Genet Suppl Ser 4:e218–e219
18. Dalsgaard S, Rockenbauer E, Buchard A, Mogensen HS, Frank-Hansen R, Børsting C, Morling N (2014) Non-uniform phenotyping of D12S391 resolved by second generation sequencing. Forensic Sci Int Genet 8:195–199
19. Eduardoff M, Santos C, de la Puente M, Gross TE, Fondevila M, Strobl C et al (2015) Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the ion PGM™. Forensic Sci Int Genet 17:110–121

20. Eduardoff M, Gross TE, Santos C, de la Puente M, Ballard D, Strobl C et al (2016) Inter-laboratory evaluation of the EUROFORGEN global ancestry-informative SNP panel by massively parallel sequencing using the ion PGM™. Forensic Sci Int Genet 23:178–189
21. Fordyce SL, Avila-Arcos MC, Rockenbauer E, Børsting C, Frank-Hansen R, Petersen FT, Willerslev E, Hansen AJ, Morling N, Gilbert MT (2011) High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche genome sequencer FLX platform. BioTechniques 51:127–133
22. Fordyce SL, Mogensen HS, Børsting C, Lagac'e RE, Chang CW, Rajagopalan N, Morling N (2015) Second-generation sequencing of forensic STRs using the ion TorrentTM HID STR 10-plex and the ion PGMTM. Forensic Sci Int Genet 14:132–140
23. Frumkin D, Wasserstrom A, Budowle B, Davidson A (2011) DNA methylation-based forensic tissue identification. Forensic Sci Int Genet 5:517–524
24. Gelardi C, Rockenbauer E, Dalsgaard S, Børsting C, Morling N (2014) Second generation sequencing of three complex STRs D3S1358, D21S11 and D12S391 in Danes and a proposal for nomenclature of sequenced STR alleles. Forensic Sci Int Genet 12:38–41
25. Gettings KB, Aponte RA, Vallone PM, Butler JM (2015) STR allele sequence variation: current knowledge and future issues. Forensic Sci Int 18:118–130
26. Gettings KB, Kiesler KM, Faith SA, Montano E, Baker CH, Young BA et al (2016) Sequence variation of 22 autosomal STR loci detected by next generation sequencing. Forensic Sci Int Genet 21:15–21
27. Grunau C, Clark SJ, Rosenthal A (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res 29:E65–E65
28. Guo F, Zhou Y, Liu F, Yu J, Song H, Shen H et al (2016) Evaluation of the early access STR kit v1 on the ion torrent PGM™ platform. Forensic Sci Int Genet 23:111–120
29. Gymrek M, Golan D, Rosset S, Erlich Y (2012) lobSTR: a short tandem repeat profiler for personal genomes. Genome Res 22:1154–1162
30. Hares DR (2012) Expanding the CODIS Core Loci in the United States. Forensic Sci Int Genet 6:e52–e54
31. Hares DR (2015) Selection and implementation of expanded CODIS core loci in the United States. Forensic Sci Int Genet 17:33–34
32. Holland MM, McQuillan MR, O'Hanlon KA (2011) Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy. Croat Med J 52:299–313
33. Jeffreys AJ, Brookfield JF, Semeonoff R (1985a) Positive identification of an immigration test-case using human DNA fingerprints. Nature 317:818–819
34. Jeffreys AJ, Wilson V, Thein SL (1985b) Individual-specific 'fingerprints' of human DNA. Nature 314:67–74
35. Jobbing MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. Nat Rev Genet 5:739–751
36. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12:179–192
37. Kim EH, Lee HY, Yang IS, Jung SE, Yang WI, Shin KJ (2016) Massively parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons. Forensic Sci Int Genet 22:1–7
38. King JL, LaRue BL, Novroski NM, Stoljarova M, Seo SB, Zeng X et al (2014) Highquality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Sci Int Genet 12:128–135
39. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, Howard E, de Lima RL et al (2002) Mutations in IRF6 cause van der woude and popliteal pterygium syndromes. Nat Genet 32:285–289
40. Li C, Zhao S, Zhang N, Zhang S, Hou Y (2013) Differences of DNA methylation profiles between monozygotic twins' blood samples. Mol Biol Rep 40:5275–5280

41. Li R, Montpetit A, Rousseau M, Wu SY, Greenwood CM, Spector TD, Pollak M, Polychronakos C, Richards JB (2014) Somatic point mutations occurring early in development: a monozygotic twin study. J Med Genet 51:28–34
42. Loreille O, Koshinsky H, Fofanov VY, Irwin JA (2011) Application of next generation sequencing technologies to the identification of highly degraded unknown soldiers' remains. Forensic Sci Int Genet Suppl Ser 3:e540–e541
43. McElhoe JA, Holland MM, Makova KD, Su MS, Paul IM, Baker CH et al (2014) Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Sci Int Genet 13:20–29
44. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33:5868–5877
45. Meldrum C, Doyle MA, Tothill RW (2011) Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev 32:177–195
46. Mikkelsen M, Rockenbauer E, Wächter A, Fendt L, Zimmermann B, Parson W, Abel Nielsen S, Gilbert T, Willerslev E, Morling N (2009) Application of full mitochondrial genome sequencing using 454 GS FLX pyrosequencing. Forensic Sci Int Genet Suppl Ser 2:518–519
47. Paliwal A, Vaissiere T, Herceg Z (2010) Quantitative detection of DNA methylation states in minute amounts of DNA from body fluids. Methods 52:242–247
48. Parson W, Strobl C, Huber G, Zimmermann B, Gomes SM, Souto L et al (2013a) Reprint of: evaluation of next generation mtGenome sequencing using the ion torrent personal genome machine (PGM). Forensic Sci Int Genet 7:632–639
49. W. Parson, C. Strobl, G. Huber, B. Zimmermann, S.M. Gomes, L. Souto, L. Fendt, R. Delport, R. Langit, S. Wootton, R. Lagace´, J. Irwin, Evaluation of next generation mtGenome sequencing using the ion torrent personal genome machine (PGM), Forensic Sci Int Genet 7 (2013b) 543–549
50. Phillips C, Gelabert-Besada M, Fernandez-Formoso L, García-Magariños M, Santos C, Fondevila M, Ballard D, Syndercombe Court D, Carracedo A, Lareu MV (2014) New turns from old STaRs: enhancing the capabilities of forensic short tandem repeat analysis. Electrophoresis 35:3173–3187
51. Pitterl F, Schmidt K, Huber G, Zimmermann B, Delport R, Amory S, Ludes B, Oberacher H, Parson W (2010) Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous south African and central Asian populations. Int J Legal Med 124:551–558
52. Planz JV, Sannes-Lowery KA, Duncan DD, Manalili S, Budowle B, Chakraborty R, Hofstadler SA, Hall TA (2012) Automated analysis of sequence polymorphism in STR alleles by PCR and direct electrospray ionization mass spectrometry. Forensic Sci Int Genet 6:594–606
53. Rockenbauer E, Hansen S, Mikkelsen M, Børsting C, Morling N (2014) Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing. Forensic Sci Int Genet 8:68–72
54. Santana Q, Coetzee M, Steenkamp E, Mlonyeni O, Hammond G, Wingfield M et al (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. BioTechniques 46:217–223
55. Schadt EE, Turner S, Kasarskis A (2010) A window into third-generation sequencing. Human Mol Genet 19:2
56. Scheible M, Loreille O, Just R, Irwin J (2011) Short tandem repeat sequencing on the 454 platform. Forensic Sci Int Genet Suppl Ser 3:e357–e358
57. Scheible M, Loreille O, Just R, Loreille O, Just R, Irwin J (2014) Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers. Forensic Sci Int Genet 12:107–109
58. Seo SB, King JL, Warshauer DH, Davis CP, Ge J, Budowle B (2013) Single nucleotide polymorphism typing with massively parallel sequencing. Int J Legal Med 127:1079–1086

59. Shewale JG, Qi L, Calandro LM (2012) Principles, practice, and evolution of capillary electrophoresis as a tool for forensic DNA analysis. Forensic Sci Rev 24:79–100
60. Van Neste C, Van Nieuwerburgh F, Van Hoofstat D, Deforce D (2012) Forensic STR analysis using massive parallel sequencing. Forensic Sci Int Genet 6:810–818
61. Voelkerding KV, Dames SA, Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. Clin Chem 55:641–658
62. Wang Z, Zhang S, Bian Y, Li C (2015) Differentiating between monozygotic twins in forensics through next generation mtGenome sequencing. Forensic Sci Int Genet Supp Ser 5:e58–e59
63. Wang Z, Zhou D, Cao Y, Hu Z, Zhang S, Bian Y et al (2016) Characterization of microRNA expression profiles in blood and saliva using the ion personal genome machine® system (ion PGM™ system). Forensic Sci Int Genet 20:140–146
64. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL et al (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37:853–862
65. 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
66. Zeng X, King J, Hermanson S, Patel J, Storts DR, Budowle B (2015a) An evaluation of the PowerSeq™ auto system: a multiplex short tandem repeat marker kit compatible with massively parallel sequencing. Forensic Sci Int Genet 19:172–179
67. Zeng X, King JL, Stoljarova M, Warshauer DH, LaRue BL, Sajantila A et al (2015b) High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing. Forensic Sci Int Genet 16:38–47
68. Zhao X, Li H, Wang Z, Ma K, Cao Y, Liu W (2016) Massively parallel sequencing of 10 autosomal STRs in Chinese using the ion torrent personal genome machine (PGM). Forensic Sci Int Genet 25:34–38

Chapter 16
Unique Individualistic Microflora: The Future of DNA Fingerprinting Technique

Pankaj Shrivastava, Hirak R. Dash, Sonia Kakkar, Mahendra K. Gupta, and Toshi Jain

Abstract Microbial forensics is a newly emerging discipline in forensic science, which is the amalgam of both classical and advanced microbiology. Microbial phylogenetics and bioinformatics also play an important role in microbial forensic analysis. The omnipresence of bacteria and its uniqueness to a particular individual makes microbial genome analysis a potential tool for personal identification in addition to human genome analysis. Most of the studies on microbial forensics are based upon bioweapons; however, use of bacterial community for individualization and their possible role in body fluid degradation resulting into failure of even the most sensitive DNA fingerprinting technique is also of major concern and needs to be explored. From microbial forensics point of view, in addition to the conventional practices such as 16S rRNA and other housekeeping gene sequencing, metagenomic analysis by using high-throughput sequencing and polyphasic taxonomic approach can be employed for a better output in criminal investigation. This review unveils the current status of microbial forensics and takes the account of future requisites that should be inculcated in the present technology to probe forensic microbiology in criminal casework more efficaciously.

Keywords Forensic · Microbiology · Biocrimes · Human identification · NGS

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

S. Kakkar
Department of Forensic Medicine, PGIMER, Chandigarh, India

M. K. Gupta · T. Jain
School of Studies in Microbiology, Jiwaji University, Gwalior, Madhya Pradesh, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_16

16.1 Introduction

An eyewitness in criminal cases may become antagonistic, but scientific evidences collected during scientific investigation and their obtained laboratory results become solid evidence which will never turn deceptive. Forensic science can be related to a tool for criminal investigation using various branches of core science such as chemistry, biology, and serology, ballistics, physics, photography, and DNA fingerprinting. During this process, the corresponding scientific knowledge and technologies are used as witnesses in the court of law either to identify the accused or exonerate the innocence. Microbial forensics is the relatively new discipline of forensic science which deals with the identification, origin, and potential effects of the potent microbes and to present them in the criminal justice system. This evidence is crucial to the successful execution of cases of human host identification, bioterrorism, and medical negligence. This field attained the limelight after the anthrax attack in New York (US) in 2001 [2]. After 2 years of anthrax attack, quality assurance guidelines were formalized and published by FBI's Scientific Working Group for Microbial Genetics and Forensics [14]. The major role of microbial forensic is to identify the biological threat and its source of origin. Identification of the microbial/biological threat further helps in inferring the potential risks involved with it. Taking into account the above information the preventive measures can be opted, which can further control the havoc. Finally, these findings can be used to reach on to the final conclusion of "who committed the crime?" Microbial forensics exploits the principles of microbial genomics, phylogenetics, and bioinformatics to probe various biocrimes.

The human body is a dwelling to millions of microbes residing exogenously as well as endogenously. Unique microorganisms in terms of their number and variety are present in a human body which is dependent on the daily activities, antibiotics, and medicine use of the individual [81]. Prenatal skin is completely sterile; however, colonization of microbes starts just after birth [24, 75]. Exogenous flora, also known as the transient microbial flora, is a temporary microbial flora of human skin due to the environment which a person is exposed to, while the endogenous, also known as the resident microbes, are the microbes that reside over the body permanently and are a result of the endogenous factors. Microbial colonization depends on the natural balance between skin surface and the host's variables such as age and sex of the host and other environmental factors. Natural immunity response of an individual may also alter skin microbial diversity [35]. Variability in microbial flora can also be observed due to changes in the diet, stress, sexual behavior, medication, hormonal changes, and other host-related factors [45, 54, 55, 68, 73].

16.2 The Human Skin Microbiome

Numerous eukaryotic fungi and/or protists colonize on human skin, saliva, and oral and gut mucosa and on most of the body parts (Fig. 16.1). Diverse groups of commensal microbes perform many functions in the host's body, e.g., provide essential

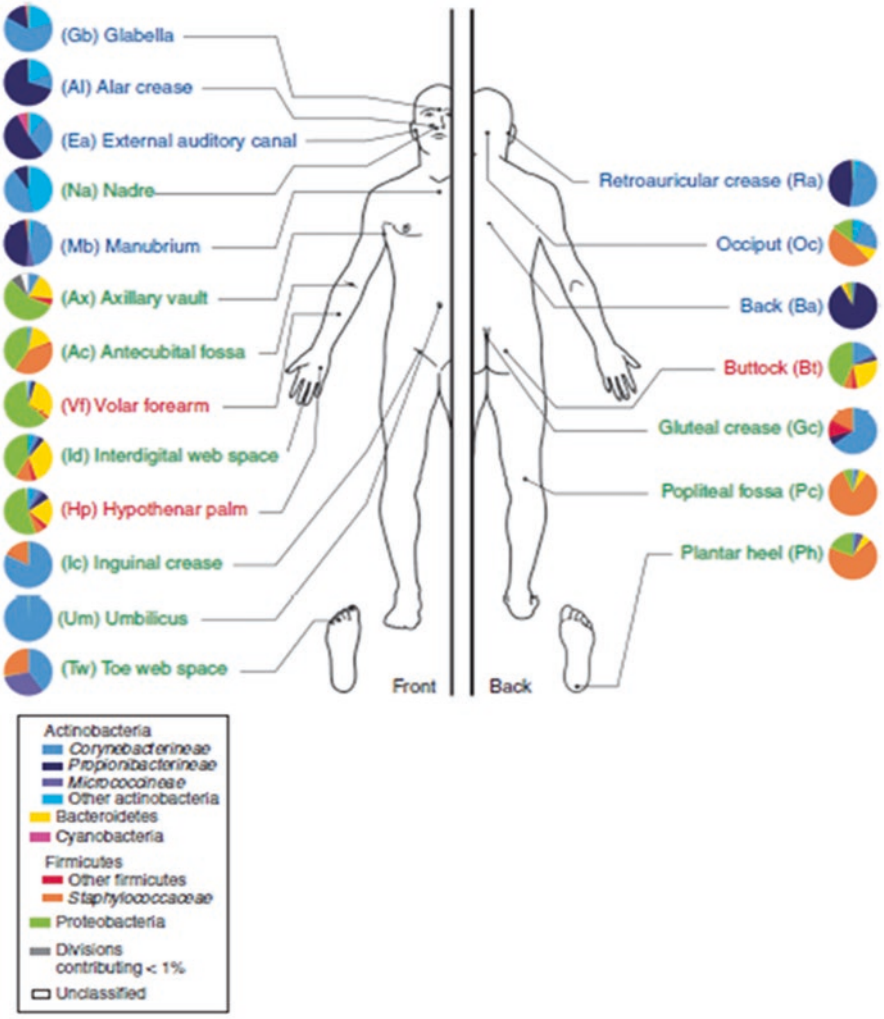

Fig. 16.1 Topographical distribution of bacteria on various skin sites of human. (Reprinted from Nat Rev. Microbiol. Author manuscript available in PMC 2013 January 2003)
Note: *Sites are grouped as* sebaceous or oily (blue circles), moist (typical skin creases) (green circle), and dry, flat surfaces (red circles). The sebaceous sites are the glabella (between the eyebrows), alar crease (side of the nostril, external auditory canal, inside the ear), retroauricular crease (behind the ear), occiput (back of the scalp), manubrium (upper chest), and back. *Moist sites are* the nare (inside the nostril), axillary vault (armpit, antecubital fossa (inner elbow), interdigital web space (between the middle and ring fingers, inguinal crease (side of the groin), gluteal crease (topmost part of the fold between the buttocks), popliteal fossa (behind the knee), plantar heel (bottom of the heel of the foot), toe web space, and umbilicus (navel). *Dry sites are* the volar forearm (inside of the mid-forearm, hypothenar palm (palm of the hand proximal to the little finger), and buttock. Data from reference [36].

nutrients to the hosts, help in the digestion of food particles, and provide defense against opportunistic pathogens [7, 56, 87, 91]. Some of the microorganisms live in symbiotic association with skin which further prevents the invasion of other harmful microbes. *Staphylococcus epidermidis* and other coagulase-negative Staphylococci are considered to be the initial bacterial colonizers of human skin. Other healthy skin colonizers include *Corynebacterium* sp., *Propionibacterium* sp., *Dermabacter* sp., and *Brevibacterium* sp. in a human body [19]. The group corynebacteria may be of two types, i.e., lipophilic or nonlipophilic. As the name suggests, lipophilic microbes reside on the areas of lipid abundance, e.g., axilla of a human body. The signature characteristic features of corynebacteria include being nonmotile and nonproduction of toxins. Some common examples of corynebacteria residing on the human body include *Arcanobacterium haemolyticum*, *C. xerosis*, *C. minutissimum*, and *C. striatus*. Out of these, the anaerobic coryneform *Propionibacterium acne* is most abundantly found on the skin. Being anaerobic in nature, *Propionibacterium acne* grows deep in the adnexal structures. Unique microbiota is present for many cutaneous invaginations and appendages such as apocrine and eccrine glands, sebaceous glands, and hair follicles [51]. Eccrine glands are abundant on skin surfaces in comparison to the apocrine glands. The composition of the secretions of these glands includes salt and water, and they continuously bathe the skin surface with their secretions. Thermoregulation is the major function associated with sweat. Water and electrolyte secretion from the eccrine glands acidify skin, which further prevents the colonization and growth of microorganisms. Apocrine glands are present in many axillary regions such as nipples, axillary vault (armpit), and genital areas. Adrenaline regulates these glands to synthesize many odorless, milky, viscous secretions. These secretions contain pheromones, molecules which are responsible for behaviors like sex or alarm in the individuals [20]. Sebaceous glands secrete a lipid-rich substance called sebum; sebum forms a hydrophobic coating over the skin and hair and provides an antibacterial shield. These glands enhance the growth of facultative anaerobic, common skin commensal bacteria, i.e., *Propionibacterium acnes* [51, 53]. The whole-genome sequencing of *P. acnes* has found multiple genes encoding lipase enzyme which breaks skin lipids of sebum [8]. *P. acnes* hydrolyses the triglyceride found in sebum and releases fatty acids. Subsequently, bacteria stick to this free fatty acid and helps in colonization of microbes on the sebaceous gland [34]. Also, these free fatty acids decrease the skin pH. This decreased pH prevents the colonization of skin pathogens like *Staphylococcus aureus* and *Streptococcus pyogenes* and the growth of coagulase-negative Staphylococci and *Corynebacterium* [25, 47]. Skin occlusions can result in higher skin pH and favor the growth of *Staphylococcus aureus* and *Streptococcus pyogenes* [5]. Skin microbiome is dependent on the density of sebaceous glands in the region. Thus, body areas rich in sebaceous gland such as the back, chest, and face enhance the growth of lipophilic microbes [72]. The skin of the arm and leg is relatively shrivelled, and surface temperature fluctuation is low, so compared with other skin sites, these sites were found to be invaded by fewer organisms (quantitatively) than other moist areas of the body [53]. The colonization of microbes on the skin varies topographically as the skin surface

anatomy is regionally variable (Fig. 16.1). Skin regions like axillary vault and groin show a higher temperature and humidity, thus promoting the growth of organisms surviving in such moist conditions such as Gram-negative *Bacilli*, *Staphylococcus aureus*, and coryneforms [72].

16.3 Microbial Colonization of the Skin Forming the Grounds for Forensic Identification

The reason behind the emphasis laid on health and hygiene by healthcare practitioners is the bacteria colonizing the human skin can be dislodged and transferred to the objects on touching [44, 66]. Principal of exchange given by Locard also supports it, which says that wherever you go, you leave a trace and carry a trace from it [17]. This transfer of bacteria from the skin surface to objects lays the foundation of microbial forensic investigations. These bacteria may persist on the objects for prolonged periods as they are resistant to environmental stress, which includes moisture, temperatures, and UV radiations [11, 82]. The main motive of comparing the genetic profile of evidentiary and reference microbial sample is to find out whether the source or lineage of the samples is the same or different. Many studies have confirmed diversity of microflora and their variation among individuals' skin [2, 21, 28, 31]. A study by Fierer et al. [28] revealed the sharing of 17% of species-level phylotype between two hands of an individual, whereas 13% sharing observed for interindividual level. The principal reason for observed variation in skin microbiota of hands could be due to frequent washing of hands, exposure to environmental conditions, and/or handedness preferences. A higher bacterial diversity is observed in females than males. However, the varied microflora among different sex has not understood properly which may be due to difference in physiological factors, hygiene, and cosmetic use [28, 31]. 16S rRNA metagenomic sequencing revealed the higher dynamics in intrapersonal skin microflora than that of the interpersonal variation [21, 31]. In an experiment [79] collected transient microflora (TM) from the physical fingerprints prior washing the hands and resident microflora after washing the hands with soap; both the microflora were analyzed using PFGE (pulse field gel electrophoresis). TM showed a higher number of PFGE types than RM. Reason behind that was explained to be exposure to what?? and picking up of nonendogenous species from the environment. It was shown that resident microflora collected after washing the hands were very similar intraindividually and very different interindividually. The molecular variations between the similar strains can help in determining the origin and transmission routes of a particular sample. So if the resident or endogenous microflora can be collected from the crime scene evidences, it may assist the investigator in deciphering the geographical location of human host involved in a crime, which may help in narrowing down the investigation procedures. Stability and interpersonal variability of bacteria make it valuable for personal identification. A huge diversity of human fingerprint microflora raised skepticism in their use for human identification [79].

16.4 Applications of Microbial Forensics

Huge bacterial diversity is resident to human skin. Bacterial flora present on the human skin provides huge information of forensic relevance such as individualization of host or its geographical origin [79]. It has been proved that the transient/exogenous, i.e., the temporary microflora on the skin, differs frequently from resident or endogenous microflora of the same individual [79]. Microbial DNA fingerprinting of the physical fingerprints left on the touched objects at the crime scene can give potential information about the human host inference. Getting information about the human host can be helpful in narrowing down the investigations. Endogenous or resident microflora retrieved from the physical fingerprints can intimate the investigator about the geographical origin of the perpetrator. Because of its restriction to a particular region it helps in deciphering the source of origin. The information regarding geographical origin is helpful in cases devoid of suspects or where DNA match result is exclusion. It can direct the investigation to a particular geographical location [79]. Fierer et al. [29] studied the bacterial community associated with the skin which can be recovered from the touched objects like computer keys, mice, etc. Bacteria such as *Streptococcus salivarius* and *Streptococcus mutans* have shown immense potential in identification of saliva on the skin, clothes, cigarette butts, bitten apples, and mixed body fluid differentiation. Polymerase chain reaction (PCR) analysis results in the revealing of *Streptococcus salivarius* and *Streptococcus mutans* in 83.5% and 67% of the saliva-contaminated samples. Additionally, the absence of both these bacteria in other body fluids [4] suggests their superior usability for forensics applications.

Martini et al. have found the potential skin pathogens in used clothes. Panties and bras were found to have the highest count of bacteria and fungi. Cotton present in the clothes and other materials like bedsheets, etc. acts as the carrier of bacteria in them from human contact, mostly due to the presence of moisture from varied origin, i.e., semen, saliva, blood, vaginal secretions, secretions from wounds, or spilled drinks. So when the bedsheets are changed, the deeper layers always carry the pathogen as the cotton threads act as wick to transfer the pathogens through moisture [60]. Analysis of Nelly et al. [62] showed the survival of *Staphylococci* and *Enterococci* for an extended period on various materials used by patients or the fabrics used in hospitals. Takashima et al. [85] demonstrated the better penetration of microbes on polyester or acrylic fabrics better than cotton. It was illustrated by Horswell et al. [38] that the DNA profile from the microbes can be generated from the soil samples associated with shoe sole and from the stains on the clothing. The microbial DNA profiles obtained from aforementioned sources represented the possible site of sample collection and can be utilized as a supportive evidence to provide the clue between the perpetrator and spot. Turnbaught et al. [88] showed that even identical twins harbor substantially variable microbial flora which were until recently could not be differentiated even by DNA fingerprinting technology. The omnipresence of bacteria and its uniqueness to a particular individual make microbial genome analysis a better selection for personal identification than human genome analysis.

16.5 Tools and Techniques to Study Microbial Forensics

16.5.1 16S rRNA Gene-Based Identification of Bacterial Strains

Earlier studies of comprehensive skin microflora were limited due to the nonavailability of proper growth conditions for isolation and growth of fastidious microbes. In vitro cultivation of only 1% of bacteria on earth has been carried out till date, popularly known as the "great plate count anomaly" [6, 39, 84]. These shortcomings of the culture-based techniques instigated the researchers to introduce a change, and, hence, the16S rRNA-based metagenomics has been inducted to study the microbes. The use of genomic technique to identify bacterial strains proved out to be more efficacious as it helped in revealing much more diversity in the bacterial community [21, 28, 31].

With the advent of culture-independent techniques, mixed bacterial strains have been characterized with a great ease from environmental and other biomass samples such as animals as well as the humans. The sequence of events during this procedure includes isolation of metagenomic DNA from the sample, followed by amplification of housekeeping gene such as 16S rRNA gene by PCR, cloning of the amplified product, and sequencing for identification [32, 63]. Thus, huge varieties of novel microbial phylotypes have been detected from a pool of bacterial communities such as seawater, soil, and humans [1, 59, 69, 94].

16S rRNA may be found either in bacteria, archaea, mitochondria, or chloroplasts in plants. Sequencing of 16S rRNA gene is mostly used for identification of prokaryotic organisms, hence deducing their phylogenetic relationship. Many advantages of using 16S rRNA gene-based identification of bacteria include the presence of ribosomes and ribosomal RNA in all cells, conserved nature of 16S rRNA gene, and absence of microbial cell culture in sequencing techniques [23].

Certain specific nucleotide base sequences found exclusively in all groups are called as signature sequences. These signature DNA sequences of about five to ten bases long, found categorically in the 16S rRNA gene, are unique in many groups of prokaryotes, archaea, as well as eukarya. The hypervariable regions present in these 16S rRNA gene sequences are the sequences which showed the divergence/evolution of organisms over time. In most of the organisms, conserved sequences flank the variable regions in the 16S rRNA gene. Hence, primers are designed in such a way to bind at the conserved region followed by amplification of the variable regions for further use. Till date, 16S rRNA gene sequence has been analyzed for a substantial number of microbial species in comparison to any other gene, thus increasing its usability. In this regard, over ten thousands of 16S rRNA gene sequences have been deposited at National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) and the Ribosomal Database Project (http://rdp.cme.msu.edu/) for reference from varied sources and comparing new sequences to the available database.

The 16S rRNA gene is the least variable DNA in all prokaryotic cells. Thus, the variability of this sequence is used to study the identification, phylogeny, and determination of species divergence rates among bacteria. Additionally, as this DNA is not translated into proteins, hence, analysis 16S rRNA gene sequences is regarded as the most suitable tool for microbial taxonomy analysis.

16.5.2 Use of Other Housekeeping Genes

Though sequencing of 16S rRNA gene is regarded as the yardstick for bacterial identification and phylogeny-related studies, however, in certain instances, it does not produce concrete differentiation at interspecies level [43, 67]. Additionally reporting of horizontal gene transfer of 16S rRNA gene in various groups of bacteria such as *Streptococcus* sp. has questioned the authenticity of this technique [46, 78]. Thus, in addition to 16S rRNA gene sequencing, various other housekeeping gene sequences have also been used to compare the phylogeny and relatedness among bacterial strains. These housekeeping genes can very well be applied in the field of microbial forensics which may include *rpoA*, *rpoB*, *rpoC*, *rpoD*, *gyrA*, *gyrB*, *recA*, *recN*, *dnaJ*, and many others [23]. *rpo* genes have been widely employed in distinguishing various species of bacterial strains that include *Enterococcus* sp., lactic acid bacteria, *E. coli*, and *Mycobacterium tuberculosis* which are the key microbes from forensic investigation point of view. Additionally the ubiquitous and essential nature of gene encoding DNA topoisomerase/DNA gyrase, i.e., *gyrA* of *gyrB*, makes them a suitable target to study bacterial phylogeny. *Bacillus* group of microorganisms has been discriminated very well by targeting these genes in them, i.e., *Bacillus subtilis* and *Bacillus cereus*. Sequencing of *gyr* genes is also in practice to differentiate species of *E. coli* and *Pseudomonas putida*. In addition to the aforementioned gene sequences, the other housekeeping genes that are in practice for bacterial identification have been mentioned in Table 16.1.

16.5.3 Newly Emerging Trends in Microbial Forensic Analysis

The aim of analyzing microbial forensic evidence is to find out the source of the sample. Analysis of microbial forensic evidence can be done by both genetic and nongenetic methods. Chemical and physical analysis methods may help in determining the procedure used to prepare, store, and spread the pathogen. Epidemiology is the science which deals with spread and control of a disease. Epidemiological studies help in deducing causative pathogen and its etiology, i.e., the source of a particular pathogen/disease. So while analyzing microbial forensic evidence, epidemiology should also be considered as an imperative part of investigation as it has the potential to solve the purpose of attribution [18, 58, 86].

The main motive of comparing the genetic profile of evidentiary and reference microbial sample is to find out whether the source or lineage of the samples is the

Table 16.1 Most commonly used housekeeping genes for identification of bacterial entities

Target gene	Function of the encoded protein	Target group of microorganisms	References
16S rRNA	Small subunit of ribosome	All bacteria, archaea	Janda and Abott [43] and Woo et al. [93]
rpoA/B/C/D	β subunit of bacterial RNA polymerase	*Enterococcus* sp., lactic acid bacteria, *Mycobacterium tuberculosis*, *E. coli*	
gyrA/B	Type II DNA topoisomerase/ DNA gyrase	*B. subtilis*, *B. cereus*, *E. coli*, *Pseudomonas putida*	Das et al. [23]
gapA	Synthesis of D-glyceraldehyde-3-phosphate dehydrogenase	*E. coli*, *Mycoplasma* sp.	Goh et al. [33]
pyrH	UMP phosphorylation by UMP kinase	*Vibrio vulnificus*, *Vibrio* sp.	
dnaA/K/J	Protein responsible for initiation of DNA replication	*Mycobacterium* sp., *Alphaproteobacteria*	Alexandre et al. [3]
pheS	Encodes phenylalanyl-tRNA synthase	*Enterococcus* sp., *Lactobacillus* sp.	Naser et al. [61]

same or different. If the lineage is different, the extent of difference has to be recognized to rule out the similarity or difference in the source of origin. The pathogen maintained in laboratory under specific conditions shows a little diversity than the same pathogen samples present in nature; some genes in laboratory-maintained bacteria tend to have 100-fold more mutation rate. So if the biological weapon is laboratory maintained, even single genetic difference between the evidentiary and control samples can be considered as significant [48, 50, 77, 83]. Analysis of microbial forensic evidence is a little cumbersome task as in environment, some species may exchange genetic material, and some pathogens are significantly similar to their non-pathogenic strains. However, it is highly important to store the genomic profile of non-pathogenic strains also [37, 40, 41, 64, 70, 76].

16.5.4 Genomic Approach to Study Human Microbiome

Human microbiome is the genetic catalog of various microbes that are commensal to the human body [89]. The term microbiota differs from microbiome and is considered to be the second-line genome of an organism [12]. The term microbiome was first introduced by Joshua Lederberg in 2001. The Human Microbiome Project was started in 2008 to study the microbiome of individuals and their variations [65]. It is also established that after completion of the Human Microbiome Project, there exists a huge microbial genomic variations among individuals [89]. Thus, there exist huge potential applications of human microbiome in terms of personal medicine as well as diagnosis of dysbiosis-related disorders [26].

16.5.5 Application of High-Throughput Sequencing (HTS)/ Next-Generation Sequencing (NGS) in Microbial Forensics

The quality assurance guidelines were formalized and published by FBI's Scientific Working Group for Microbial Genetics and Forensics [14] to investigate any biocrime and for a better emergency response to any terror attack. Cummings et al. [22] confirmed suspects by the use of NGS and sequencing of four strains, *Bacillus anthracis* and *Yersinia pestis* by the use of SOLiD. 454 sequencing system was used by Brenig et al. [9] for the identification of traces using metagenomic analysis followed by sequencing. This confirms the utility of these techniques in forensic implications. Fierer et al. [29] studied the skin microflora left on articles by using NGS metagenomics and proved that bacteria transferred to objects from the skin have forensic relevance. With the advent of technology, HTS technique was inculcated in QA guidelines to probe the biocrimes and bioterror activity. HTS was developed to study human genome [49, 90]. Initially, bacterial genome was first sequenced in 1995 with the perspective of using this technique for microbial forensic probes [30]. Culturing of the bacteria is not required in HTS, so it makes metagenomics quite fast. The possibility of assigning the features like identity, strain, virulence profile, etc., which can lead to the attribution of biocrime evidence, is greatly enhanced by HTS analysis. Also HTS can aid the investigators in recognizing an unknown microbial agent, genetically engineered microorganism present mixed in other highly sensitive samples [15]. HTS possesses few unique features like high-throughput reduced cost per nucleotide and automation capability which makes it better than Sanger sequencing [73, 92].

HTS allows millions of reactions to take place in a parallel fashion in a single instrument [10, 52, 57, 69, 80]. When associated with bioinformatics tools, HTS detects desired microbes even in less quantity [16]. HTS technology for microbial forensics includes assembly of completed single genome of microbes, metagenomics, targeted site sequence, and source attribution by profiling and sample comparison [16].

Whole-genome shotgun metagenomic sequencing can describe the complete genetic diversity of microbes which enables the prediction of gene attributes associated with the skin microflora [35]. NGS offeres many advantages over capillary sequencing or PCR-based techniques, as it allows culturefree detection of most of the microbial entities present in a sample and analysis of multiple samples at a time.It is cost-effective as well. Studies have found NGS to be effective for high throughput and accuracy along with multiplexing. Additionally, it is in an upper hand over conventional practices as it can generate the results in less than 24 h (Fig. 16.2). The abovementioned features of NGS make it a valuable invention for the swift whole-genome sequencing of the pathogens for forensic and epidemiological purpose.

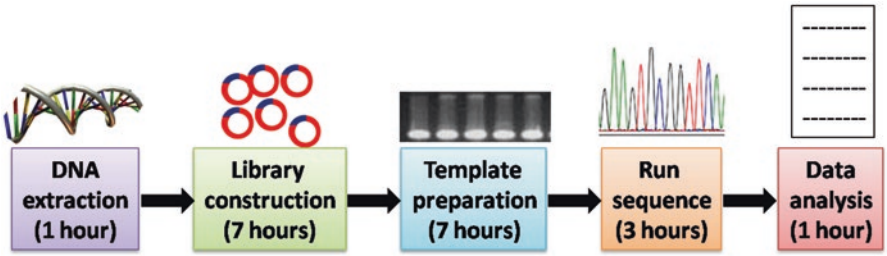

Fig. 16.2 NGS workflow: sample to results in less than 24 h

16.6 Human Forensic Genetic vs Microbial Forensic Genetic Analysis

There exist similarities and differences both in human and microbial forensic genetics which have been discussed in detail below. However, both can solve the same purpose of individualization and relatedness, unraveling the link between the host, suspect, victim, and culprit (Fig. 16.3).

16.6.1 Similarities

Many molecular biology techniques have similarity between the microbiome study and the human genome study, i.e., extraction of nucleic acids, PCR, or multiplex PCR and genotyping and/or sequencing of nucleic acids. Other similar aspects include use of population database, qualitative inference, as well as the utilization of quality assurance (QA) and quality control (QC) practices [42].

16.6.2 Differences

There exist many significant differences mostly in terms of data interpretation and statistical approach of microbial forensic genetics cases. Budowle and Chakraborty [13] pointed many differences such as size and composition of database, techniques of statistical evaluation, and interpretation. The biggest difference between the eukaryotes and the prokaryotes is that most of the viruses and bacteria are haploid, whereas human beings are normally diploid in nature. Hence, the statistical tools used for diploid organisms such as humans cannot be applied to the prokaryotes like viruses and bacteria. Unlike human forensic genetics, the identity cannot be established in case of microbial forensic genetics analysis because of

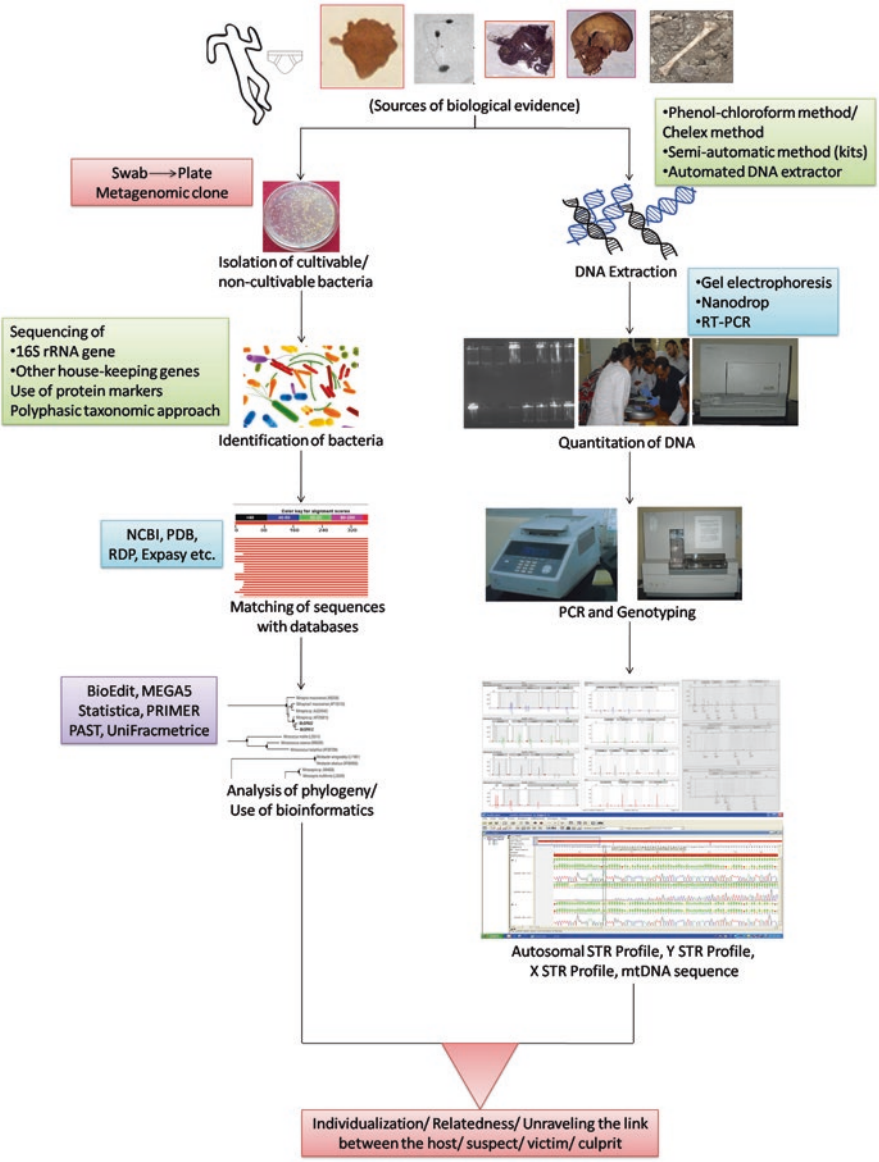

Fig. 16.3 Flowchart of human forensic DNA typing and microbial forensic analysis

their asexual reproduction and they being clonal in nature. The inference that can be made is of lineage and not the identity. Thus, the aforementioned differences affect the structure and methodology used for microbiome-based forensic application of the technique.

16.7 Challenges in Microbial Forensic Analysis

The biggest challenge in microbial forensic is having a database of enormously diverse microbial flora residing over the human skin from various geographical regions. Additionally, keeping the data record of mutation rates in every microorganism is even more challenging. The laboratory-made microbes mutate 100 times faster than the naturally existing ones [48, 50, 77, 83]. Different statistical approaches have to be designed to interpret the microbial forensic analysis results as the microbes are haploid and the statistical parameters used for human identification are based on diploid genome structure. Some pathogenic strains are similar to their non-pathogenic strains as they can transfer their genetic material, so it's a challenging task to identify the pathogenic strain over the non-pathogenic so that situation can be taken under control promptly [37, 40, 41, 64, 70, 76].

Utmost attention and strict preventive measures are to be taken while dealing with lethal microbes. The challenge is to develop such laboratories in developing countries where people are starving even for basic amenities.

16.8 Conclusion

For better investigations into microbial forensic cases, lineage-based models will be required where population structures differ from the human models. Though other methods for statistical evaluations are available, they need to be authenticated with experimental genomic data results using control organisms. A database for molecular variations in similar strains of microbes should be made which can help in determining the origin and transmission root of the particular sample. As suggested by Ensernik and Ferber [27], at least three different strains of each pathogen and up to 20 for highly pathogenic species should be sequenced and recorded in form of a database. Developing countries should pay more attention toward collecting the genomic data and detailed information for pathogenic strains persisting in their environment. Partnerships within the labs nationally and internationally should be encouraged so that technology usage can be learned and exploited satisfactorily. Bioinformatics tools should also be applied to probe in microbial forensic cases.

References

1. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE (2005) Defining the normal bacterial flora of the oral cavity. J Clin Microbiol 43:5721–5732
2. Aggarwal P, Chopra AK, Gupte S, Sandhu SS (2011) Microbial forensic-an upcoming investigative discipline. J Indian Acad Forensic Med 33:163–165

3. Alexandre A, Laranjo M, Young JPW, Oliveira S (2008) *dnaJ* is a useful phylogenetic marker for alphaproteobacteria. Int J Syst Evol Microbiol 58:2839–2849
4. Ali MM, Shorky DA, Zaghlout HS, Rashed LA, Nanda MG (2013) PCR application in identification of saliva samples exposed to different conditions (streptococci detection based). Pak J Biol Sci 16:575–579
5. Aly R, Shirley C, Cunico B, Aibach HI (1978) Effects of prolonged occlusion on the microbial flora, pH, carbon dioxide and transepidermal water loss in human skin. J Invest Dermatol 71:378–381
6. Amann RI, Ludwig W, Schleifer KH, Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169
7. Bakhtiar SM, LeBlanc JG, Salvucci E, Ali A, Martin R, Langella P (2013) Implications of the human microbiome in inflammatory bowel diseases. FEMS Microbiol Lett 342:10–17
8. Braggemann H, Henne A, Hoster F, Liesegang H, Wiezer A, Strittmatter A (2004) The complete genome sequence of Propionibacterium acnes a commensal of human skin. Science 305:671–673
9. Brenig B, Beck J, Schutz E (2010) Shotgun metagenomics of biological stains using ultra-deep DNA sequencing. Forensic Sci Int Genet 4:228–231
10. Brenner S, Williams SR, Vermaas EH, Storck T, Moon K, McCollum C (2000) In vitro cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. Proc Natl Acad Sci USA 97:1665–1670
11. Brooke JS, Annand JW, Hammer A, Dembkowski K, Shulman ST (2009) Investigation of bacterial pathogens on 70 frequently used environmental surfaces in a large urban U.S. university. J Environ Health 71:17–22
12. Brüls T, Weissenbach J (2011) The human metagenome: our other genome. Hum Mol Genet 20(R2):R142–R128. https://doi.org/10.1093/hmg/ddr353
13. Budowle B, Chakraborty R (2004) Genetic considerations for interpreting molecular microbial forensic evidence. Int Congress Ser 1261:56–58
14. Budowle B, Schutzer SE, Einseln A, Kelly LC, Walash AC, Smith JA (2003) Public health, building microbial forensic as a response to bioterrorism. Science 301:1852–1853
15. Budowle B, Schmedes S, Murch RS (2013) The microbial forensics pathway for use of massively-parallel sequencing technologies in the science and applications of microbial genomics. The National Academies Press, Washington, DC, pp 117–133
16. Budowle B, Connell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J (2014) Validation of high throughput sequencing and microbial forensic applications. Investig Genet 5:9. https://doi.org/10.1186/2041-2223-5-9 eCollection 2014
17. Burney I (2013) Our environment in miniature: dust and the early twentieth-century forensic imagination. Representations (Berkeley) 121:31–59
18. Butler JC, Cohen ML, Friedman CR, Scripp RM, Watz CG (2008) Collaboration between public health and law enforcement: new paradigms and partnerships for bioterrorism planning and response. Emerg Infect Dis 8:1152–1156
19. Chiller K, Selkin BA, Murakawa GJ (2001) Skin microflora and bacterial infections of the skin. J Investig Dermatol SYP Proc 6:170–174
20. Cohn BA (1994) In search of human skin pheromones. Arch Dermatol 130:1048–1051
21. Costello EK, Lauber CL, Hamad M, Fiere N, Gordo JI, Knight R (2009) Bacterial community variation in human body habitats across space and time. Science 326:1694–1697
22. Cummings CA, Bormann Chung CA, Fang R, Barker M, Brzoska P, Williamson PC (2010) Accurate, rapid and high-throughput detection of strain-specific polymorphisms in *Bacillus anthracis* and *Yersinia pestis* by next-generation sequencing. Investig Genet 1:5
23. Das S, Dash HR, Mangwani N, Chakraborty J, Kumari S (2014) Understanding molecular identification and polyphasic taxonomic approaches for genetic relatedness and phylogenetic relationships of microorganisms. J Microbiol Met 103:80–100

24. Dominguez-Bello MG, Costello EK, Contrera M, Magri M, Hidalgo G, Fierer N, Knight R (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci USA 107:11971–11975
25. Eliaz PM (2006) The skin barrier formation and recovery in skin disorders. J Clin Invest 116:1150–1158
26. Eloe-Fadrosh EA, Rasko DA (2013) The human microbiome: from symbiosis to pathogenesis. Ann Rev Med 64:145–163
27. Enserink M, Ferber D (2003) Microbial forensic: reports spells out how to fight bio crimes. Science 299:1164–1165
28. Fierer N, Hamady M, Lauber CL, Knight R (2008) The influence of sex, handedness and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci USA 105:17994–17999
29. Fierer N, Lauber CL, Zhou N, Donald DM, Costello EK, Knight R (2010) Forensic identification using skin bacterial communities. PNAS 107:6477–6481
30. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR (1995) Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496–512
31. Gao Z, Tseng CH, Pei Z, Blaser MJ (2007) Molecular analysis of human forearm superficial skin bacterial biota. Proc Natl Acad Sci USA 104:2927–2932
32. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63
33. Goh MS, Gorton TS, Forsyth MH, Troy KE, Geary SJ (1998) Molecular and biochemical analysis of a 105 kDa *Mycoplasma gallisepticum* cytadhesin (GapA). Microbiol 144:2971–2978
34. Gribbon EM, Cunliffe WJ, Holand KT (1993) Interaction of *Propionibacterium acnes* with skin lipids in vitro. J Gen Microbiol 139:1745–1751
35. Grice EA, Serge JA (2011) The skin microbiome. Nat Rev Microbiol 9:244–253
36. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young CA (2009) Topographical and temporal diversity of the human skin microbiome. Science 324:1190–1192
37. Hill KK, Ticknor LO, Okinaka RT, Asay M, Blair H, Bliss KA (2004) Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. Appl Environ Microbiol 70:1068–1080
38. Horswell J, Cordiner SJ, Maas EW, Martin TM, Sutherland KBW, Speir TW (2002) Forensic comparison of soils by bacterial community DNA profiling. J Forensic Sci 47:350–353
39. Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774
40. Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci USA 96:3801–3806
41. Jain R, Rivera MC, Moore JE, Lake JA (2002) Horizontal gene transfer in microbial genome evolution. Theor Popul Biol 61:489–495
42. Jain S, Kumar A, Gupta P, Prasad R (2005) Microbial forensic: a new forensic discipline. JIAFM 27:112–116
43. Janda JM, Abott S (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, Perils and Pitfalls. J Clin Microbiol 45:2761–2764
44. Jarvis WR (1994) Handwashing- the Semmelweis lesson forgotten? Lancet 344:1311–1312
45. Keane FE, Ison CA, Taylor-Robinson D (1997) A longitudinal study of the vaginal flora over a menstrual cycle. Int J STD AIDS 8:489–494
46. Kitahara K, Miyazaki K (2013) Revisiting bacterial phylogeny: natural and experimental evidence for horizontal gene transfer of 16S rRNA. Mob Genet Elem 3(1):e24210
47. Korting HC, Hubner K, Greiner K, Hamm G, Braun- Falco O (1990) Differences in the skin surface pH and bacterial microflora due to long term application of synthetic detergent preparations of pH 5.5 and pH 7.0. Result of a crossover trial in healthy volunteers. Acta Derm Venereol 70:429–431
48. Kotewicz ML, Brown EW, LeClerc JE, Cebula TA (2003) Genomic variability among enteric pathogens: the case of the *mutS–rpoS* intergenic region. Trends Microbiol 11:2–6

49. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Sougnez C (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
50. Le Clerc JE, Li B, Payne WL, Cebula TA (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science 274:1208–1211
51. Leeming JP, Holland KT, Cunliffe WJ (1984) The microbial ecology of pilosebaceous units isolated from human skin. J Gen Microbiol 130:803–807
52. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380
53. Marples M (1965) The ecology of human skin. Bannerstone House, Springfield
54. Marshall JC (1999) Gastrointestinal flora and its alterations in critical illness. Curr Opin Clin Nutr Metab Care 2:405–411
55. Martin HL Jr, Richardson BA, Nyange PM, Lavreys L, Hillier SL, Chohan B (1999) Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. J Infect Dis 180:1863–1868
56. Martín R, Miquel S, Ulmer J, Kechaou N, Langella P, Bermúdez-Humarán LG (2013) Role of commensal and probiotic bacteria in human health: a focus on inflammatory bowel disease. Microb Cell Fact 12:71. https://doi.org/10.1186/1475-2859-12-71
57. Metzker ML (2010) Sequencing technologies-the next generation. Nat Rev Genet 11:31–46
58. Morse SA, Khan AS (2005) Epidemiologic investigation for public health, biodefense, and forensic microbiology. In: Breeze RG, Budowle B, Schutzer SE (eds) Microbial forensics. Academic, Amsterdam, pp 157–171
59. Munson MA, Pitt Ford T, Chong B, Weightman AJ, Wade WG (2002) Molecular and cultural analysis of the microflora associated with endodontic infections. J Dent Res 81:761–766
60. Muthiani YM, Matiru VN, Bii C (2010) Potential skin pathogens on second hand clothes and the effectiveness of disinfection methods, Nairobi, Kenya. In: Proceedings of 2010 JKUAT Scientific Technological and Industrialization Conference
61. Naser SM, Dawyndt P, Hoste B, Gevers D, Vandemeulebroecke K, Cleenwerck I, Vancanneyt M, Swings J (2007) Identification of lactobacilli by pheS and rpoA gene sequence analyses. Int J Syst Evol Microbiol 57(12):2777–2789
62. Nelly AN, Orloff M (2001) Survival of some medically important Fungi on hospital fabrics and plastics. J Clin Microbiol 39:3360–3361
63. Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734–740
64. Pennisi E (2004) Researchers trade insights about gene swapping. Science 305:334–335
65. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA (2009) NIH HMP Working Group. The NIH human microbiome project. Genome Res 19:2317–2323
66. Pittet D, Allegranzi B, Boyce J (2009) The World Health Organization guidelines on hand hygiene in health care and their consensus recommendations. Infect Contrl Hosp Epidemiol 30:611–622
67. Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS One 9(4):e93827
68. Priestley CJ, Jones BM, Dhar J, Goodwin L (1997) What is normal vaginal flora? Genitourin Med 73:23–28
69. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R (2008) A large genome center's improvements to the Illumina sequencing system. Nat Methods 5:1005–1010
70. Radnedge L, Agron PG, Hill KK, Jackson PJ, Ticknor LO, Keim P (2003) Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*. Appl Environ Microbiol 69:2755–2764
71. Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. Ann Rev Microbiol 57:369–394
72. Roth RR, James WD (1988) Microbial ecology of the skin. Ann Rev Microbiol 42:441–464
73. Salminen S, Isolauri E, Onnela T (1995) Gut flora in normal and disordered states. Chemotherapy 41:5–15

74. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
75. Sarkany I, Gaylarde CC (1968) Bacterial colonization of the skin of a newborn. J Pathol Bacteriol 95:115–122
76. Schleifer KH (2004) Microbial diversity: facts, problems, prospects. Syst Appl Microbiol 27:3–9
77. Schofield MJ, Hsieh P (2003) DNA mismatch repair: molecular mechanisms and biological function. Ann Rev Microbiol 57:579–608
78. Schouls LM, Schot CS, Jacobs JA (2003) Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. J Bacteriol 185:7241–7246
79. Sebastian T, Willem VW, Hubert PE, Alex VB, Kayser M (2010) Microbial DNA fingerprinting of human fingerprints: dynamic colonization of fingertip microflora challenges human host inference for forensic purposes. Int J Legal Med 124:477–481
80. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145
81. Shrivastava P, Jain T, Gupta MK (2015) Microbial forensics in legal medicine. SAS J Med 1:33–40
82. Smith SM, Eng RHK, Padberg FT (1996) Survival of nosocomial pathogenic bacteria at ambient temperature. J Med 27:293–302
83. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of E. coli. Nature 387:703–705
84. Staley JT, Konopka A (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Ann Rev Microbiol 39:321–346
85. Takashima M, Shirai F, Sageshma M, Ikeda OY, Dohi Y (2004) Distinctive bacteria binding property of cloth materials. Am J Infect Contrl 32:27–30
86. Treadwell TE, Koo D, Kuker K, Kahn AS (2003) Epidemiologic clues to bioterrorism. Public Health Rep 118:92–98
87. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. Nature 449:804–810
88. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE (2009) A core gut microbiome in obese and lean twins. Nature 457:480–484. https://doi.org/10.1038/nature07540
89. Ursell LK, Metcalf JL, Parfrey LW, Knight R (2012) Defining the human microbiome. Nutr Rev Suppl 70(Suppl 1):S38–S44. https://doi.org/10.1111/j.1753-4887.2012.00493.x
90. Venter JC, Adam MD, Myers EW, Li PW, Mural RJ, Zinder N (2001) The sequence of human. Genome Sci 291:1304–1351
91. Weinstock GM (2012) Genomic approaches to studying the human microbiota. Nature 489:250–256
92. Wetterstrand, K.S. (2013) DNA sequencing costs: data from the NHGRI Large-Scale Genome Sequencing Program. Available at: https://www.genome.gov/sequencingcosts/
93. Woo PC, Lau SK, Teng JL, Tse H, Yuen KY (2008) Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect 14:908–934
94. Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, Forney LJ (2004) Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. Microbiol 150:2565–2573

Chapter 17
Microbial Forensics: Beyond a Fascination

Vijay Nema

Abstract Microbiology has seen a great transition from culture-based identification of microbes using various biochemical and microscopic observations to identify and functionally characterize the microbes by just collecting the DNA and sequencing it. This advancement has not only moved in and around microbiology but has found its applications in fields which were earlier considered to be the remote ones. Forensics is one such field, where tracing the leftover evidence on a crime scene can lead to the identification and prosecution of the culprit. When leftover microbes in the biological material or objects used by the culprit or the person in question are used to correlate the identity of the individual, it takes us to the new field of science—"microbial forensics." Technological advances in the field of forensics, molecular biology, and microbiology have all helped to refine the techniques of collecting and processing of the samples for microbiological identification using DNA-based methods followed by its inference in the form of evidence. Studies have supported the assumption that skin or surface microflora of an individual is somewhat related with the microflora found on the objects used by that individual and efforts are ongoing to see if this is found consistently in various surroundings and with different individuals. Once established, this technique would facilitate accurate identification and differentiation of an individual or suspect to guide investigations along with conventional evidence. Legal investigations are not only the field where microbial forensic could help. Agriculture, defense, public health, tourism, etc. are the fields wherein microbial forensics with different names based on the fields are helping out and have potential to further support other fields.

Keywords Microbial forensics · Next-generation sequencing · Metagenomics · Microbiome · Culture-independent techniques · DNA

V. Nema (✉)
Division of Molecular Biology, National AIDS Research Institute, Pune, India

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_17

17.1 Introduction

Tracing the leftover evidences on a crime scene remained the only way of getting to the culprit. The history and evolution of forensic sciences has been very fascinating and interesting for even those who do not understand the science behind it. The journey started with William James Herschel sometimes in 1858 when he recognized that the fingerprints remain unique to the individuals. It was used on legal and administrative documents then and was published in 1916 with all evidences and their analysis [1]. It was gradually picked up for all possible uses including crime scene investigations. Fingerprints remained the landmark evidence in all forensic investigations and are still playing important roles. Additional techniques came in to support forensics in the form of DNA fingerprinting, wherein DNA matching became useful in various cases. DNA fingerprinting is used to establish a link between biological evidence and a suspect in a criminal investigation [2–6]. DNA—or genetic—fingerprinting relies heavily on the principle that no two individuals share the same genetic code.

Recently, there has been a new discipline co-emerged with culture-independent techniques of identifying microbes. This new discipline looks for microbes and tries to co-relate them with individuals as like fingerprints or DNA. There is a constant interaction of individuals with microbes in their surrounding, and they leave microbes into their surroundings. Whenever there is a physical contact, bacteria hop across from the skin to the material used. The microbial communities attached with an individual's skin or other sites are being explored, and preliminary evidences suggest that they are unique and can identify individuals or the material used by them in a few cases [7–9]. This became the base for coining of a new term "microbial forensics." However the field has other dimensions too and would be discussed in detail in this chapter. The microbial forensics rely on the inputs from various fields of basic and applied sciences. These include microbiology, genetics, bioinformatics, forensic science, immunology, population genetics, biochemistry, molecular biology, epidemiology, etc. along with the law enforcement, public health, policy, and intelligence communities.

17.2 What It Entails

Bacterial density on the human skin may be as high as 107 cells/cm^2 [10] and can be freely transferred to surface which comes in contact. Evaluating the traces of skin microbiome left on questioned objects may be useful for forensic identification. Calculating the distances from samples and their donators, it seems possible to estimate whether items or palm prints belong to one specific person or not. This has become possible because of the culture-independent method of tracing all microbial species present in a given environment. This technique used here is called next-generation sequencing, and the population of microbes no matter cultivable or non-culturable is called microbiome. In this process, total DNA of the given environment

or object is taken out, and the specific gene for prokaryotes, i.e., 16SrRNA gene, is amplified. This gene has a property like a clock with fast-moving arm and slow-moving arm. This concept has been explained by the work of Karl Woose who named this gene as a biological clock [11–14]. The gene contains highly conserved regions to identify the microbes with consistency and at the same time some variable regions to differentiate between closely related microbes. The technique has advantages over DNA from individuals as the microbiome DNA is abundant in touch DNA, and these organisms are much more stable because of complicated cell wall structures. Hence, it might be easier in certain instances to extract bacterial DNA than human DNA from surfaces target samples. Another good reason for this is the higher abundance of bacterial cells on the skin and shed epidermal cells as compared to human cells. There are some surfaces like fabrics, smudged surfaces, or highly textured surfaces from where obtaining clear fingerprints is very difficult. Bacterial DNA can still be found there and can help in solving the purpose [7].

17.3 Technical Terms and Their Explanation

Microbiology and forensic science were always considered to be different fields, and establishing a link between the two for better investigative power was not thought of earlier. Microbiology is simply defined as study of microorganisms wherein microorganisms are the organisms that exist as single cells or cell clusters and must be viewed individually with the aid of a microscope. Microbiology classifies microbes into various groups, and based on their characteristics and physiology, they could be assigned different genera and species. Most of the microbes, especially the saprophytic ones, depend on other substrates and interact specifically with their environment and habitat to gain important nutrients for their survival. This requirement remains very specific and unique with respect to one community of microbes and here comes a link with forensic sciences. Forensic science explores about those unique things (belongings or body remains) which could be co-related with the career (criminal) of those things, for instance, fingerprints or DNA. Because microbial communities on a particular individual's body are expected to be unique, identifying that individual with leftover microbial traces on the material used by the individual is a possibility and is being confirmed using various experiments. However, culturing of all those microbes and identifying them is a near impossible task. The reason is that all microbes do not grow on known culture media, and some which can grow are sometimes overtaken by the competition with other microbes in the same community. Hence collecting the total DNA and amplifying for the microbial signature gene is a good solution (Fig. 17.1). This technique is called metagenomics. Further, a sequencing technique that sequences the metagenome is called next-generation sequencing (NGS). NGS also known as high-throughput sequencing, represents different modern sequencing technologies for sequencing DNA and RNA much faster than Sanger sequencing and with lower economical and technical inputs.

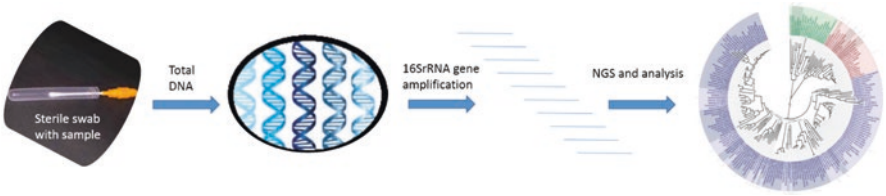

Fig. 17.1 The work flow of metagenomics from sample collected by swabbing a surface

17.4 How Does It Go Along with Other Techniques?

Identifying individuals is important in forensic science, and various developments have supported it from time to time. However, while making use of DNA, its quantity becomes crucial for accurate detection with desired quality for prosecution of crime. A smart offender can be cautious enough to decrease or degrade the traces of biological leftovers like blood, semen, etc. along with the fingerprints from the crime scene, which can complicate offender detection. Hence although fingerprints and DNA fingerprints remain very precise techniques, the availability of raw material becomes limiting factor in their use. Microbes can help in such scenario as their DNA is not as easily destroyed as a human DNA and remains available on surfaces at crime scene or on objects used. The pattern of bacterial DNA is dictated by the surrounding environment and the individual's microbiome [15, 16]. It is possible that the different bacterial patterns or the type of bacteria with typical physiology could discriminate individuals with different lifestyles. Hence, bacterial DNA analysis may serve as a complimentary technique in cases where standard DNA identification is partially informative [17]. Along with this, technologies like PCR, real-time PCR, MLST (multilocus sequence typing), MLVA (multilocus VNTR analysis), FISH (fluorescence in situ hybridization), and microarrays are being employed based on the need of the case or the availability of the infrastructure. Additional new methodologies like matrix-assisted laser desorption/ionization-time of flight (MALDITOF), gas chromatography-mass spectroscopy (GC-MS), and liquid chromatography-mass spectroscopy (LC-MS) are also well established in resolving minor difference in bacterial strains.

17.5 Scope of its Use

Investigation of a crime scene is not just potential application wherein microbial forensics can play an important role. Bioterrorism, biosecurity, biometry, medical forensics, etc. are the upcoming fields wherein definitive detection of microbes and their correlations may help a lot.

17.5.1 Criminology

Research on the transmission of microbes between human and surrounding environments has proved to be a property with potential for microbiome to be used in forensic investigations. In some cases, typical human microbial associations have been used to relate individuals to objects they have used [7]. Pattern of associated microbes with the surfaces at home and the family living and working with them have shown a predictive correlation to an extent that family's home and that individuals within a home can be differentiated [18]. Similar is the case with smartphones which rather remain in contact with specific individual and for longer durations [19]. Interestingly the microbial communities were different on the top and bottom of the phone differentiating the individuals and surface flora. An interesting property of the deposited microbiome is that it changes with a constant rate on a particular surface and can be utilized for forensic calculations. Work by Metcalf and co-workers have revealed that postmortem, the microbiome of animal hosts changes radically, but the pattern is much more predictable [20]. This can help tracing the time and direction of the events. Lax and co-workers worked to determine if the surface type and individuals have an effect on the microbial communities and have found that microbial community structure was determined both by surface type and participant.

Another aspect of microbial forensics is its role in bioterrorism. Herein the microbial forensics could be defined as "the discipline of applying scientific methods to the analysis of evidence related to bioterrorism, biocrimes, hoaxes, or the accidental release of a biological agent or toxin for attribution purposes" [21]. Microbial forensics, while dealing with bioterrorism, concentrate on identification of the agent or toxin and/or the mode of its production and dissemination. In addition, traditional forensic methods are used in conjunction to reach the goal of identifying the perpetrators of the crime. Around 14,000 microbial species or strains are listed as dangerous for humans [22]. Building individual diagnostic methods for these many numbers of agents is an impossible task, and hence NGS is the only tool to do massive parallel sequencing and to identify unknown pathogens, microorganisms modified to create panic, and pathogens in complex communities or samples with low abundance.

17.5.2 Agriculture and Medicine

With the global expansion of trade and communication, frequent movement of individuals from one country to other is unavoidable. In such scenario carriage of an endemic pathogen or drug resistance to a new geographical region is a looming threat. Recent outbreaks and transmission incidents of SARS and EBOLA have raised an alert. Microbial forensic has a crucial role in such cases wherein the status is not declared by individuals, is not detected by routine quarantine, or is a new

pathogen altogether. Detection of microbial drug resistance or the emerging resistance is becoming increasingly important for human health. Similar things are true in the case of plant pathogens and food borne diseases [23, 24].

The foodborne diseases have always been a substantial global challenge to public health. A huge population worldwide become sick of foodborne illnesses every year with a substantial burden on public health as well as on economy. Addressing this problem has many steps, out of which one age-old problem is the rapid identification of the food source of the contamination. In a classic laboratory study, we could trace back the source of the foodborne outbreak, but the finding could not be utilized in helping the troubled ones immediately [25]. This is due to the infrastructural limitations and technical challenges in identifying the pathogens. A technology which was considered to be reliable and was used till recent past was pulsed-field gel electrophoresis (PFGE). However its resolution in pinpointing the source of the outbreak has not been satisfactory. Recent employment of whole-genome sequencing (WGS) in such investigations has shown promise. A retrospective study by US Food and Drug Administration's Center for Food Safety and Applied Nutrition (FDA-CFSAN) in 2012 could provide a far better resolution of the causal factors. All the isolates were sequenced on the Illumina MiSeq. WGS using Illumina could distinguish all of the isolates which looked exactly the same. Ultimately they concluded that the isolates from the outbreak were most closely related to a 5-year-old historical isolate that was linked to a processing facility only 8 km away from the source of the outbreak [26]. This could not only allow newer findings but also traced back the source of contamination to further allow the rectification. Such cases are sometimes accidental but are often criminal too, and tracing back the source would prevent such cases. In bio crimes, serious disease outbreak by natural occurrence or intentional may result in harm or death, causing disruption, creating fear, and affecting economic well-being. Microbial forensics thus plays an important role in consumer protection, food security, and even in litigation.

Agriculture and agricultural goods are also the susceptible area for microbial interventions and hence are important in terms of microbial forensics. There could be deliberate misuse of microbes or their products affecting flora and fauna which is important for agriculture. This could be given a name such as "agroterrorism" [27]. A field of investigation emerged against this threat for investigating into the violations and used scientific knowledge and technology to do so. This was given the name of bioforensics [28].

Another aspect of microbial forensic application to the foodborne pathogens is to trace pathogens in cash crops especially spices and other costly ones. Van Doren and co-workers have studied 14 reported illness outbreaks from Canada, Denmark, England and Wales, France, Germany, New Zealand, Norway, Serbia, and the United States which occurred due to consumption of pathogen-contaminated spice during 1973–2010. The outbreaks were reported from a few developed countries only. The reason for not including other countries was that those countries did not have updated technology to investigate and report similar findings. It was reported that these outbreaks resulted in 1946 human illnesses, 128 hospitalizations, and 2

deaths. Infants/children were the primary population segments impacted by 36% (5/14) of spice-attributed outbreaks [29]. The economic aspect associated here is the detection of pathogens after shipment and then the recall of the material. This involved a huge cost.

Recent development in microbial forensics in agricultural sciences also aids in pest control as well as deliberate introduction of pests along with food imports or use of pathogens as anticrop bioweapons. Different companies are coming up with molecular detection tools for rapid detection of specific pathogens in such products. For instance, Hu and co-workers have compared and evaluated the effectiveness of the molecular methods (3M Molecular Detection System (MDS) and ANSR Pathogen Detection System (PDS)) for the detection of *Salmonella* in egg products and compared the same with culture methods to find that the molecular methods are the superior and faster ones [30]. Budowle and co-workers have established a criterion comprising a foundation for investigators to establish, validate, and implement high-throughput sequencing (HTS) as a tool in microbial forensics [31]. Likewise, design principles for an effective microbial forensics program for law enforcement and national security purposes have been provided [32].

Microbial pathogens or toxins can be used to commit acts of terror; they can be used as weapons for execution of a crime. In biological warfare, transmissible lethal agents are used to attack the targeted populations. The impact of the bioterrorism was seriously considered after the anthrax attack in the United States in 2001. This incidence in the United States helped the world understand that bioterrorism can have drastic and global impacts. Microbial forensics has a role in such cases by applying scientific methods for the analysis of evidence from such a bioterrorism attack. Microbial forensic in conjunction with epidemiology can try to decipher if an outbreak is natural, accidental, or intentional [33, 34]. For instance, study by Price and co-workers found that the *Bacillus anthracis* injectional anthrax cases were originated from heroin users in Scotland [35]. Ou and co-workers used molecular tracking of HIV to report for the first time about the passage of HIV infection from dentist to patient after invasive healthcare procedure [36]. A Spanish anesthetist infected 275 patients with hepatitis C virus which could be found using phylogenetic and molecular clock analysis [37]. In the 2014 Ebola outbreak, the origin and transmission could be traced using bioforensic methods [38], etc. Other important and relatively new aspect of microbial forensics in medicine and medicolegal field is "thanatomicrobiome" (*thanatos*—death) that studies the microorganisms found in internal organs and cavities upon death. The thanatomicrobiome tries to investigate the total microbial communities including bacterial and fungi from all the body locations of decomposing corpses. These studies are important in providing evidence in medicolegal death investigations [39]. By doing this, the concept of human postmortem microbiome project (HPMP) has also been introduced which would create a consortium of research projects to identify and characterize the thanatomicrobiome and epinecrotic communities (e.g., epithelial tissues, body cavities, and the alimentary canal), relating to human decomposition with a potential of finding a state-of-the-art, more dependable, and molecular way of determining the time of death [40].

17.6 Potential to Be an Independent Field: Emerging Evidence

It has been shown that typical characteristics and pattern of the human microbiome might identify individuals and remain constant in the individual but differ from the others [7, 41, 42]. This means that individuals might be specifically and consistently identified using their microbiome. However microbiome-based identifiability is still a long way to go. Franzosa and co-workers have suggested a few means to achieve the target of identifiability using microbes. The identification of a "metagenomic code" that remains unique for an individual for a longer duration of time and stands true for a sizable population is the key [43]. Hence microbiome establishment, structure, personalization, and temporal stability are the standard terms and areas to work upon.

There are some untraveled avenues which start from microbial forensics. Emergence of antimicrobial resistance, prediction and prevention of future outbreaks, etc. are some of the fields wherein a surveillance system making use of the principles of microbial forensics can help. Antimicrobial resistance in microorganisms emerges naturally. However, antimicrobial exposure due to human practices in healthcare, agriculture, sanitation, industrial processes, travel, and other fields contributes significantly. Timely detection of pathogens harboring resistance can mitigate the onward transmission among individuals and among different geographical locations. Timely detection of pathogens like HIV, severe acute respiratory syndrome (SARS) virus, and pandemic influenza could have avoided the big health emergencies which we have witnessed in recent past. Similar is the case with the pandemic spread of SARS coronavirus in 2003 and H1N1 influenza in 2009 resulting in substantial economic loss. Microbial forensics can play a crucial role in such cases by checking the emergence in real time and suggesting measures to prevent transmission. However, a lot of investment is required to place such services at all vulnerable points or check points. Technological advances especially in the field of metagenomics have paved way for identification of potential human pathogens among other species and have attained good predictive power regarding transmissibility and virulence of the novel microbes.

17.7 Limitations

The science of microbial forensics is in its infancy and needs much more that what has already been done. A few things which needs to be taken care as preparation before we take this as a routine science are protocols and procedures for collecting specimens at the attack site, recognizing that an attack is occurring and diagnosing the disease, analysis of specimens in contained facilities, quality assurance and control. Although microbial forensics involve techniques or methodologies from basic laboratory sciences, the problems in question, processes engaged in and expected outcomes need more than that. There are efforts and continuous need for validating

all the tools and techniques involved which should be acceptable to peers and stakeholder from scientific, legal, and policy making side. The optimization of methods to answer key questions pertaining to investigative and legal needs is a must to satisfy the criterion of acceptability. Meeting these challenges will allow the establishment of a complementary and reliable method to compensate for the lacuna of DNA fingerprinting and fingerprinting as discussed earlier.

Meeting the challenges needs the consorted efforts from global communities of workers from basic sciences, epidemiologists, forensic experts, medical experts, legal experts, and a big team of technology developers. The most reliable technique till date for microbial forensics is metagenomics—a culture-independent approach for identifying and enumerating microbes. Metagenomic have been an outstanding technique for sequencing the genomes of unculturable microbes, which represent the vast majority of microorganisms, particularly from environmental samples. However, technological advancement identifies the rare taxon is awaited.

As microbial metagenomics is undergoing a formative phase as a diagnostic technique, optimization of methods and their validation remain a challenge. Other aspects of this are the leadership role and generic availability of tools and techniques. Countries with major resources would be able to take lead in basic research, while the resource-limited setting may not be able to adapt microbial forensics owing to its monitory needs. Availability of equipment and techniques for rapid and precise molecular diagnosis is important for controlling and responding to the needs of microbial forensics. Till date, next-generation sequencing is the only technology that seems promising for microbial forensics. But the instrument and the reagents remain very costly as compared to biochemical tests and a few basic molecular assays being used in forensic laboratories. Moreover, expert workers and bioinformatics analysis of the huge data generated after massive parallel sequencing or next-generation sequencing require a devoted facility and expertise.

17.8 Conclusion

Microbial forensics may, in most of the cases, be associated with the detection of causal pathogen in cases of biological terrorism, but microbiome associated with individuals and objects used or touched by them at crime scenes may also be used as a tool in providing forensically relevant information. This science has got its diverse utility in the field of medicine, agriculture, trade especially in food articles, etc. as discussed in this chapter. The evidence and the literature available till date indicate a definitive linkage between an individual and microbial communities inhabiting that individual skin or other body parts. The science is progressing toward identifying these communities, next steps of which would be to see if these communities have something in common and that common character has something to do with the individual. Later work in this field may identify chemical entities and their microbial connections to land up to the final conclusions about signature communities of microbes and their forensic potential. However, we may have to continue our search to see those days soon.

References

1. Herschel WJ (1916) The origin of fingerprinting. Humphrey Milford Oxford University Press, London
2. Dixson AF, Hastie N, Patel I, Jeffreys AJ (1988) DNA 'fingerprinting' of captive family groups of common marmosets (Callithrix jacchus). Folia Primatol (Basel) 51(1):52–55
3. Gill P, Jeffreys AJ, Werrett DJ (1985) Forensic application of DNA 'fingerprints'. Nature 318(6046):577–579
4. Helminen P, Ehnholm C, Lokki ML, Jeffreys A, Peltonen L (1988) Application of DNA "fingerprints" to paternity determinations. Lancet 1(8585):574–576
5. Jeffreys AJ (2013) The man behind the DNA fingerprints: an interview with professor sir Alec Jeffreys. Investig Genet 4:21
6. Wetton JH, Carter RE, Parkin DT, Walters D (1987) Demographic study of a wild house sparrow population by DNA fingerprinting. Nature 327(6118):147–149
7. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R (2010) Forensic identification using skin bacterial communities. Proc Natl Acad Sci U S A 107(14):6477–6481
8. Goga H (2012) Comparison of bacterial DNA profiles of footwear insoles and soles of feet for the forensic discrimination of footwear owners. Int J Legal Med 126(5):815–823
9. Lax S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, Gilbert JA (2015) Forensic analysis of the microbiome of phones and shoes. Microbiome 12(3):21
10. Fredricks DN (2001) Microbial ecology of human skin in health and disease. J Investig Dermatol Symp Proc 6:167–169
11. Morell V (1997-05-02). Microbiology's scarred revolutionary. Science 276(5313):699–702
12. Woese CR, Fox GE (1977-11-01) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci 74(11): 5088–5090
13. Woese CR, Magrum LJ, Fox GE (1978) Archaebacteria. J Mol Evol 11(3):245–251
14. Woese CR, Kandler O, Wheelis M (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A 87(12):4576–4579
15. Lee SY, Woo SK, Choi GW, Hong HJ, Eom YB (2015) Microbial forensic analysis of bacterial fingerprint by sequence comparison of 16S rRNA gene. J Forensic Res 6:297
16. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G et al (2013) Cohabiting family members share microbiota with one another and with their dogs. elife 2:e00458
17. Leake SL (2013) Is human DNA enough?-potential for bacterial DNA. Front Genet 4:282
18. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science 2014;345(6200):1048–1052
19. 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
20. Metcalf JL, Wegener Parfey L, Gonzalez A, Lauber CL, Knights D, Ackermann G et al (2013) A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. elife 2:e01104
21. Schmedes SE, Sajantila A, Budowle B (2016) Expansion of microbial forensics. J Clin Microbiol 54(8):1964–1974
22. Taylor LH, Latham SM, Woolhouse MEJ (2001) Risk factors for human disease emergence. Philos Trans R Soc Lond Ser B Biol Sci 356:983–989
23. Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R (2016) Practical value of food pathogen traceability through building a whole-genome sequencing network and database. J Clin Microbiol 54(8):1975–1983
24. Chen Y, Luo Y, Pettengill J, Timme R, Melka D, Doyle M, Jackson A, Parish M, Hammack TS, Allard MW, Brown EW, Strain EA (2017) Singleton sequence type 382, an emerging clonal group of Listeria monocytogenes associated with three multistate outbreaks linked to contaminated stone fruit, caramel apples, and leafy green salad. J Clin Microbiol 55:931–941

25. Nema V, Agrawal R, Kamboj DV, Goel AK, Singh L. Isolation and characterization of heat resistant enterotoxigenic Staphylococcus aureus from a food poisoning outbreak in Indian subcontinent. Int J Food Microbiol 2007 10;117(1):29–35
26. Hoffmann M, Luo Y, Monday SR, Gonzales-Escalona N, Ottensen A, Muruvanda T, Wang C, Kastanis G, Keys C, Janies D, Senturk I, Catalyurek UV, Wang H, Hammack TS, Wolfgang WJ, Schoonmaker-Bopp D, Chu A, Myers R, Haendiges J, Evans P, Meng J, Strain E, Allard MW, Brown EW (2016) Tracing origins of the Salmonella Bareilly strain causing a food-borne outbreak in the United States. J Infect Dis 213:502–508
27. Latxague E, Sache I, Pinon J, Andrivon D, Barbier M, Suffert F (2007) A methodology for assessing the risk posed by the deliberate and harmful use of plant pathogens in Europe. EPPO Bull 37:427–435
28. Suffert F. (2017) Characterization of the Threat Resulting from Plant Pathogen Use as Anti-crop Bioweapons: An EU Perspective on Agroterrorism. In: Gullino M., Stack J., Fletcher J., Mumford J. (eds) Practical Tools for Plant and Food Biosecurity. Plant Pathology in the 21st Century, vol 8. Springer, Cham
29. Van Doren JM, Neil KP, Parish M, Gieraltowski L, Gould LH, Gombas KL (2013 Dec) Foodborne illness outbreaks from microbial contaminants in spices, 1973-2010. Food Microbiol 36(2):456–464
30. Hu L, Ma LM, Zheng S, He X, Wang H, Brown EW, Hammack TS, Zhang G. Evaluation of 3M Molecular Detection System and ANSR Pathogen Detection System for rapid detection of Salmonella from egg products. Poult Sci. 2017; 96(5):1410–1418
31. Budowle B, Connell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, Forsman M, Kadavy DR, Markotic A, Morse SA, Murch RS, Sajantila A, Schmedes SE, Ternus KL, Turner SD, Minot S (2014) Validation of high throughput sequencing and microbial forensics applications. Investig Genet 30(5):9
32. Murch RS (2014) Designing an effective microbial forensics program for law enforcement and national security purposes. Arch Immunol Ther Exp 62(3):179–185
33. Keim P, Budowle B, Ravel J. 2011. Microbial forensic investigation of the anthrax-letter attacks, Budowle B, Schutzer SE, Breeze RG, Keim PS, Morse SA, Microbial forensics, 2nd. Academic, Amsterdam, 15–25
34. National Research Council (2011) Review of the scientific approaches used during the FBI's investigation of the 2001 anthrax letters. The National Academies Press, Washington, DC
35. Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, Mason J, Vincent G, Drees KP, Beckstrom-Sternberg SM, Phillippy AM, Koren S, Okinaka RT, Chung W-K, Schupp JM, Wagner DM, Vipond R, Foster JT, Bergman NH, Burans J, Pearson T, Brooks T, Keim P (2012) Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe. Emerg Infect Dis 18:1307–1313
36. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins JI, Schochetman G, Berkelman RL, Economou AN (1992) Molecular epidemiology of HIV transmission in a dental practice. Science 256:1165–1171
37. Schloissnig S et al (2013) Genomic variation landscape of the human gut microbiome. Nature 493(7430):45–50
38. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S, Moses LM, Yozwiak NL, Winnicki S, Matranga CB, Malboeuf CM, Qu J, Gladden AD, Schaffner SF, Yang X, Jiang P-P, Nekoui M, Colubri A, Coomber MR, Fonnie M, Moigboi A, Gbakie M, Kamara FK, Tucker V, Konuwa E, Saffa S, Sellu J, Jalloh AA, Kovoma A, Koninga J, Mustapha I, Kargbo K, Foday M, Yillah M, Kanneh F, Robert W, Massally JLB, Chapman SB, Bochicchio J, Murphy C, Nusbaum C, Young S, Birren BW, Grant DS, Scheiffelin JS, et al. 2014. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 345: 1369–1372
39. Javan GT, Finley SJ, Abidin Z, Mulle JG (2016a) The Thanatomicrobiome: a missing piece of the microbial puzzle of death. Front Microbiol 7:225

40. Javan GT, Finley SJ, Can I, Wilkinson JE, Hanson JD, Tarone AM (2016b) Human Thanatomicrobiome succession and time since death. Sci Rep 6
41. Faith JJ et al (2013) The long-term stability of the human gut microbiota. Science 341(6141):1237439
42. González-Candelas F, Bracho MA, Wróbel B, Moya A (2013) Molecular evolution in court: analysis of a large hepatitis C virus outbreak from an evolving source. BMC Biol 11:76
43. Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP, Bohannan BJ, Huttenhower C (2015) Identifying personal microbiomes using metagenomic codes. Proc Natl Acad Sci U S A 112(22):E2930–E2938

Chapter 18
Implications of Microbes in Forensic DNA Fingerprinting

Pankaj Krishna

Abstract It was the human genomic and mitochondrial DNA which was used in DNA fingerprinting in forensic case studies. However, many of the time, the human DNA is not available for analysis due to many reasons, so additional biomarker may be of immense help for forensic scientists. The microbiome analysis appears to hold that promise, and several research publications are suggestive of the use of microbial fingerprint to link an individual with the object. The microbiome which is basically microbes harbouring humans is analysed using DNA tools. In microbiome analysis, the samples are subjected to DNA extraction, amplification, parallel sequencing of the 16S rRNA region and analysis. Many studies showed that the individual person has a unique microbial fingerprint which makes them distinct in the population. The microbial population could be studied to understand the geographic location, post-mortem interval application, nature of transmission and cause of death including the establishing individual identity.

Keywords Microbiome · Markers · DNA · 16S rRNA gene · Systematics · Metagenomics · Human Microbiome Project

18.1 Introduction

Microorganisms, which basically consist of bacteria (which are further classified phylogenetically into Bacteria and Archaea domains), viruses, prions, protists and fungi, are vital to the function of all ecosystems [14]. Eukarya is believed to be more closely related to Archaea [19]. The diverse microbes exist abundantly (estimated bacteria is 5×10^{30} worldwide) and so have massive activity, function and accumulative mass [53]. The coexistence of animals and microbes are estimated since a billion years ago [6]. While the question remains about the domestication of microbes in animals, importantly after the Human Microbiome Project (broadly known as International Human Microbiome Consortium (IHMC)), efforts are being made to

P. Krishna (✉)
Avantor Performance Materials India Limited, Dehradun, India
e-mail: pankaj.krishna@avantormaterials.com

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1_18

understand the human inhabited microbes for host interaction, health and survival. However, according to recent estimation, the humans are remarkably similar at the phylum level, while at the genus, species and strain level of the given population, the diversity is profoundly specific for each individual [9, 13, 15, 18, 30].

The bacterial characterization was done historically by describing its phenotypic properties, such as morphological structure, cultivational and nutritional requirements, carbon source utilization, biochemical metabolism, physiology, pathogenicity, growth condition, antigenic properties and its ecosystem [16]. These approaches have been useful to extract information on reasonable scale diversity but are biased due to limited information on media, cultivation condition and growth strategy of bacteria. Hence, a large but significant population of natural microbial communities could not be isolated in laboratory condition limiting their growth and isolation [43]. The general estimates predict that more than 99% of the naturally occurring prokaryotes (bacteria within the domains Bacteria and Archaea) present in many natural environments are not readily culturable; it indicates that only 1% bacteria are culturable and is available for our knowledge and accessible for further basic and applied research [2]. This laid emphasis on an alternative method to identify and access the microbes without isolating in laboratory conditions.

Several classical methodologies are being employed for calculation and assessment of microbial diversity [3]. Another approach was to directly observe the morphologically distinct groups, but this method has serious limitation for the analysis and degree of diversity required [43]. The potential limitation and biasness associated with cultural methods are widely accepted as the cultural methods reveal only those physiological, nutritional, growth and metabolism information which are compatible with the laboratory cultural environment [48]. The great diversity of the microbial flora makes it extremely difficult to determine accurately the microbial population quantitatively. More significantly, designing of growth media and devising conditions to cultivate the uncultivable microbes seems to be distant possibility in near future.

Naturally existing microbes could be analysed by molecular biological techniques with little or no bias for its structure, community and species composition [39]. The genes are appeared to have far more attractive and accurate concept element than the species for measuring (micro) structural and functional biological variety, but it is not devoid of problem and bias [51]. The phylogenetic relationships among microorganisms were inferred based on the sequence variation in rRNA genes [55]. Analysis of the 16S rRNA nucleotide base sequence of bacterial DNA makes it possible to construe phylogenetic relationships among bacteria and community structural pattern in an ecosystem. The cistrons particularly of 16S rRNA have changed more slowly and remained conserved during evolution than the bulk of the bacterial genome, only to allow them to be used as molecular/evolutionary indicators to check the relatedness of the species. The 16S rRNA cistrons, present in all bacteria, could be used as biomarker for the speciation and diversity analysis and is useful tracer in evolutionary biology [55].

The culture-independent genomic analyses of microbial communities are called as metagenomics. Metagenomics can be used to analyse the environment where the

uncultivable microbes are 99% with little bias [2]. This approach builds on advances in microbial genomics by employing direct isolation of nucleic acid, polymerase chain reaction (PCR) amplification, size-based separation, cloning of genes and short and parallel sequencing directly from environmental samples [41, 46]. The isolation of nucleic acid from an environmental sample bypasses culturing the organisms in laboratory condition, and cloning of its cistrons into a cultured organism could be used for function- and structure-driven biotechnology study. The metagenomics analysis is categorized as function- and sequence-driven approaches. In function-driven approach, the directly isolated nucleic acid is converted into metagenomics library and screened for expressed proteins or traits. But in sequence-driven approach, the similarities of a particular DNA sequence are screened for to establish the identity and structure of the community. Further refinement in function-driven approach especially for the candidate genes specific to function may lead the accelerated discovery of useful molecules [57].

The classical species information is less meaningful than the gene sequences from one environment. Metagenomics or microbial genomics is a tool that hunts for habitat than the species and is a right complementary tool to analyse "organismal genomics" [45].

18.2 Markers in Molecular Biology

The rRNA provides a dominant mass in ribosomes which are made up of ribonucleoprotein particle. The 70S bacterial ribosomes have been studied in great detail; however, it believed that shapes of all 70S ribosomes are similar. The secondary structure and sequences are used for comparison and to dig out the conserve features of rRNA in a variety of organisms [29]. Now data from the sequence analysis of the rRNA gene from many organisms have shown that common sense is a fallible guide for understanding the relationship between organisms. 16S rRNA is now popular with molecular systematics because this class of RNA is found in all organisms and the molecules are sufficiently long about 1660 nucleotides to allow accurate interspecies comparisons.

For ten million ribosomes, an estimated ten million copies of each type of rRNA molecule are required in growing higher eukaryotes in an each cell generation. A cell contains multiple copies of the rRNAs which are required to produce adequate quantities of rRNA. Seven copies of rRNA genes are found in *E. coli* to meet the cell needs for ribosomes. Unlike the prokaryotes the eukaryotes carry the multiple copies of the highly conserved rRNA genes which are tandemly arranged on a chromosome and are separated by a nontranscribed region known as spacer DNA which can vary greatly in length and sequence (8000–13,000 nucleotides pairs long, depending on the organism). The tandemly arranged multiple copies of these genes are likely

to co-evolve [31, 36, 54]. “The rRNA genes are transcribed by RNA polymerase I and each gene produces the same primary RNA transcripts” [1].

The 18S rRNA of fungus on which little fungal diversity has been reported possesses limited information as compared to 16S rRNA gene of prokaryotes. The internal transcribed spacer (ITS) is analysed for greater specificity in the analysis of fungal diversity [50]. The ITS are noncoding regions of DNA sequence that are tandemly arranged and separate genes coding for the 28S, 5.8S and 18S ribosomal RNAs. These ribosomal RNA (rRNA) genes are preserved over the taxa, while the spacers between them may be species-specific. The conserved nucleotides of the rRNA genes have complementary region for the “universal” primers for the easy amplification by the polymerase chain reaction. The spacers with variation have demonstrated value for recognizing among a wide variety of hard-to-distinguish and hard-to-identify taxa. The ITS region is relied upon to indicate intraspecific variety over expansive scale since (1) it changes quickly, and (2) a species population can be reproductively confined at huge scales. Like 16S rRNA gene of bacteria, ITS region of fungi is most broadly sequenced region [28].

It has ordinarily been the most valuable for molecular systematics at the species level and even within species (e.g. to distinguish geographic races). Variation among individual rDNA repeats can in some cases be seen inside both the ITS and IGS regions on account of its higher degree of variation than different regions of rDNA. Investigation of 16S rRNA gene is currently broadly utilized for examination of bacterial population, and investigation of 18S genes and internal transcribed spacer (ITS) regions is progressively being utilized for the fungal identification and community structure analysis. Direct extraction of DNA from naturally occurring environment and amplification of 16S rDNA has provided precious information of unculturable population and understanding of the ecosystem and its interaction with the microbes. Many primers sequences related to SSU-rRNA for the PCR amplification and gene sequencing have been published since the past 25 years [10, 12, 26, 52].

Some primers were designed to amplify the taxa, while others were not specific enough for taxa but intended to amplify all prokaryotic SSU-rRNA genes and are denoted as “universal” [4, 24, 34, 40].

The DNA extraction yield is associated with certain bias in the molecular technique used. The first step of molecular technique for DNA extraction is the lysis whose efficiency varies between different bacterial, fungal and microbial groups. It also varies between spores, mycelia, age and regions of mycelia.

Regardless of the bacterial genome, the 16S rRNA imparts information of just 0.05% on the average of the total prokaryotic genome/cell and has very limited value to predict the functions of the cell such as physiology, niche, biotechnological properties and other phenotypic characters [44]. Numerous 16S rRNA genes of the same species have indistinguishable or demonstrate just minor contrasts. The reality of the matter is that the databases of 16S rRNA genes are tremendous and that fundamentally the same sequences are recovered over and over [20] demonstrating that possibly this approach is achieving plateau, in the most considered biotypes.

18.3 Deep Sequencing

A new direction was attained in the analysis of the microbial diversity analysis after advent of the next-generation sequencing technology. It was feasible to do the rapid sequencing directly from the PCR amplicon. Many platforms were launched for massively parallel DNA sequencing and are available commercially: 454 (Roche), Illumina (Illumina), SOLiD (Life Technologies), HeliScope (Helicos), Ion Torrent (Life Technologies) and PacBio (Pacific Biosciences). In fact, few of these platforms are discontinued today and no more available for sale. Each platform typifies the complex relationship of enzymology, molecular biology, chemistry, physics (high resolution optics), hardware and software engineering [33]. This next-generation sequencing is based either on synthesis or ligation where read length varies from 35 to 1100 bp. These machines are either based on single molecule or based on emulsion/bridge PCR for amplification [17]. Nowadays, Illumina and Ion Torrent are the popular choice for microbiome research. van Dijk et al. [49] have reviewed and given a brief overview of both machines as "Ion Torrent: semiconductor sequencing very similar to 454 sequencing except that, instead of pyrophosphate, proton release during nucleotide incorporation is detected using ion sensors; no imaging technology is required. After incorporation of a nucleotide, which is identified by its color, the 30 terminator on the base and the fluorophore are removed, and the cycle is repeated".

18.4 Human Microbiome

The understanding of the human biology is incomplete without the study of its associated surrounding environment. It is widely believed that the microbes and human have billion years long association and microbes have modulating effect on host immunological and other responses. The microbiome is characterized as the aggregate genomes of the organisms (made out of microorganisms, virus, bacteriophage, protozoa and fungi) that inhabit inside and on the human body. With the advent of the high-throughput sequencing, it was easier to sequence the microbial communities harbouring human without culturing them. Taking advantage of high-throughput sequencing, studying the microbiome of various human sites and their modulating impact, the Human Microbiome Project (HMP) was incepted and initiated by NIH [25].

The HMP and other studies showed that there is greater degree of variability in the microbial population at different sites of the same person and inter-person analysis. Many epidemiological factors such as past history, life style, the surroundings and the occupation have impact on the composition and structure of microbial community of a person. Clarke et al. [8] cited that a person's lifestyle can be a major driver of the composition of the microbiome communities across many body sites, including what they eat [11], with whom they live [47], whether they have pets [37], where they are from [56], if they smoke [5, 38], their health status [7] and whom they kiss [27].

Humans have more microbes than their own cells and can be used as forensic analysis. Microbial fingerprint of humans are unique and are used to as a forensic tool. The two examples shown as case studies in the beginning of this chapter are to understand the potential implications of the microbes as a forensic tool. In the first case study, the phylogeny and relatedness of HIV-1 is analysed, and the second case study discusses about the unique microbial fingerprint of a person which is being used for the linking of a person with his object. The potential ability of this microbial fingerprint for linking specific human subject to an object has created a new possibility to explore and exploit this new branch of science. There are many similar studies carried out which discusses about the microbial fingerprint in several niche environment such as shoes, skin, hair, cell phone, rooms, bathrooms, offices and classroom and even used to predict information about a person interacting with the environment, such as the sex of the inhabitant of dorm rooms [32]. The temporal studies were conducted to understand the changes in the microbial community structure, and it was confirmed that the microbial fingerprint is detectible for up to several months.

Gunn and Pitt [19] in their review paper have indicated the potential microbes as forensic indicators for the – post-mortem toxicology, identification of the individual, as a cause of death, transmission (hospital acquired, natural, sexual and reckless transmission) and body fluid. There are microbial forensic studies which include the identification of personal identity, country of origin and time of death by using microbial signature analysis from the sample collected from the diverse locations such as the hair, skin and vagina [8].

In absence of the human DNA evidence, the microbiome left behind by the suspect on a victim or object could be useful evidence in forensic investigation. It was possible to identify a single individual in large population and to link a person based on skin, hair and vaginal microbiome to the victim/object which was touched based on microbial fingerprint. These studies indicate the discrimination power based on microbial fingerprint and link them to sources of potential physical evidence. Remarkably, the power of microbial fingerprints for identification purposes extends to ethnicity [8]. Clarke et al. [8] further suggested the applications of forensic microbial applications in deciphering the geolocations and use in post-mortem interval estimation. Due to varied environmental condition, the microbial population differs at different geographical locations. The signature microbiome population of a host may help in identifying its geographical location and to check the human trafficking and to identify nonindigenous disease outbreak [8, 21].

The microbiome data is preliminary and not ready for the use in forensics, and there are several challenges associated with the use of microbiome in forensic analysis from the time of sampling to analysis. The sampling in itself is crucial which may be affected with the time of collection and depth of sequencing and temporal variations. However, it is believed that the storage may not have greater impact on the analysis of the microbiome. Analysis also requires application of right statistical model, determining the quality score and rare taxa present in the sample. Molecular technique has also bias in terms of extracting DNA which largely associated with the lysis efficiency which varies due to the cell wall composition between the different

microbial groups. The amplification part has bias in terms of primer designing and its specificity to amplify the representative population. One bacterium may have different copy number of the same gene which may lead to the overestimation of the population. In any case the 16S rRNA gene represents only a 0.05% on the average of the genome of a prokaryotic cell, and the next-generation sequencing applies amplification and sequencing of around 200 bps. There is requirement for quality control (QC) and quality assurance (QA) measures, like those utilized by medicinal research centres and laboratories [42] to check the trueness and precision of the bacteriological outcomes [19].

18.4.1 Case I

State of Louisiana vs. Richard J. Schmidt was a case in view of circumstantial evidence showing that gastroenterologist Dr. Richard J. Schmidt on August 4, 1994, made a blend of blood or blood items from two patients under his care, one contaminated with HIV-1 and the other with hepatitis C, and injected his ex-girlfriend by intramuscular infusion. This case is interesting because of the fact that the one part of the criminal investigation was supported by molecular phylogenetic analysis of HIV viruses and was the first occasion when that phylogenetic molecular analysis has been utilized as a confirmation in a US criminal proceeding [35].

Dr. Schmidt, 48-year-old, was a gastroenterologist doctor of repute, married, with three children. He had a thriving practice at Lafayette General Hospital and was popular with the patients. Janice Trahan, a registered nurse, married, with an infant son, joined the same hospital and fell in love with Schmidt. They were in a relationship for a decade. The affair may be termed as turbulent marked with four abortions but she bore him a son. Frustrating with not keeping his promises and not seeing the future of their relationship, Trahan decided to break-up. Even after the break-up, they were seeing each other, Schmidt used to give her B-12 shots to boost energy. On one humid night, when Trahan was sleeping, Schmidt appeared with a hypodermal needle to give her the B-12 shot. She felt a sharp needle prick despite her protest (Newsweek; People.com).

In January 1995, the pregnant Trahan developed some symptoms and visited her doctor which in turn proved to be no good, and she was declared HIV-1 infected. Going through the trauma, she could relate the only one incidence the last B-12 shot. She has undergone the abortion and approached the police and accused the doctor of injecting the deadly viruses (Newsweek; People.com). A microbial forensic investigation began which involved the epidemiological factors such as lifestyle, occupational, nonoccupational, local population and past health records. Trahan has reported having sexual contacts with the doctor and other six men, but none of them were tested to be positive for HIV-1. At many instances, she has donated blood to the blood bank and tested negative for HIV infection. To support the hypothesis of injecting the HIV-infected blood in the victim, the police has identified the patient infected with HIV-1 whose blood sample collected in the doctor's office was

believed to be injected in Trahan. The patient was a homosexual male and was a teacher in Middle School.

HIV is a rapidly evolving lentivirus, which is made up of RNA, and is dynamic, evolving, changing and mutating whose phylogenetic analysis is complex, but it is very unlikely that two persons will have exactly the same RNA sequence. As a result of the mutation rate of HIV-1, phylogenetic examination of HIV-1 DNA sequences is a powerful apparatus for the distinguishing proof of closely related viral strains that might be utilized to deduce transmission between people [35]. Metzker et al. [35] has chosen to co-amplify pol and env region of HIV-1. It was sequenced, and a relatedness study was carried out to establish the phylogenetic relationship among the victim, suspected transmission pair and molecular clones. The molecular phylogeny evidence shown by Metzker et al. [35] has brought about the conviction of the gastroenterologist of Louisiana on the second-degree attempt to murder charge.

18.4.2 Case II

Fierer et al. [13] has worked to test the hypothesis that bacterial communities are diverse and unique to a person which could be used a microbial fingerprint in forensic identification. They have conducted two sets of experiments. Three computer keyboards and their owners were analysed for the relatedness in the microbial communities; and nine computer mice, the owners hand and the local communities were analysed for linking object with its owner. They demonstrated that skin related microscopic organisms can be promptly recuperated from surfaces (single PC keys and PC mice) and that the structure of these microbial communities can be utilized to separate articles took care of by various people, regardless of whether those items have been submitted untouched for over to 2 weeks at room temperature [13]. In the preliminary work, they have shown that the skin microbial community structure and the object associated with that person have high-level degree of similarity and could be used as a tool in forensic identification.

The DNA recovery and analysis for microbial analysis is easier because of the microbial DNA abundance (on the skin surface and on shed epidermal cells) as compared to human DNA for STR analysis. Besides, the method may be valuable for distinguishing objects from which clear fingerprints can't be acquired (e.g. cloths, smirched and textured surfaces) [13].

References

1. Alberts B, Bray D, Lewis J, Roff M, Roberts K, Watson JD (1994) Molecular biology of cell. Garland Publishing, New York
2. Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169

3. Atlas RM (1984) Diversity of microbial communities. Adv Microbial Ecol 7:1–47
4. Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994) Remarkable archael diversity detected in Yellowstone National Park hot spring environment. Proc Natl Acad Sci 91:1609–1613
5. Bizzarro S, Loos BG, Laine ML, Crielaard W, Zaura E (2013) Subgingival microbiome in smokers and non-smokers in periodontitis: an exploratory study using traditional targeted techniques and a next-generation sequencing. J Clin Periodontol 40:483–492. https://doi.org/10.1111/jcpe.12087
6. Blaser MJ (2010) Harnessing the power of the human microbiome. Proc Natl Acad Sci 107(14):6125–6126
7. Casarin RCV, Barbagallo A, Meulman T, Santos VR, Sallum EA, Nociti FH, Duarte PM, Casati MZ, Gonçalves RB (2013) Subgingival biodiversity in subjects with uncontrolled type-2 diabetes and chronic periodontitis. J Periodontal Res 48:30–36. https://doi.org/10.1111/j.1600-0765.2012.01498.x
8. Clarke TH et al (2017) Integrating the microbiome as a resource in the forensics toolkit. Forensic Sci Int Genet 30:141–147
9. Costello EK et al (2009) Bacterial community variation in human body habitats across space and time. Science 326:1694–1697
10. Dassarma S, Fleischmann EF (1995) Archea: a laboratory manual – halophiles. Cold Springer Harbour Laboratory Press, New York, pp 269–272
11. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563. https://doi.org/10.1038/nature12820
12. Elwood HJ, Olsen GJ, Sogin ML (1985) The small subunit rRNA gene sequences from the hypotrichous ciliates Oxytricha nova and Stylonychia pustulata. Mol Biol Evol 2:399
13. Fierer N, Hamady M, Lauber CL, Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci U S A 105:17994–17999
14. Fuhrman JA (2009) Microbial community structure and its functional implications. Nature 459(7244):193–199
15. Gao Z, Tseng CH, Pei ZH, Blaser MJ (2007) Molecular analysis of human forearm superficial skin bacterial biota. Proc Natl Acad Sci U S A 104:2927–2932
16. Gillis PBM, De Leg J (1992) The genus *Aquaspirillum*. In: Balows A, truper HG, Dworkin M, Harder W, Schleifer KH (eds) The prokaryotes: a handbook on the biology of bacterial ecophysiology, isolation, identification, applications. Springer-Verlag, New York, pp 2569–2582
17. Glenn TC (2011) Field guide to next-generation DNA sequencers. Mol Ecol Resour 11(5):759–769
18. Grice EA, et al.; NISC Comparative Sequencing Program (2009) Topographical and temporal diversity of the human skin microbiome. Science 324:1190–1192
19. Gunn A, Pitt SJ (2012) Review paper microbes as forensic indicators. Trop Biomed 29:311–330
20. Hagstrem A, Pommier T, Fohwer F, Simu K, Stolle W, Svensson D, Zweifel UL (2002) Use of 16S rDNA for delineation of marine bacterioplankton species. Appl Environ Microbiol 68:3628–3633
21. Hewitt KM, Gerba CP, Maxwell SL, Kelley ST (2012) Office space bacterial abundance and diversity in three metropolitan areas. PLoS One 7:e37849. https://doi.org/10.1371/journal.pone.0037849
22. http://people.com/archive/act-of-revenge-vol-51-no-19/
23. http://www.newsweek.com/deadly-attraction-177422
24. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998) Novel division level bacterial diversity in a Yellowstone hot spring. J Bacteriol 180:366–376
25. Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. Nature 486(7402):207–214
26. Kolganova TV, Kuznetsov BB, Tourova TP (2002) Designing and testing oligonucleotide primers for amplification and sequencing of Archaeal 16S rRNA genes. Mikrobiologiya 71:283–286

27. Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G (2014) Shaping the oral microbiota through intimate kissing. Microbiome 2:41. https://doi.org/10.1186/2049-2618-2-41
28. Krishna P, Reddy MS, Satyanarayana T (2006) Molecular techniques for understanding the microbial community structure in Mycorrhizosphere. Microbial activity in the rhizosphere. Springer, Berlin/Heidelberg, pp 173–198
29. Lewin B (2000) Genes VII. Oxford University Press Inc., New York
30. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI (2008) Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol 6:776–788
31. Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49:727–764
32. Luongo JC, Barberán A, Hacker-Cary R, Morgan EE, Miller SL, Fierer N (2016) Microbial analyses of airborne dust collected from dormitory rooms predict the sex of occupants. Indoor Air 27:338–344. https://doi.org/10.1111/ina.12302
33. Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402
34. Martinez-Murcia AJ, Acinas SG, Rodriguez-Valera F (1995) Evalution of prokaryotic diversity by restrictase digestion of 16S rDNA directly amplified from hypersaline environments. FEMS Microbiol Ecol 17:247–256
35. Metzker ML et al (2002) Molecular evidence of HIV-1 transmission in a criminal case. Proc Natl Acad Sci 99(22):14292–14297
36. Miller OL (1981) The nucleolus, chromosomes and visualization of genetic activity. J Cell Biol 91:15s–27s
37. Misic AM, Davis MF, Tyldsley AS, Hodkinson BP, Tolomeo P, Hu B, Nachamkin I, Lautenbach E, Morris DO, Grice EA (2015) The shared microbiota of humans and companion animals as evaluated from Staphylococcus carriage sites. Microbiome 3:2. https://doi.org/10.1186/s40168-014-0052-7
38. Moon JH, Lee JH, Lee JY (2015) Subgingival microbiome in smokers and nonsmokers in Korean chronic periodontitis patients. Mol Oral Microbiol 30:227–241. https://doi.org/10.1111/omi.12086
39. Muyzer G, de Waal EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695–700
40. Nielsen AT, Liu W-T, Filipe C, Grady L, Molin S, Stahl DA (1999) Identification of a novel group of bacteria in a sludge from a deteriorated biological phosphorus removal reactor. Appl Environ Microbiol 65:1251–1258
41. Pace NR, Stahl DA, DJ L, Olsen GJ (1985) Analyzing natural microbial populations by rRNA sequences. ASM News 51:4–12
42. Pitt SJ, Cunningham JM (2009) An introduction to biomedical science in clinical and professional practice. Wiley-Blackwell, Chichester
43. Prosser, J. A. 2002. Molecular functional diversity in soil micro-organisms. In Diversity and integration in mycorrhizas, S. A. Smith and F. A. Smith, Kluwer Academic Publishers, Dordrecht 9–17
44. Rodriguez-Valera F (2002) Approaches to prokaryotic biodiversity: a population genetics perspective. Environ Microbiol 4:628–633
45. Rodriguez-Valera F (2004) Environmental genomics, the big picture? FEMS Microbiol Lett 231:153–158
46. Schloss PD, Handelsman J (2003) Biotechnological prospects from metagenomics. Curr Opin Biotechnol 4:303–310
47. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, Gordon JI, Fierer N, Knight R (2013) Cohabiting family members share microbiota with one another and with their dogs. eLife 2:e00458. https://doi.org/10.7554/eLife.00458

48. Torsvik V, Sørheim R, Goksøyr J (1996) Total bacterial diversity in soil and sediment communities—a review. J Ind Microbiol 17(3-4):170–178
49. van Dijk EL et al (2014) Ten years of next-generation sequencing technology. Trends Genet 30(9):418–426
50. Viaud M, Pasquier A, Brygoo Y (2000) Diversity of soil fungi studied by PCR-RFLP of ITS. Mycol Res 104:1027–1032
51. Vitkup D et al (2001) Completeness in structural genomics. Nat Struct Mol Biol 8(6):559
52. Watanabe K, Kodama Y, Harayama S (2001) Design and evaluation of PCR primers to amplify 16S rDNA fragments used for community fingerprinting. J Microbiol Methods 44:253–262
53. Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95:6578–6583
54. Williams SM, Robbins LG (1992) Molecular genetic analysis of drosophila rDNA arrays. Trends Gent 8:335–340
55. Woese CR (1987) Bacterial evolution. Microbiol Rev 51(2):221
56. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI (2012) Human gut microbiome viewed across age and geography. Nature 486:222–227. https://doi.org/10.1038/nature11053
57. Yun J, Ryu S (2005) Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microbiol Cell Fact 4:8

Appendix

Scientific Working Group on DNA Analysis Methods (SWGDAM) guidelines for DNA fingerprinting analysis

The Scientific Working Group on DNA Analysis Methods, better known by its acronym of SWGDAM, is a group of approximately 50 scientists representing Federal, State, and Local forensic DNA laboratories in the United States and Canada. During meetings, which are held twice a year, Committees discuss topics of interest to the forensic DNA community and often develop documents to provide direction and guidance for the community. This is the latest guidelines issued by the committee in November 2016 to address Next Generation Sequencing (NGS) technologies. The SWGDAM Executive Board approved posting of this document, with the minor revisions, in December 2016.

1. General Considerations

1.1 Validation is a process by which a procedure is evaluated to determine its efficacy and reliability for forensic casework and/or database analysis.

1.2 There are two types of validation required to implement or modify technologies for forensic DNA analysis – developmental and internal. The application of existing technology to the analysis of forensic samples does not necessarily create a new technology or methodology. Developmental validation studies in other fields may sufficiently address forensic applications.

1.2.1. Developmental validation is the acquisition of test data and determination of conditions and limitations of a new or novel DNA methodology for use on forensic, database, known or casework reference samples.

1.2.1.1. Peer-reviewed publication of the underlying scientific principle(s) of a technology shall be required.

1.2.1.2. Peer-reviewed publication (or other means of dissemination to the scientific community, such as presentation at a scientific meeting) of developmental validation studies is encouraged.

H. R. Dash et al. (eds.), *DNA Fingerprinting: Advancements and Future Endeavors*, https://doi.org/10.1007/978-981-13-1583-1

However, validated technologies or procedures may be implemented without such publication.

1.2.2. Internal validation is an accumulation of test data within the laboratory to demonstrate that established methods and procedures perform as expected in the laboratory. Prior to using a procedure for forensic applications, a laboratory shall conduct internal validation studies.

1.2.2.1. Internal validation studies should be sufficiently documented and summarized.

1.2.2.2. Quality assurance parameters and interpretation guidelines shall be derived from internal validation studies. For example, lower template DNA may cause extreme heterozygote imbalance; as such, empirical heterozygote peak-height ratio data could be used to formulate mixture interpretation guidelines and determine the appropriate ratio by which two peaks are determined to be heterozygotes. In addition to establishing an analytical threshold, results from sensitivity studies could be used to determine the extent and parameters of quality control tests that reagents require prior to their being used in actual casework.

1.2.2.3. For laboratory systems that consist of more than one laboratory, each of the laboratories shall perform, document and maintain studies which may be impacted by location-specific factors (such as precision, sensitivity, contamination, etc.). Studies that are not location-specific may be shared among all locations.

2. Developmental Validation

The developmental validation process shall include, where applicable, the following studies:

2.1 Characterization of genetic markers: The basic characteristics (described below) of a genetic marker should be determined and documented.

2.1.1 Inheritance: The mode of inheritance of DNA markers demonstrated through family studies.

2.1.2 Mapping: The genomic location of the genetic marker.

2.1.3 Detection: Technological basis for identifying the genetic marker (e.g., capillary electrophoresis, DNA sequencing, hybridization assays, etc.).

2.1.4 Polymorphism: Type of variation (e.g. sequence and/or length variants)

2.2 Species specificity: The ability to detect genetic information from non-targeted species (e.g., detection of microbial DNA in a human assay) should be determined. The detection of genetic information from non-targeted species does not necessarily invalidate the use of the assay, but may help define the limits of the assay.

2.3 Sensitivity studies: The ability to obtain reliable results from a range of DNA quantities, to include the upper and lower limits of the assay, should be evaluated.

2.4 Stability studies: The ability to obtain results from DNA recovered from biological samples deposited on various substrates and subjected to various environmental and chemical insults should be evaluated. In most instances, assessment of the effects of these factors on new forensic DNA procedures is not required. However, if substrates and/or environmental and/or chemical insults could potentially affect the analytical process, then the process should be evaluated to determine the effects of such factors.

2.5 Precision and accuracy of the assay should be demonstrated: Precision characterizes the degree of mutual agreement among a series of individual measurements, values and/or results. Precision depends only on the distribution of random errors and does not relate to the true value or specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results.

Accuracy is the degree of conformity of a measured quantity to its actual (true) value. Accuracy of a measuring instrument is the ability of a measuring instrument to give responses close to a true value.

2.5.1 Repeatability: Precision and accuracy of results (e.g., quantitative and/or qualitative) of the same operator and/or detection instrument should be evaluated.

2.5.2 Reproducibility: Precision and accuracy of results (e.g., quantitative and/or qualitative) among different operators and/or detection instruments should be evaluated.

2.6 Case-type samples: The ability to obtain reliable results should be evaluated using samples that are representative of those typically encountered by the testing laboratory. Where appropriate, consistency of typing results should be demonstrated by comparing results from the previous procedures to those obtained using the new procedure.

2.7 Population studies: The distribution of genetic markers in populations should be determined in relevant population groups. When appropriate, databases should be tested for independence expectations.

2.8 Mixture studies: The ability to obtain reliable results from mixed-source samples should be determined. These studies will assist the laboratory to establish guidelines for mixture interpretation, which may include determination of the number of contributors to the mixture, determination of the major and minor contributor profiles, and contributor ratios or proportions.

2.9 PCR-based studies

2.9.1. Publication of the sequence of individual primers is not required in order to appropriately demonstrate the reliability and limitations of PCR-based technologies. However, availability of the primer sequences

is encouraged in order to aid in the identification of potential primer binding site variants and troubleshooting.

2.9.2. The reaction conditions needed to provide the required degree of specificity and robustness should be determined. These include, but are not limited to, thermal cycling parameters, the concentration of primers, magnesium chloride, DNA polymerase, and other critical reagents.

2.9.3. The potential for differential amplification among loci, preferential amplification of alleles in a locus, and stochastic amplification (i.e., excessive allelic signal imbalances due to the random sampling and amplification of low template quantities) should be assessed.

2.9.4. The effects of multiplexing should be assessed.

2.9.5. Appropriate controls should be assessed.

2.9.6. Criteria for detection of amplified product should be determined based on the platform and/or method.

2.9.7. Appropriate measurement standards (qualitative and/or quantitative) for characterizing the alleles or resulting DNA product should be established.

2.10 NGS-specific studies

2.10.1 The effects of barcoding/indexing samples and subsequent bioinformatic sample separation should be assessed, to include, when appropriate, the effects of novel barcodes/indices that have not otherwise undergone developmental validation.

2.10.2 Sensitivity studies should address limit of detection as it relates to both starting DNA input as well as the extent of varying the quantity (extent of sample multiplexing in sequencing) and/or quality of libraries pooled in the sequencing reaction.

2.10.3 NGS instrumentation should be assessed for the possibility of signal cross talk during sequencing, and sample carryover between runs.

3. Internal Validation

The internal validation process shall include the studies detailed below. If conducted within the same laboratory, developmental validation studies may satisfy some elements of the internal validation guidelines. The laboratory should evaluate the appropriate sample number and type, based on the methodology and/or application necessary to demonstrate the potential limitations and reliability. The laboratory should determine the suitability of each study based on the methodology and may determine that a study is not necessary. The recommended internal validation studies are summarized in Table 1.

System includes methodology, chemistries and instrumentation. *Mixture studies will be required if the assay is intended to distinguish different contributors (male/female/major/minor etc.).

3.1. Known and nonprobative evidence samples or mock evidence samples: Methods intended for casework samples should be evaluated and tested using known samples and nonprobative evidence samples or mock case samples.

Table 1 Summary of recommended studies for internal validation

	Extraction system	Quantitation system	Amplification system	Detection system
Known / Non-Probative Samples	X	X	X	X
Precision and Accuracy: Repeatability	X	X	X	X
Precision and Accuracy: Reproducibility	X	X	X	X
Sensitivity Studies	X	X	X	X
Stochastic Studies	X	X	X	X
Mixture Studies	X*	X*	X	X
Contamination Assessment	X	X	X	X

Methods intended for database samples should be evaluated and tested using known samples. Results from these studies should be compared to the previous results of known samples and/or nonprobative evidence or mock case samples to ensure concordance.

3.2. Sensitivity and Stochastic Studies: The laboratory should demonstrate sensitivity levels of the test. Sensitivity studies are used to determine the dynamic range, ideal target range, limit of detection, limit of quantitation, heterozygote balance (e.g., peak height ratio) and the signal to noise ratio associated with the assay. Sensitivity studies can also be used to evaluate excessive random (stochastic) effects generally resulting from low quantity and/or low quality samples.

3.3. *Precision and accuracy of the assay should be demonstrated: Precision* characterizes the degree of mutual agreement among a series of individual measurements, values and/or results. Precision depends only on the distribution of random errors and does not relate to the true value or specified value. The measure of precision is usually expressed in terms of imprecision and computed as a standard deviation of the test results.

Accuracy is the degree of conformity of a measured quantity to its actual (true) value. Accuracy of a measuring instrument is the ability of a measuring instrument to give responses close to a true value.

3.3.1 Repeatability: Precision and accuracy of results (e.g., quantitative and/or qualitative) of the same operator and/or detection instrument should be evaluated.

3.3.2 Reproducibility: Precision and accuracy of results (e.g., quantitative and/or qualitative) among different operators and/or detection instruments should be evaluated.

3.4. Mixture studies: Mixed DNA samples that are representative of those typically encountered by the testing laboratory should be evaluated. These studies will assist a casework laboratory to establish guidelines for mixture interpretation, which may include determination of the number of contributors to the mixture,

determination of the major and minor contributor profiles, and contributor ratios or proportions. A simplified mixture study may also assist a databasing laboratory to recognize mixtures and/or contamination.

3.5. Contamination assessment: The laboratory should evaluate, using both controls and known samples, the detection of exogenous DNA (including allele drop-in and heteroplasmy) originating from reagents, consumables, other samples, operator and/or laboratory environment.

3.6. NGS-specific studies

3.6.1 Sensitivity studies should address limit of detection as it relates to both starting DNA input as well as the extent of varying the quantity (extent of sample multiplexing in sequencing) and/or quality of libraries pooled in the sequencing reaction.

4. Material Modification

A material modification is an alteration of an existing analytical procedure that may have a consequential effect(s) on analytical results; for example, a decrease in reaction volume of an amplification test kit that is already in use by the laboratory or a change in injection time for a genetic analyzer. A material modification shall be evaluated by comparing the results from the original procedure to the results of the modified procedure to ensure concordance. The laboratory should evaluate the appropriate sample number, sample type, and the studies necessary to demonstrate this.

5. Performance Check

A performance check is a quality assurance measure to assess the functionality of laboratory instruments and equipment that affect the accuracy and/or validity of forensic, database, known or casework reference sample analysis. This may be required after repairs and/or scheduled maintenance. The laboratory should evaluate the appropriate sample number and type to demonstrate the reliability of the instrument or equipment. The laboratory should also determine the suitability of each study and may determine that a study is not necessary.

5.1 If the physical location or the environment of the instrument has been changed (e.g., instrument moved to another room, significant remodeling of the room, etc.), a performance check should be completed.

5.2 After an internal validation has been performed on a critical instrument, each additional critical instrument of the same make and model shall require a performance check. The performance check should demonstrate that results are reproducible on the new critical instrument and that values from the internal validation can still be obtained. For example, the performance check of a new critical instrument should demonstrate that the sensitivity level is consistent with the sensitivity level obtained from an internal validation, but need not demonstrate whether or not the new critical instrument is more sensitive.

6. Software

6.1 New software or significant software changes (including bioinformatics tools) that may impact interpretation, the analytical process, or sizing algorithms shall require a validation prior to implementation. Depending on the function and application of the software, the laboratory should determine the appropriate validation studies to identify its reliability and limitations.

6.2 A software upgrade that would not impact interpretation, the analytical process, or sizing algorithms shall require a performance check.

7. Further Reads/Suggested Readings

Butler, J.M. Quality Assurance and Validation. In: Advanced Topics in Forensic DNA Typing: Methodology. Elsevier, 2011.

FBI. Quality Assurance Standards for Forensic DNA Testing Laboratories (September 1, 2011) available at http://www.fbi.gov/about-us/lab/codis/qas- standards-for-forensic-dna-testing-laboratories-effective-9-1-2011

Scientific Working Group on DNA Analysis Methods. Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Approved at the Scientific Working Group on DNA Analysis Methods meeting, Fredericksburg, Virginia, January 2010. Available at http://www.swgdam.org/Interpretation_Guidelines_January_2010.pdf