

Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample

C. Robino · F. Crobu · C. Di Gaetano ·
A. Bekada · S. Benhamamouch · N. Cerutti ·
A. Piazza · S. Inturri · C. Torre

Received: 12 April 2007 / Accepted: 31 August 2007 / Published online: 2 October 2007
© Springer-Verlag 2007

Abstract The distribution of Y-chromosomal single nucleotide polymorphism (SNP) haplogroups and short tandem repeat (STR) haplotypes was determined in a sample of 102 unrelated men of Arab origin from northwestern Algeria (Oran area). A total of nine different haplogroups were identified by a panel of 22 binary markers. The most common haplogroups observed in the Algerian population were E3b2 (45.1%) and J1 (22.5%). Y-STR typing by a 17-loci multiplex system allowed 93 haplotypes to be defined (88 were unique). Striking differences in the allele distribution and gene diversity of Y-STR markers between haplogroups could be found. In particular, intermediate alleles at locus DYS458 specifically characterized the haplotypes of individuals carrying haplogroup J1. All the intermediate alleles shared a common repeat sequence structure, supporting the hypothesis that the variant originated from a single mutational event.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-007-0203-5) contains supplementary material, which is available to authorized users.

C. Robino (✉) · S. Inturri · C. Torre
Department of Anatomy, Pharmacology and Legal Medicine,
University of Turin,
Turin, Italy
e-mail: carlo.robino@unito.it

F. Crobu · C. Di Gaetano · N. Cerutti · A. Piazza
Department of Genetics, Biology and Biochemistry,
University of Turin,
Turin, Italy

A. Bekada · S. Benhamamouch
Département de Biotechnologie, Faculté des Sciences,
Université d'Oran,
Oran, Algeria

Keywords Y chromosome · Single nucleotide polymorphism · Short tandem repeat · Algeria · DYS458

Introduction

Over the last decade, the analysis of Y-chromosomal short tandem repeat (STR) loci has emerged as a powerful tool for paternity testing and forensic casework [1]. The need of information regarding the distribution of Y-STR haplotypes in human populations to obtain reliable frequency estimates has produced an enormous amount of literature, most of which was conveyed in online reference databases like the Y Haplotype Reference Database (YHRD) [2]. As novel Y-STRs become conveniently available in multiplex polymerase chain reaction (PCR) format, in addition to those traditionally included in forensic databases forming the so-called minimal haplotype (minHt), population data on extended Y-chromosomal haplotypes are required. In recent years, there has been an increasing interest in the forensic community towards complementing Y-STR data with information on Y-chromosomal single nucleotide polymorphisms (SNPs) [3]. These binary markers are characterized by a low mutation rate, and therefore their occurrence represents a unique event in human evolution. These properties, combined with the absence of recombination in most parts of the Y chromosome and the consequent accumulation of genetic diversity among lineages and populations, make Y-SNPs a valuable tool for predicting the geographic or ethnic origin of unknown samples found in forensic casework [4, 5]. In this study, combined Y-chromosomal SNP and STR variation was analyzed in a sample of Algerian individuals of Arab descent from the Oran area (northwest Algeria).

Materials and methods

Blood samples were drawn from 102 healthy adult men of Arab descent from the Oran area (northwest Algeria). Informed consent was obtained from all subjects, and sampling was anonymous. DNA extraction was carried out after the salting-out method [6]. Y-SNP genotyping was performed by heteroduplex analysis of PCR products, using the primers described by Cinnioglu et al. [7] and the Transgenomic WAVE dHPLC system (Transgenomic, Omaha NE, USA) [8]. The set of 22 binary markers used in the study and the complete Y-SNP typing results are shown in supplementary Table S1. Data were referred to by haplogroup, according to a standardized nomenclature [9]. Diagnostic Y-SNP defining each haplogroup are indicated in supplementary Table S1. Y-STRs were amplified with the AmpF/STR Yfiler system (Applied Biosystems, Foster City CA), according to the manufacturer's instructions, and typed by capillary electrophoresis on the ABI PRISM 310 Genetic Analyzer using GeneScan and Genotyper software (Applied Biosystems). Sequencing of DYS458 intermediate alleles was performed using the Big Dye Terminator Sequencing kit (Applied Biosystems) with primers described by Schoske et al. [10]. Standard diversity indices, pairwise genetic distances (R_{ST}) and analysis of molecular variance (AMOVA) were calculated with the software Arlequin version 2.000 [11]. Discrimination capacity (DC) was determined by dividing the number of observed haplotypes by the number of sampled individuals. The Neighbor Joining (NJ) tree illustrating the relationship between Algerian Arabs and 11 reference populations based on R_{ST} values was constructed using the program PHYLIP 3.67 [12]. Haplotype-frequency surfaces were graphically computer reconstructed following the Kriging procedure [13] by use of the Surfer System version 8.05 (Golden Software).

Results and discussion

The list of observed haplogroups is provided in supplementary Table S1. E3b2 was found to be the most frequent haplogroup (45.1%) in the northwest Algerian population sample. As shown by previous studies, E3b2 is very common in northwest Africa, but its frequency sharply declines eastward across north Africa [14], and it is almost completely absent in sub-saharan Africa. Individuals carrying this haplogroup are also found, with much lower frequencies compared to those observed in northwest Africa, in Turkey, the near East, the Balkans, southern Europe and in Iberia, possibly as a consequence of recent gene flow due to Islamic occupation of the peninsula [15]. Another common haplogroup in the Algerian population is J1 (22.5%). Outside the Maghreb, haplogroup J1 was shown to have the highest concentration in Egypt and the Middle East, with declining frequencies towards the east (Indian subcontinent), south (Horn of Africa) and northwest (Turkey, the Balkans and southern Europe) [16, 17]. Haplogroups typically found in European (R1b3) and sub-saharan African (E3a) populations were also observed with frequencies of 10.8 and 7.8%, respectively.

The list of observed haplotypes has been submitted to the YHRD and is provided in supplementary Table S2. By Y-STR typing, a total of 93 different 17-loci haplotypes were found in the Algerian population sample, and 88 were unique. Three haplotypes were shared by four, three and two individuals, respectively, all belonging to haplogroup E3b2. Two haplotypes, showing a single one-repeat unit mismatch at locus DYS389II, were shared by three and two subjects, respectively, carrying haplogroup J1. No cases of haplotype sharing among different haplogroups were observed, even when only minHt loci were considered. The contribution of additional Y-STRs to haplotype diversity and power of discrimination provided by minHt loci is shown in Table 1.

Table 1 Haplotype diversity (h) and discrimination capacity (DC) of minimal and 17-loci haplotypes in the whole Algerian population sample and within major haplogroups

	Total ($n=102$)			E3b2 ($n=46$)			J1 ($n=23$)		
	GD	h	DC	GD	h	DC	GD	h	DC
MinHt	–	0.979±0.007	0.716	–	0.900±0.029	0.457	–	0.972±0.026	0.826
DYS458	0.806	0.993±0.003	0.814	0.633	0.968±0.014	0.652	0.356	0.984±0.019	0.870
DYS456	0.670	0.988±0.004	0.765	0.446	0.948±0.017	0.565	0.312	0.972±0.026	0.826
DYS439	0.644	0.982±0.007	0.755	0.552	0.917±0.027	0.543	0.372	0.972±0.026	0.826
YGATAH4	0.570	0.984±0.006	0.735	0.165	0.925±0.023	0.500	0.237	0.972±0.026	0.826
DYS448	0.461	0.982±0.006	0.735	0.166	0.918±0.025	0.500	0.170	0.972±0.026	0.826
DYS635	0.466	0.981±0.007	0.735	0.202	0.911±0.028	0.500	0.387	0.972±0.026	0.826
DYS438	0.456	0.979±0.007	0.716	0.004	0.900±0.029	0.457	0.000	0.972±0.026	0.826
DYS437	0.227	0.979±0.007	0.716	0.000	0.900±0.029	0.457	0.000	0.972±0.026	0.826
17-loci haplotype	–	0.997±0.002	0.918	–	0.990±0.008	0.870	–	0.984±0.019	0.870

Contribution to h and DC of minHt loci provided by single additional Y-STR markers ranked by gene diversity (GD) is shown.

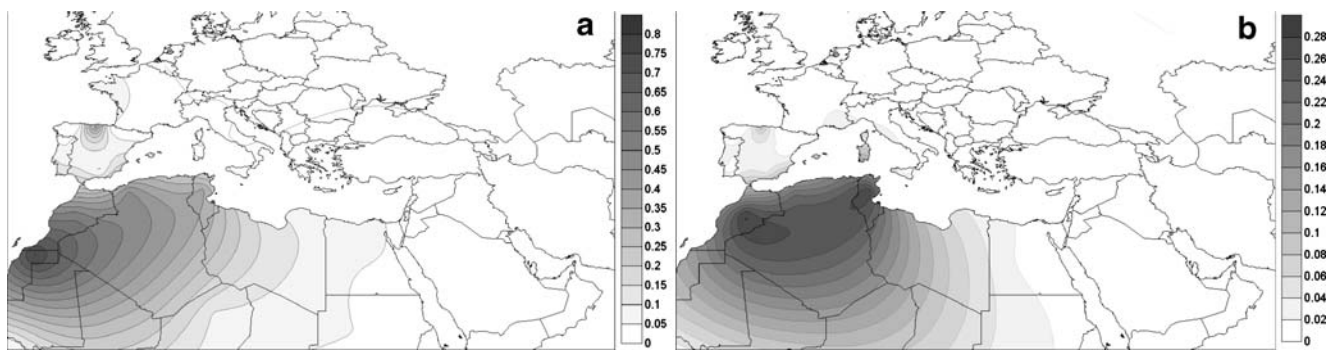


Fig. 1 Contour maps created by the Kriging method showing the frequency distributions of: **a** haplogroup E3b2, based on this study and literature data [14–16]; **b** the combination of the two most common haplotypes observed in Algerians carrying haplogroup E3b2 (DYS19:13; DYS389I:14; DYS389II:30; DYS390:24; DYS391:9/10; DYS392:11; DYS393:13), based on this study and data from three

continental metapopulations (Europe, Asia, Africa) included in the YHRD (release 21). For geographical areas not covered in the YHRD, like Morocco and Palestine, population data from Quintana-Murci et al. [18] and Nebel et al. [28] were, respectively, used. Data from Nebel et al. on Jews and Palestinian Arabs do not include locus DYS389I/II

There was good correlation between gene diversity (GD) of additional loci and their ability to distinguish between Y lineages, as assessed by DC. Such a strict correspondence was not seen for haplotype diversity (h), e.g. locus YGATAH4 although less diverse than DYS439 provided a larger percentage increase in h . With the exception of E3b2 and J1, all haplogroups were already subdivided into unique Y lineages by minHt loci. Among individuals sharing haplogroup E3b2, h as determined by minHt markers (0.900 ± 0.029) was low, if compared to the general sample (0.979 ± 0.007), but by typing additional Y-STRs, the difference was almost completely compensated. Diversity between minHts within haplogroup J1 (0.972 ± 0.026) was closer to that observed in the whole population sample, although among the additional loci only DYS458 further contributed to the increase in h and DC.

Pairwise genetic distances (R_{ST}) calculated between minHts (with the exclusion of DYS385) found in the Arab population sample from Algeria and geographical neighbours from north Africa [14, 18–20] and southern Europe

[21–23] are shown in Table S3. Significant R_{ST} values were observed between Algerian Arabs and Berbers from Morocco ($R_{ST}=0.05962$; $p<0.01$) and Tunisia ($R_{ST}=0.02791$; $p<0.01$), whereas the only significant comparisons with non-Berber populations from north Africa were with northern Egyptians ($R_{ST}=0.01996$; $p<0.05$) and Tunisian Andalusians ($R_{ST}=0.00962$; $p<0.05$). The NJ tree illustrating the relationship between populations based on R_{ST} values is shown in Fig. S1.

The apportionment of haplotype diversity among and within haplogroups in the Algerian population sample, as determined by AMOVA, showed that 71.5% ($p<0.0001$) of the total genetic variation was attributable to differences between haplogroups. The most frequent Y-STR alleles within major Algerian haplogroups are shown in supplementary Table S4. The ability of single Y-STRs included in the minHt to clearly differentiate between haplogroups has been previously described, e.g. African haplogroup E3a was shown to be typically associated with alleles with less than 22 repeat units at locus DYS390 [24]. Among

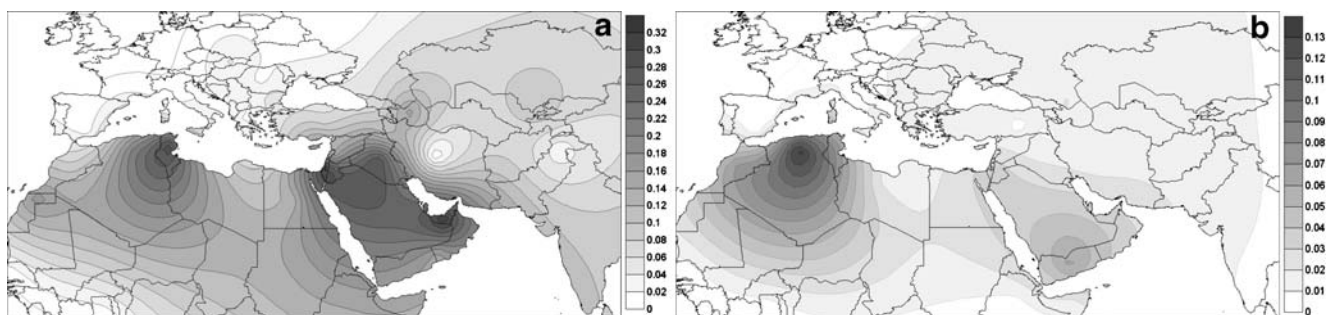


Fig. 2 Contour maps created by the Kriging method showing the frequency distributions of: **a** haplogroup J1, based on this study and literature data [16, 17]; **b** the combination of the two most common haplotypes observed in Algerians carrying haplogroup J1 (DYS19:14; DYS389I:13; DYS389II:29/30; DYS390:23; DYS391:11; DYS392:11; DYS393:12), based on this study and data from three

continental metapopulations (Europe, Asia, Africa) included in the YHRD (release 21). For geographical areas not covered in the YHRD, like Morocco and Palestine, population data from Quintana-Murci et al. [18] and Nebel et al. [28] were, respectively, used. Data from Nebel et al. on Jews and Palestinian Arabs do not include locus DYS389I/II

additional Y-STRs included in the 17-loci multiplex PCR system used in this study, DYS635 completely differentiated the European haplogroup R1b3 (alleles 23 and 24) from Mediterranean and African haplogroups, in which only alleles with less than 23 repeats were observed. Similarly, genotypes at locus DYS448 discriminated between R1b3 (alleles 18 and 19) and the African haplogroup E3a (alleles 20–22). Also noteworthy was the high frequency of intermediate alleles (18.2 and 19.2) at locus DYS458 and its constant association with haplogroup J1 in the Algerian population. Although there are data suggesting that locus DYS458 may be affected by a higher mutation rate compared to other tetrameric Y-STR loci [25], the observed microvariants seem to reflect a single mutational event. They all shared a unique Y-SNP background and according to sequencing results a common repeat sequence structure [GAAA]_{16–17}AA[GAAA]₂. These findings are consistent with other recent population studies indicating that individuals carrying intermediate alleles at locus DYS458 are confined to a distinct subclade of haplogroup J [21, 26]. Further collection of joint Y-SNP and Y-STR population data will help to better understand the meaning of this association in terms of human evolutionary history.

The absence of recombination in the Y chromosome implies that haplotype diversity within subjects carrying a certain diagnostic SNP is limited to mutations accumulated over generations in the founder haplotype. As a consequence, specific allele combinations observed in a haplotype can be used to a certain extent to predict haplogroup, and hence the geographic or ethnic origin of an unknown biological sample. For instance, the frequency distribution of modal minHts (with the exclusion of DYS385) seen in the Algerian population sample among E3b2 and J1 individuals closely resembles that of the corresponding haplogroups (Figs. 1 and 2).

In this respect, the availability of detailed information on Y-SNP and Y-STR joint distribution in human populations can provide a great benefit to forensic experts. The analysis of large Y-SNP panels for in-depth resolution of haplogroups is still a complex and DNA consuming procedure [5, 27] and cannot be routinely performed on small stains, like those commonly analyzed in forensic casework, which have to be preserved for STR multilocus profiling and individual identification. However, a-priori knowledge about the most likely SNP background of a certain Y-chromosomal haplotype can help the investigators to infer the provenience of the subject who left the evidence. At the same time, information on the geographic and ethnic distribution of Y-SNP haplogroups and their internal Y-STR variability can assist the expert with the estimate of the frequency of a particular haplotype in populations for which a specific reference database of Y-STR haplotypes is not yet available. To integrate Y-STR and Y-SNP data in

convenient on-line accessible databases, issues like the selection of binary polymorphisms to generate a standard level of tree resolution, choice of suitable technologies for Y-SNP typing, Y-SNP nomenclature, and quality control of participating laboratories will have to be addressed in the future by the forensic community.

Acknowledgements The continuing support of the Compagnia di San Paolo to C. Torre's laboratory is gratefully acknowledged. We also thank Dr Boudjema A, Medienna Benchekor S and Meghoufel N for assistance in the collection of the samples.

References

- Gusmao L, Butler JM, Carracedo A et al (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int J Legal Med* 120:191–200
- Willuweit S, Roewer L (2007) Y chromosome haplotype reference database (YHRD): Update. *FSI: Genetics* 1:83–87
- Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci Int* 118:158–162
- Wetton JH, Tsang KW, Khan H (2005) Inferring the population of origin of DNA evidence within the UK by allele-specific hybridization of Y-SNPs. *Forensic Sci Int* 152:45–53
- Brion M, Sanchez JJ, Balogh K et al (2005) Introduction of a single nucleotide polymorphism-based "Major Y-chromosome haplogroup typing kit" suitable for predicting the geographic origin of male lineages. *Electrophoresis* 26:4411–4420
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16:1215
- Cinnioglu C, King R, Kivisild T et al (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114:127–148
- Oefner PJ, Underhill PA (1998) DNA mutation detection using denaturing high performance liquid chromatography (DHPLC). In: Dracopoli NC et al (eds) *Current protocols in human genetics*, Suppl 19. Wiley, New York, pp 7.10.1–7.10.12
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612
- Schoske R, Vallone PM, Kline MC, Redman JW, Butler JM (2004) High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Sci Int* 139:107–121
- Schneider S, Roessli D, Excoffier L (2000) Arlequin Version 2.000: A Software for Population Genetics Data Analysis, Genetics and Biometry Laboratory, University of Geneva, Switzerland
- Felsenstein J (1989) PHYLIP: phylogeny inference package (version 3.2). *Cladistics* 5:164–166
- Delfiner P (1976) Linear estimation of non-stationary spatial phenomena. In: Guarasio M, David M, Haijbechts C (eds) *Advanced geostatistics in the mining industry*. Reidel, Dordrecht, pp 49–68
- Arredi B, Poloni ES, Paracchini S et al (2004) A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 75:338–345
- Cruciani F, La Fratta R, Santolamazza P et al (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* 74:1014–1022

16. 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
17. Di Giacomo F, Luca F, Popa LO et al (2004) Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* 115:357–371
18. Quintana-Murci L, Bigham A, Rouba H, Barakat A, McElreavey K, Hammer M (2004) Y-chromosomal STR haplotypes in Berber and Arabic-speaking populations from Morocco. *Forensic Sci Int* 140:113–115
19. Cherni L, Pereira L, Goios A et al (2005) Y-chromosomal STR haplotypes in three ethnic groups and one cosmopolitan population from Tunisia. *Forensic Sci Int* 152:95–99
20. Frigi S, Pereira F, Pereira L et al (2006) Data for Y-chromosome haplotypes defined by 17 STRs (AmpFLSTR Yfiler) in two Tunisian Berber communities. *Forensic Sci Int* 160:80–83
21. Alves C, Gomes V, Prata MJ, Amorim A, Gusmao L (2007) Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR YFiler) in Portugal. *Forensic Sci Int* DOI 10.1016/j.forsciint.2006.10.002
22. Robino C, Inturri S, Gino S et al (2006) Y-chromosomal STR haplotypes in Sicily. *Forensic Sci Int* 159:235–240
23. Gonzalez-Neira A, Gusmao L, Brion M, Lareu MV, Amorim A, Carracedo A (2000) Distribution of Y-chromosome STR defined haplotypes in Iberia. *Forensic Sci Int* 110:117–126
24. Bosch E, Calafell F, Santos FR et al (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *Am J Hum Genet* 65:1623–1638
25. Mulero JJ, Chang CW, Calandro LM, Green RL, Li Y, Johnson CL, Hennessy LK (2006) Development and validation of the AmpFISTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci* 51:64–75
26. Onofri V, Alessandrini F, Turchi C, Fraternali B, Buscemi L, Pesaresi M, Tagliabracci A (2007) Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis. *Int J Legal Med* 121:234–237
27. Sobrino B, Brion M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154:181–194
28. Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am J Hum Genet* 69:1095–1112