

# A revised root for the human Y chromosome differentiation and diversity landscape among North African populations

## Abstract

North Africa is a crucial area in the settlement history of modern humans because it represents a possible connection between Africa and Europe. So far, genetic data were inconclusive about the fact that this region constitutes a barrier to gene flow, as previous results were highly variable depending on the genetic locus studied. The present study evaluates the impact of North Africa in reducing gene flow between populations from Africa and Europe, by comparing formally various genetic loci in the Y chromosome and analyzing several parameters of population differentiation taking into account the effects of both demography and natural selection at some loci. The Y chromosome is inherited through the paternal line and remains virtually unchanged through many generations. The combination of the analysis of Y chromosome STR and SNP will allow the analysis of human evolution in paternal lineages in different time scales. The forensic implication of the results are discussed.

**Keywords:** North Africa, Y chromosome, STR, SNP, ancestry, population, gene flow

Volume 5 Issue 1 - 2018

**K Bentayebi, A Hajitou**

Department of Medicine, Imperial College London, UK

**Correspondence:** Kaoutar Bentayebi, Phage Therapy Group, Division of Brain Sciences, Department of Medicine, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, UK, Tel +07388288672, Email kaoutar.bentayebi@gmail.com

**Received:** June 16, 2017 | **Published:** September 07, 2018

## Introduction

The Y chromosome has a curious role in population Genetics and forensic genetics. It is male specific and constitutively haploid. It passes from father to son, and, unlike other chromosomes, largely escapes meiotic recombination. Two segments (the pseudoautosomal regions) do recombine with the X, but this amount to less than 3 Mb of its ~60-Mb length, which makes it easier to study than the autosomes.<sup>1</sup> These traits, explains its use in an increasing number of studies, especially those that address the history of the human population and makes it easier to study than the autosomes. Nevertheless, the Y chromosome continues to be overshadowed in these studies by the autosomal chromosome and mtDNA, despite containing far more genetic information than either. One reason for the lagging position of the Y chromosome has been a delay in the generation of useful data. The first PHYLOGENETIC TREE for human mtDNA was published in 1987,<sup>2</sup> and the first for the Y chromosome was published in 1989.<sup>3</sup> Since then, Y-chromosome-based studies have slowly accumulated; with the maturation of sequencing and genotyping technologies, a tremendous increase in all kinds of genetic data looms. It is therefore an appropriate time to reflect on what the Y chromosome can offer to population-genetic studies.

## Overview on Y chromosome in North Africa

The western Mediterranean region has a wide genetic interest in population genetics and forensic genetic studies. It has been a focus of genetic research for many years, due to its geostrategic position distinguished by its ancient and shifting history. The sequence differentiation of human Y chromosomes has occurred during and after the process of colonization and diffusion into the different geographic regions and continents, and its dissection is a useful tool for the investigation of range expansions, migrations and other forms of gene flow in prehistoric and historic times. In other words, the sequence variation of modern Y chromosomes represents a unique

record of the history of human males, and of the relationships between past populations. A number of recent studies carried out in the Moroccan, Algerian and Tunisian populations have shown that most Y chromosomes can be classified into monophyletic units (haplogroups), which tend to be specific to each continent and major ethnic group.<sup>4-6</sup> This study takes advantage of this ethnic/geographic specificity to define better the origins and relationships of the populations living in the Western Mediterranean basin.

We are interested in using two types of YDNA markers in our study: STR and SNP. Applications of these markers include phylogeny, forensics, paternity tests, genetic marker for disease, gene mapping, selection of traits in breeding, etc. i) STR stands for “short tandem repeat”. The STRs are repetitions of 2 or more nucleotides that occur mostly in the non-coding parts of the DNA. The number of repeats is called allele value. STRs are highly polymorphic which means that the amount of repetitions in a specific STR can be different from person to person. The same number of repeats between persons across a wide range of STR markers indicates that they are most likely related. ii) SNP stands for “single nucleotide polymorphism”. This means that there is a variation or mutation in one single nucleotide at a specific location in the genome (either in a gene or in a non-coding region). Unlike Y-STR markers which can have multiple repeats, SNP markers often only exist in 2 forms – the ancestral and the mutated allele. Y chromosome-specific single nucleotide polymorphisms (SNPs) are particularly useful for identification of stable paternal lineages because of their low rate of parallel and back mutations. By testing highly polymorphic Y-STR markers that can inform of recent ancestry in addition to the Y-SNP markers an even more detailed DNA profile can be obtained.<sup>7-9</sup> Data here were collected from previous analysis on three North African populations (Morocco, Algeria and Tunisia). Y chromosomes were typed using 17 Short Tandem Repeats polymorphisms and SNP polymorphism for 52 Tunisian,<sup>10</sup> 102 Algerian<sup>11</sup> and 267 Moroccan.<sup>12,13</sup> All 421 individuals

were tested for seventeen STRs (DYS19, DYS385 a/b, DYS389 I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 (Y GATA C4), and YGATA H4. dinucleotide repeat markers (YCAIIa, YCAIIb, DYS413a, and DYS413b).

From the extreme north western Africa to the east, data showed a decreasing pattern of diversity. The Moroccan, Algerian and Tunisian populations showed the respective values of different haplotypes: 257 (96.25%), 93 (91.18%) and 39 (75%) of which 247, 88 and 29 were unique. Forensic parameters displayed in table 1 showed a striking difference in discrimination capacity among the three North African populations for the 17 Y-STR markers, with a low value in Tunisia (0,750) compared to Algeria (0.918) and Morocco (0.963). Generally, lower levels or variation are observed on the Y chromosome in the Tunisian population. The smaller sample size of the Tunisian population included in this study could have affected the results and lead to show a lower diversity within this population. We couldn't increase de sample size because we didn't find studies in the literature describing wider Tunisian population sample. Further analysis on this population including these markers has to be done. The increasing pattern of genetic diversity from the North eastern part of Maghreb to the western shores, may be seemingly be explained by the out of Africa model. Indeed, the genetic segments of Y chromosome might had further time interval to acquire genetic differences and variation in Morocco compared to Tunisia. Two microvariant alleles (18.2 and 19.2) were observed at the locus DYS458 in the three studied populations, with a lower frequency than that observed in the Upper (Southern) Egyptian population. Although there are data suggesting that locus DYS458 may be affected by a higher mutation rate compared to other tetrameric Y-STR loci,<sup>14</sup> the observed microvariants seem to reflect a single mutational event (Table 1).

**Table 1** Statistical parameters of 17 Y-chromosomal short tandem repeat loci in 421 North African male samples

Population	Sample	UH	DH	GD	HD	DC
Morocco	267	247	257	DYS385: 0.887	0.9991	0.963
Algeria	102	88	93	DYS458: 0.806	0.997	0.918
Tunisia	52	29	39	-	-	0.750

\*Abbreviations: UH, unique haplotype; DH, different haplotype; GD, gene diversity; HD, haplotype diversity; DC, discrimination capacity.

In order to get a deeper insight on the Y chromosome diversity in the Maghreb region, 41 Y-SNP were analyzed in a Tunisian population, and 22 biallelic markers respectively in Algeria and Morocco. The haplogroup E-M81, formally, E1b1b1b, E3b1b and E3b2 was the major Y chromosome haplogroup abundantly and respectively found in Morocco, Tunisia and Algeria. This haplogroup is the most common Y chromosome haplogroup in the Maghreb, dominated by its subclade E-M183. It is thought to have originated in the area of North Africa 5,600 years ago.<sup>15,16</sup> This haplogroup reaches a mean frequency of 42% in North Africa, decreasing in frequency from approximately 80% or more in some Moroccan Berber populations, including Saharaw is, to approximately 10% to the east of this range in Egypt.<sup>16-18</sup> Because of its prevalence among these groups and also others such as Mozabite, Middle Atlas, Kabyle and other Berber groups, it is sometimes referred to as a genetic "Berber marker". Pereira et al.,<sup>19</sup> report high levels amongst Tuareg in two Saharan populations - 77.8%

near Gorom-Gorom, in Burkina Faso, and 81.8% from Gosi in Mali. There was a much lower frequency of 11.1% in the vicinity of Tanout in the Republic of Niger. E-M81 is also quite common among North African Arabic-speaking groups. It is generally found at frequencies around 45% in coastal cities of the Maghreb (Oran, Tunis, Tizi Ouzou, Algiers).<sup>11,16</sup> It is also seen, although at low frequency, in the Iberian Peninsula (4%) and Sicily (3%) due to recent gene flow from Northwest Africa through Gibraltar strait.<sup>20,21</sup> The second most abundant haplogroup in the studied population was J1 that showed a decreasing prevalence from Tunisia to Morocco with respectively, 35%, 22.5% and 10%. The J1 haplogroup, defined by the single nucleotide polymorphism (SNP) M267, is most frequent in the Arabian Peninsula especially in Yemen (76%).<sup>22</sup> This could be attributed to the early medieval period during which the Semitic expansion spread J1 out of Arabia into North Africa (Figure 1).<sup>23</sup>



**Figure 1** Distribution and movement of the major E-M215 lineage. E-M215 also known as E1b1b was dated by Cruciani in 2007 to about 22,400 years ago in the Horn of Africa. E-M215 is defined as one of the major patriline ages of humanity, supporting the "out of Africa" theory. Recent discovery showed that the earliest homo sapiens was found in Morocco and has about 300,000 years old. In Africa, E-M215 is distributed in highest frequencies in the Horn of Africa and North Africa, whence it has in recent millennia expanded as far south as South Africa, and northwards into Western Asia and Europe (especially the Mediterranean and the Balkans). Studies on Y chromosome showed that EM81 is the most frequent lineage in North Africa. E-M78 is more present in the east and decline toward the west in contrast with E-M81.6,24-28

## Discussion

The analysis of a large panel of Y-STR and Y-SNP are very useful in population genetics to characterize the male lineage of human populations and constitute a benefic tool to forensic expert. The previous studies showed that the genetic background of Maghrebien populations is very similar and close each other. The genetic diversity has an increasing pattern from East to west of North Africa, seemingly supporting the hypothesis of out of Africa model. Most results showed that the Maghrib is characterized by the predominance of the EM-81 haplogroup that is specific and native to this region. Its presence in some region of southern Europe may be due to the colonization of North African Arabic and berbers to the Iberian Peninsula through the strait of Gibraltar. The forensic values encourage widely the use of these markers in forensic genetics to resolve some complex kinship cases. Also, 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 YSTR 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. This study confirmed that precious informations might come both from YSNPs haplogroup distribution and microsatellite variability to characterize the genetic background of the North West African population.

## Acknowledgements

None.

## Conflict of interest

The author declares that there is no conflict of interest.

## References

1. Jobling MA, Tyler-Smith C. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet.* 2003;4(8):598–612.
2. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature.* 1987;325(6099):31–36.
3. Lucotte G, Guerin P, Halle L, et al. Y chromosome DNA polymorphisms in two African populations. *Am J Hum Genet.* 1989;45(1):16–20.
4. Rosser ZH, Zerjal T, Hurler ME, et al. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet.* 2000;67(6):1526–1543.
5. Underhill PA, Jin L, Lin AA, et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* 7(10):996–1005.
6. Underhill PA, Shen P, Lin AA, et al. Y chromosome sequence variation and the history of human populations. *Nat Genet.* 2000; 26(3):358–361.
7. Yuta H, Sayako K, Noriko S, et al. Analysis of Y chromosome haplogroups in Japanese population using short amplicons and its application in forensic analysis. *Leg Med (Tokyo).* 2014;16(1):20–25.
8. Yi L, Yine H, Xiujiang L, et al. Allele frequency of 19 autosomal STR loci in the Bai population from the southwestern region of mainland China. *Electrophoresis.* 2015;36(19):2498–2503.
9. Butler J. Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. *Forensic Sci Rev.* 2003;15(2):91–111.
10. Onofri V, Alessandrini F, Turchi C, et al. Y-chromosome markers distribution in Northern Africa: High-resolution SNP and STR analysis in Tunisia and Morocco populations. *Forensic Sci Int Genet SS.* 2008;E235–e236.
11. Robino C, Crobu F, Di Gaetano C, et al. Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. *Int J Legal Med.* 2008;122(3):251–255.
12. Aboukhalid R, Bouabdellah M, Abbassi M, et al. Haplotype frequencies for 17 Y-STR loci (AmpFISTR Y-filer) in a Moroccan population sample. *Forensic Sci Int Genet.* 2009;4(3):e73–e74.
13. Aboukhalid R, Bouabdellah M, Lai R, et al. Y chromosomal SNP analysis using the minisequencing strategy in a Moroccan population samples. *Journal of forensic research.* 2010;1(2):1000113.
14. Mulero JJ, Chang CW, Calandro LM, et al. Development and validation of the AmpFISTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system. *J Forensic Sci.* 2006;51(1):64–75.
15. Cruciani F, La Fratta R, Santolamazza P, et al. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet.* 2004;74:1014–1022.
16. Arredi B, Poloni ES, Paracchini S, et al. A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa. *Am J Hum Genet.* 2004;75(2):338–345.
17. Alvarez L, Santos C, Montiel R, et al. Y-chromosome variation in South Iberia: insights into the North African contribution. *Am J Hum Biol.* 2009;21(3):407–409.
18. Bosch E, Calafell F, Comas D, et al. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between north-western Africa and the Iberian Peninsula. *Am J Hum Genet.* 2001;68(4):1019–1029.
19. Pereira Č, Viktor C, Maria S, et al. Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. *Eur J Hum Genet.* 2010;18(8):915–923.
20. Bosch E, Calafell F, Comas D, et al. High Resolution Analysis of Human Y-Chromosome Variation Shows a Sharp Discontinuity and Limited Gene Flow between Northwestern Africa and the Iberian Peninsula. *Am J Hum Genet.* 2001;68(4):1019–1029.
21. Adams SM, Bosch E, Balaesque PL, et al. The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet.* 2008;83(6):725–736.
22. Alshamali F, Pereira L, Budowle B, et al. Local population structure in Arabian Peninsula revealed by Y-STR diversity. *Hum Hered.* 2009;68(1):45–54.
23. Semino O, Magri C, Benuzzi G, et al. 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.* 2004;74(5):1023–1034.
24. Cruciani F, La Fratta R, Trombetta B, et al. Tracing Past Human Male Movements in Northern/Eastern Africa and Western Eurasia: New Clues from Y-Chromosomal Haplogroups E-M78 and J-M12. *Mol Biol Evol.* 2007;24(6):1300–1311.
25. Henn BM, Gignoux C, Lin AA, et al. Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *PNAS.* 2008;105(31):10693–10698.
26. Hassan HY, Underhill PA, Cavalli-Sforza LL, et al. Y-Chromosome Variation Among Sudanese: Restricted Gene Flow, Concordance With Language, Geography, and History. *Am J Phys Anthropol.* 2008;137(3):316–323.
27. Neus SM, Carla GF, Urasin V, et al. Whole Y-chromosome sequence reveal an extremely recent origin of the most common North African paternal lineage E-M183 (81). *Scientific reports.* 2017;15941.
28. Hublin j, Ben-Ncer A, Bailey S, et al. The age of hominin fossils from Jebel Irhoud, Morocco, and the origins of the Middle Stone Age. *Nature.* 2018;558(7657):293–296.