Accepted Manuscript

Title: Rapidly mutating Y-STRs in rapidly expanding populations: Discrimination power of the Yfiler Plus multiplex in northern Africa

Authors: Eugenia D'Atanasio, Giuseppe Iacovacci, Rossella Pistillo, Maria Bonito, Jean-Michel Dugoujon, Pedro Moral, Farha El-Chennawi, Mohammed Melhaoui, Abdellatif Baali, Mohamed Cherkaoui, Daniele Sellitto, Beniamino Trombetta, Andrea Berti, Fulvio Cruciani

PII: S1872-4973(18)30437-X
DOI: https://doi.org/10.1016/j.fsigen.2018.11.002
Reference: FSIGEN 1994

To appear in: *Forensic Science International: Genetics*

Received date: 5 August 2018
Revised date: 2 October 2018
Accepted date: 2 November 2018

Please cite this article as: D'Atanasio E, Iacovacci G, Pistillo R, Bonito M, Dugoujon J-Michel, Moral P, El-Chennawi F, Melhaoui M, Baali A, Cherkaoui M, Sellitto D, Trombetta B, Berti A, Cruciani F, Rapidly mutating Y-STRs in rapidly expanding populations: Discrimination power of the Yfiler Plus multiplex in northern Africa, *Forensic Science International: Genetics* (2018), https://doi.org/10.1016/j.fsigen.2018.11.002

Rapidly mutating Y-STRs in rapidly expanding populations: Discrimination power of the Yfiler Plus multiplex in northern Africa.

Eugenia D'Atanasio[1], Giuseppe Iacovacci[2], Rossella Pistillo[1], Maria Bonito[1], Jean-Michel Dugoujon[3], Pedro Moral[4], Farha El-Chennawi[5], Mohammed Melhaoui[6], Abdellatif Baali[7], Mohamed Cherkaoui[7], Daniele Sellitto[8], Beniamino Trombetta[1], Andrea Berti[9], Fulvio Cruciani[1,8,*].

[1] Dipartimento di Biologia e Biotecnologie "C. Darwin," Sapienza Università di Roma, Rome, Italy

[2] Banca Dati Nazionale del DNA, Direzione Centrale di Polizia Criminale, Rome, Italy

[3] Centre National de la Recherche Scientifique (CNRS), Université Toulouse-3–Paul-Sabatier, Toulouse, France

[4] Department of Evolutionary Biology, Ecology, and Environmental Sciences, Biodiversity Research Institute, University of Barcelona, Spain

[5] Clinical Pathology Department, Faculty of Medicine, Mansoura University, Mansoura, Egypt

[6] Faculté des Sciences, Université Mohammed Premier, Oujda, Morocco

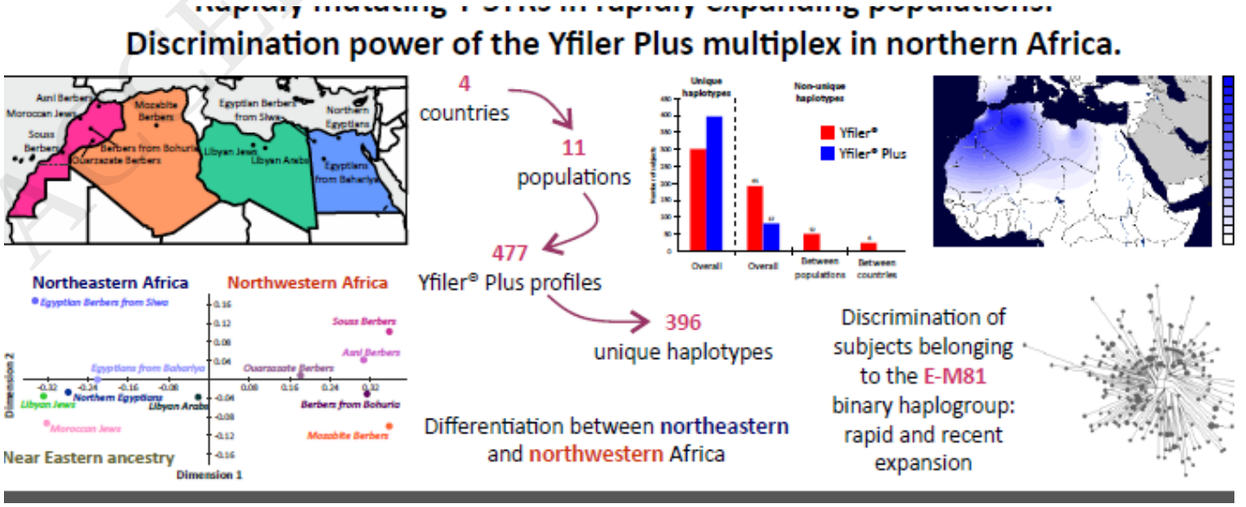
[7] Faculté des Sciences Semlalia de Marrakech (FSSM), Université Cadi Ayyad, Marrakech, Morocco

[8] Istituto di Biologia e Patologia Molecolari, CNR, Rome, Italy

[9] Carabinieri, Reparto Investigazioni Scientifiche di Roma, Sezione di Biologia, Rome, Italy

[*]**Correspondence:** fulvio.cruciani@uniroma1.it

GraphicalAbstract



Highlights

- The Y-STR profiles of 477 northern African males were obtained using the 27-loci Yfiler® Plus kit.
- About 46% of the samples belonged to the recent E-M81 Y haplogroup.
- The AMOVA analysis showed significant differences between populations and countries.
- The number of identical Y-STR profiles was drastically reduced using the Yfiler® Plus compared to Yfiler.

Abstract

The male-specific northern African genetic pool is characterised by a high frequency of the E-M81 haplogroup, which expanded in very recent times (2-3 kiloyears ago). As a consequence of their recent coalescence, E-M81 chromosomes often cannot be completely distinguished on the basis of their Y-STR profiles, unless rapidly-mutating Y-STRs (RM Y-STRs) are analysed. In this study, we used the Yfiler® Plus kit, which includes 7 RM Y-STRs and 20 standard Y-STR, to analyse 477 unrelated males coming from 11 northern African populations sampled from Morocco, Algeria, Libya and Egypt. The Y chromosomes were assigned to monophyletic lineages after the analysis of 72 stable biallelic polymorphisms and, as expected, we found a high proportion of E-M81 subjects (about 46%), with frequencies decreasing from west to east. We found low intra-population diversity indexes, in particular in the populations that experienced long-term isolation. The AMOVA analysis showed significant differences between the countries and between most of the 11 populations, with a rough differentiation between northwestern Africa and northeastern Africa, where the Egyptians Berbers from Siwa represented an outlier population. The comparison

between the Yfiler® and the Yfiler® Plus network of the E-M81 Y chromosomes confirmed the high power of discrimination of the latter kit, thanks to higher variability of the RM Y-STRs: indeed, the number of chromosomes sharing the same haplotype was drastically reduced from 201 to 81 and limited, in the latter case, to subjects from the same population.

Keywords: Yfiler® Plus, E-M81 haplogroup, Northern African populations, Y-STRs

Introduction

The northern African genetic landscape has been shaped by different migration waves over the past thousands of years. The first human settlements date back to 160 kya (kiloyears ago) and several cultures developed in Northern Africa over time, such as the Aterian, Iberomaurusian and the Capsian culture [1–4]. In historical times, the region was populated by different groups, such as the Ancient Egyptians, the Phoenicians, the Romans and the Vandals, up to the Arab invasion from the Arabian Peninsula in the 7th century C. E. [5]. These recurring waves of migrations, settlements and replacements can be ascribed to the strategic geographic position of northern Africa, which is delimited by two permeable geographic barriers (the Mediterranean sea to the north and the Sahara desert to the south) and it is a crossroad between three continents: Africa, Europe and Asia. A consequence of these events is that the Northern African gene pool can be considered a mixture of different sources, which are hardly distinguishable: for example, considering the data obtained from autosomes, there is not a clear differentiation between Berbers and non-Berbers groups and there is not a strict correlation between geographical and genetic features [6,7]. In this context, the male specific region of the human Y chromosome (MSY) can be disproportionally informative, because of its higher inter-population variability. This is due to the fact that the MSY lacks the meiotic crossing-over and is haploid: these features determine the co-inheritance of markers and an enhanced effect of the genetic drift on the Y chromosome compared to the autosomes. A result of this is that the main source of variability in the MSY is the accumulation of new mutations, especially single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs). The Y-SNPs are characterised by a low mutation rate (~ 0.76×10^{-9} site/year [8–10]), so they can be considered stable variants and can be arranged in monophyletic entities called “haplogroups”, which in turn can be gathered in a phylogenetic tree [11].

Differently from the Y-SNPs, the Y-STRs are characterised by a higher mutation rate and this feature can be exploited for forensic purposes when a male lineage differentiation is required, for example in unbalanced male-female mixtures [12,13]. The most recent multiplex for the capillary electrophoretic analysis of Y-STRs is the Yfiler® Plus, which is composed of 27 loci, including 7 rapidly-mutating STRs (RM-STRs). The RM-STRs, thanks to their high mutation rate (higher than 1×10^{-2}) [14], can allow to distinguish about ¼ of father-son pairs and ½ of brother pairs [15,16]. The inclusion of the RM-STRs in the Yfiler® Plus multiplex increased the level of within-population haplotype diversity and decreased (ideally to zero) the haplotype sharing between males from different populations [17–19]. Over the past years, an increasing number of populations (especially from Eurasia) has been analysed exploiting the Yfiler® Plus kit, but information about northern African populations is still lacking.

One of the most common northern African haplogroups is E-M81, which expanded in northern Africa in recent times and shows increasing frequencies from east to west, while the Y-STR variability decreases in the same direction [20–22]. The recent coalescence age of this haplogroup made difficult to discriminate between Y chromosomes from the same population and also between subjects from different groups or countries [22]. Our aim was to evaluate the power of the RM Y-STRs in discriminate between northern African males carrying the same binary haplogroup. To this end, we reported the Yfiler® Plus haplotypes for 477 male subjects belonging to 11 ethnic groups sampled from four northern African countries (Morocco, Algeria, Libya and Egypt). We also characterised the microvariant alleles by Sanger sequencing and we framed our samples in their phylogenetic context by analysing 72 stable Y-SNPs. In general, we observed relatively low levels of intra-population diversity, while the inter-population differentiation varied on the basis of the groups considered. The inclusion of the RM Y-STRs in the analysis led to a better differentiation between subjects: considering only the E-M81 Y chromosomes, we found a

greater number of unique haplotypes compared to the Yfiler® results and we observed haplotype sharing only within the same population.

Material and methods

DNA samples

We analysed the 27 Y-STRs of the Yfiler® Plus kit and 72 Y-SNPs in 477 northern African males, belonging to eleven populations from Morocco, Algeria, Libya and Egypt (Fig. 1 and Supplementary Table 1).

All the populations have been analysed in previous studies [10,23–28] for some of the Y-SNPs here reported and differences in sample sizes for some ethnic groups is due to subsequent DNA unavailability. The DNA samples from Moroccan Jews and Libyan Jews (31 and 24 subjects, respectively) came from the National Laboratory for the Genetics of Israeli Populations. The study was approved by the "Sapienza Università di Roma" ethical committee (document number 2755/15) and by the Carabinieri General Headquarters under the Ministry of Defense. Informed consent was obtained for all the DNA donors.

DNA extraction and quantification

Genomic DNA was extracted from blood, saliva or lymphoblastoid cell lines using standard procedures and quantified using the Quantifiler Trio Quantification Kit (ThermoFisher Scientific, Waltham, MA, USA).

SNP genotyping

The Y haplogroup affiliation of the 477 males here reported have been defined analysing 72 informative biallelic markers, selected following a hierarchical approach based on the known MSY phylogeny [26–32]. Most of the Y-SNPs have been published in previous studies [23–31,33] and analysed by means of DHPLC, RFLP or Sanger sequencing, after a PCR amplification step.

Y-STR multiplex genotyping

The analysis of the 27 Y-STRs have been performed from 1 ng of genomic DNA using the Yfiler® Plus amplification kit, according to the manufacturer's protocol. The amplified fragments have been analysed with GeneMapper IDX v.1.4, after an electrophoretic run on a 3500 XL Genetic Analyzer.

The haplotypes of the 477 male subjects (Supplementary Table 2) have been submitted to the YHRD (Y-chromosomal Haplotype Reference Database, https://yhrd.org) under the accession numbers YA004351-YA004356 (release R58).

Y-STR microvariant analysis

We observed microvariant alleles at five loci (DYS458, DYS448, DYS385, DYS449 and DYF387S1), which have been checked by PCR and Sanger sequencing (Supplementary Table 3). We were not able to analyse one microvariant (DYF387S1.1), because of the DNA unavailability of the

two samples of interest. Three out of five Y-STRs (DYS458, DYS449 and DYF387S1) have been previously analysed in eastern African samples and we used the same PCR primers [18]. The PCR primers for the other two loci have been designed using Primer3 software (http://bioinfo.ut.ee/primer3-0.4.0) on the basis of the Y chromosome reference sequence (Human Feb. 2009 - GRCh37/hg19 – Assembly) reported in the UCSC Genome Browser (http://genome.ucsc.edu) (Supplementary Table 4). The PCR reaction, purification, Sanger sequencing and chromatogram alignment have been performed as reported in Iacovacci et al. [18].

Forensic and statistical analysis

For the estimation of all the forensic and statistical parameters and for the network construction (see next section), we converted the DYS389II alleles to the DYS389B nomenclature for each subject, subtracting the number of repeats at the DYS389I from the number of repeats at DYS389II.

For each locus, we calculated the genetic diversity (GD, Supplementary Table 5) as described by Nei and Tajima [34]. We also estimated the standard population genetic parameters. The haplotype diversity (HD) was calculated according to the formula $HD = n(1-\Sigma p_i^2)/(n-1)$, where n is the sample size and p_i is the frequency of the i^{th} haplotype. The haplotype match probability (HMP) was calculated as $HMP = \Sigma p_i^2$. The discrimination capacity (DC) corresponds to the ratio between the number of different haplotypes and the total number of haplotypes.

We used Arlequin software (ver. 3.5.2.2) [35] to perform the analysis of molecular variance (AMOVA), grouping the subjects in two hierarchical levels (individuals assigned to 11 populations and populations assigned to 4 countries). For this analysis, the two multi-locus markers DYS385 and DYF387S1 were discarded, the null or supernumerary alleles at the other loci were coded as

missing data and the microvariant alleles were coded as the number of complete repeats on the basis of the sequencing data. We performed the AMOVA both considering or not the molecular distance between haplotypes (Φ-statistics or F-statistics, respectively). For the Φ-statistics, we measured the molecular distance as the sum of the number of repeat difference between two haplotypes. The significance levels for both Φ_{ST} or F_{ST} parameters were calculated comparing the actual value with the distribution of 1,000 randomized values.

Network construction, time estimates and frequency map

We used the Network 5.0 software to reconstruct a median joining network representing the phylogenetic relations among the Yfiler® Plus haplotypes [36,37]. The two multi-locus systems DYS385 and DYF387S1 were arbitrarily subdivided, assigning the smaller allele and the larger allele to two different allelic series.

The time estimates of two northern African haplogroups (E-M81 and E-V65) were performed using the ASD (Average Square Distance) method. We used two different groups of Y-STRs: 1) a set of 22 loci, including all the Yfiler® Plus markers, with the exclusion of three complex loci (the two multi-copy loci DYS385 and DYF387S1 and the interrupted marker DYS389II); 2) a set of 17 loci, after removing the RM Y-STRs from the previous 22 markers. For the time estimates, we used a mutation rate of 5.55×10^{-3} per marker per generation and 3.08×10^{-3} per marker per generation for the 22 loci set and the 17 loci set, respectively. These figures were obtained averaging the known germ line mutation rates [14,15] of all the loci included in each set.

Frequency maps were drawn on a grid with 100 rows × 78 columns using the Kriging method implemented by the Surfer 6.0 software (Golden Software, Inc., Golden, CO, USA), including the frequency data for other African and Eurasian populations reported in ref. [28].

Results and Discussion

Y haplogroup distribution in the 11 northern African populations

We analysed 72 Y-SNPs in order to reconstruct a low-resolution Y-tree to frame the Y-STR variability in a stable phylogenetic context. We identified 31 binary haplogroups or paragroups in our sample, as shown in Fig. 2.

The most frequent Y lineages in northern Africa were found to be E-M81, E-M78 (in particular, its internal clade E-V65), haplogroup J and R-V88 (Supplementary Table 2), as already observed in other studies [10,20–22,24–26,28]. In the whole set of subjects, these haplogroups reached a frequency of about 46%, 8%, 15% and 4%, respectively. Considering the distribution of these clades in the northern African populations here analysed, we observed that R-V88 was almost exclusively found in the Egyptians from Siwa oasis, which showed other peculiarities compared to the other Egyptian groups (such as the relative high proportion of B-M150 Y chromosomes). E-M78 was mainly found in the eastern countries of northern Africa (Egypt and Libya), while it is less frequent in the north-western regions [24,25]. Within this haplogroup, the most frequent sub-clades were found to be E-V65 (~ 3%) and E-V22 (~ 3%), while other lineages are less frequent (E-V13 and E-V12*, both with a frequency slightly higher than 1%) or absent (such as the eastern African haplogroup E-V32 [10,18,24,25]). The northern African haplogroup J subjects were found to belong to two main sub-clades, J-M172 (~ 6%) and J-M267 (~ 9%), whose frequencies decreased from east to west [21]. E-M81 showed the opposite clinal pattern, with its frequencies that increased moving westward (> 1% in Egypt, 66% in Morocco).

Single-locus diversity indexes

Genotypes for the 23 single-copy loci and 2 multi-copy loci were obtained for all samples, with the exclusion of three subjects showing a null allele at three different loci (DYS389II, DYS448 and DYF387S1). The genetic diversity (GD) values for the 25 loci analysed are reported in Supplementary Table 5. The highest overall GD values are observed at DYS385 and DYS481 (0.86 and 0.85, respectively). In contrast to previous observations that reported the highest GD values in both multi-copy loci DYS385 and DYF387S1, in our set of data DYF387S1 showed the fourth highest value, which was comparable to the GD values of the other RM Y-STRs. The low overall GD observed at DYF387S1 is mainly due to the low GD value reported for this locus in the single Algerian population analysed in this study (Mean GD: ~ 0.42). Indeed, the allelic range for DYF387S1 in the Algerian Mozabite Berbers is quite narrow (from 35 repeats to 39 repeats), in contrast to the wider allelic range observed for the other countries (33-41 in Morocco, 34-42 in Libya and 33-45.2 in Egypt). In general, the Algerian population showed the lowest GD values for all loci but one (DYS635). These results can be explained considering two main factors: 1) in our sample, the Algerian populations is the only one composed entirely by a single ethnic group (Fig. 1 and Supplementary Table 1); 2) the large majority of the Algerians (55 out of 64 subjects, ~ 86%) were found to belong to the recent E-M81 haplogroup [22]. On the contrary, the highest GD values for each locus are observed in Egypt and, to a lesser extent, in Libya, where we also observed a higher Y-haplogroup diversification (Fig. 2 and Supplementary Table 2).

Molecular analysis of microvariants

We observed microvariant alleles at five different loci: DYS458, DYS448, DYS385, DYS449 and DYF387S1 (Supplementary Table 2). We analysed these microvariants by Sanger resequencing of the Y chromosomes involved (Supplementary Table 3 and Supplementary Table 4).

For the DYS458 marker, we observed the DYS458.2 variant in 44 subjects, which has been already described and linked to the J-M267 Y haplogroup [38,39]. Consistently with the data reported in the literature, we found a perfect association between the DYS458.2 variant and the haplogroup J-M267 (Supplementary Table 2). We confirmed the molecular structure of this microvariant by resequencing 5 out of the 44 DYS458.2 subjects (Supplementary Table 3).

DYS448 showed two different microvariants in our set: we observed a DYS448.2 variant in a northern Egyptian and a DYS448.4 variant in a Moroccan Jew. The DYS448.2 allele was due to a dinucleotide deletion in the non-repetitive portion of the Y-STR locus, while the DYS448.4 variant was caused by a more complex rearrangement, involving two short deletions in different positions. Moreover, the Moroccan subject with the DYS448.4 allele also showed a single nucleotide substitution in the non-repetitive core of the locus (Supplementary Table 3). The differences in the molecular features of these two alleles and in the geographic origin and haplogroup affiliation of the subjects suggested that the two microvariants originated independently.

The DYS385 marker has been analysed using two pairs of PCR primers, in order to distinguish the two different copies of the locus. Thanks to this approach we have been able to describe the DYS385.2 microvariant, which was due to a dinucleotide deletion in the $(GAAA)_n$ repetitive motif, and to assign it to a specific copy of the locus (Supplementary Table 3).

DYS449 is a RM Y-STR with a structure characterised by different elements (repetitive and non-repetitive) with a variable length. This complex organisation is the cause of several molecularly different microvariants reported in the literature for samples belonging to different haplogroups [18,40]. In the present study, this locus showed a new DYS449.2 variant in a northern Egyptian belonging to the Y haplogroup E-M123* (Supplementary Table 2 and Supplementary Table 3).

In the present study, we reported two different microvariant alleles for the multi-copy locus DYF387S1. We observed the DYF387S1.2 allele in 18 Egyptians from Siwa. This microvariant was already described in a previous study [18] and, considering the phylogenetic relationships among the subjects, it was proposed that it originated in the Y haplogroup B-M182 background. Using the same experimental approach (Supplementary Table 4), we could distinguish the two copies of the marker and we could recognise that the microvariant had the same molecular structure and arose in the same DYF387S1 copy as already described [18] (Supplementary Table 3). Moreover, our DYF387S1.2 samples all belong to the B-M182/M150 haplogroup, confirming the correlation between this Y-STR microvariant and the B-M182 haplogroup. We also found a DYF387S1.1 allele in two northern Egyptians of haplogroup J-M267. We could not perform the molecular characterisation of this microvariant because of DNA unavailability. However, the samples were identical for all the 25 Y-STR loci but one (DYS518), which differed for only one repetitive unit. These features suggest that the DYF387S1.1 microvariant probably originated in a common ancestor of these two subjects.

Intra-population Y-STR haplotype variability

The full list of 477 Yfiler® Plus haplotypes/haplogroups is reported in Supplementary Table 2. We calculated the standard forensic parameters using the variability of the 27 Y-STR loci of this multiplex (Table 1).

Considering the whole northern African population, we observed a DC value (0.9078) that is in the lower range of the DC values reported for other populations analysed with the same multiplex [17,19,41]. However, the northern African DC is comparable to the DC reported for the whole eastern Africa, which is the only other African group analysed so far with the 27 Y-STR kit

[18]. It is also worth noting that the DC values obtained were found to be higher than the results obtained for the same subjects using only the 17 loci of the Yfiler® (overall DC values: 0.9078 *vs.* 0.7358; number of different haplotypes: 433 *vs.* 351) (Supplementary Table 6), confirming the higher power of resolution of the Yfiler® Plus. The greater deviation between the DC figures obtained with the two Y-STR multiplexes is observed for the Berbers of Bouhria, which showed an increase in the DC value of about 44%. Nevertheless, this ethnic group showed the lowest DC value using both Y-STR multiplexes, along with the Asni Berbers (Yfiler® Plus DC = 0.8868; increase compared to the Yfiler® DC: ~ 15%). In this context, it is worth noting that the other people with a low DC value were the Berbers from the Siwa oasis in Egypt, which are also characterised by a less variable Y haplogroup background compared to the neighbouring Egyptian populations (Fig. 2) [10,26]. The quite low number of different haplotypes in these populations are probably due to their prolonged isolation, which increased the effects of the genetic drift, causing a lower intra-population variability.

Among the 433 different haplotypes reported in our sample collection (Table 1), 396 were unique, while the other 37 were observed in two or more subjects (Supplementary Table 7). These haplotypes were shared only between subjects from the same ethnic group, with no sharing between populations nor between countries. On the contrary, the Yfiler® multiplex was not able to distinguish several samples from different populations or different countries (Fig. 3).

The comparison between the data obtained with the Yfiler® Plus and the Yfiler® confirmed the higher power of discrimination of the former kit, thanks to the much higher variability introduced by the RM Y-STRs. However, it must be noted that Yfiler® Plus failed to completely

distinguish subjects, especially from populations that are characterised by isolation or are organised in a patrilineal society, as already observed in the patrilineal clans in eastern Africa [18].

Inter-population Y-STR haplotype variability

In order to understand the degree of genetic differentiation between the northern African populations here analysed, we performed an AMOVA analysis both considering or not the molecular differences among haplotypes (Φ-statistics and F-statistics, respectively) (Table 2).

Considering the whole northern Africa, we could observe a significant differentiation between populations and between countries, considering both the Φ-statistics and F-statistics ($p < 0.05$). The differences were found to be significant also in the pair-wise analysis between the four countries (Supplementary Table 8 and Supplementary Table 10), and in most of the pair-wise comparisons of the 11 populations (Supplementary Table 9 and Supplementary Table 11). It is worth noting that, with the exception of the Jews groups, the few pairs of populations that resulted not to be differentiated were from the same country. We used the Φ_{ST} distance values to draw a multidimensional scaling (MDS) plot (Fig. 4).

The first dimension separated northwestern Africa (Morocco and Algeria; to the right) and northeastern Africa (Libya and Egypt; to the left). The exception to this general pattern was represented by the Moroccan Jews, which lied in the left part of the graph, as expected considering the Φ_{ST} values (Supplementary Table 11). This group showed a first dimension position similar to the Libyan Jews and the observed proximity of the two Jews populations to the northeastern African groups can be interpreted as a consequence of a near eastern ancestry [42]. Indeed, recent studies observed a higher near eastern component in the Egyptians compared to the other northern African groups [43,44], which was probably due to the prolonged contacts and the geographical proximity between Egypt and the Near East. With the exclusion of the separation along an east-to-west axis, the plot did not highlight any ethno-geographic sub-structure within the two northern African macro-regions. The northwestern samples were scattered in the right part of the graph and the northeastern groups did not form any clusters as well. The Egyptians Berbers from Siwa can be considered an outlier, being differentiated from the other populations

by both dimensions of the MDS plot, and this was in line with the peculiar Y-SNP haplogroup composition showed by these people (Fig. 2). It is worth noting that the Egyptian Berbers from Siwa also showed a mtDNA haplogroup composition different from neighbour populations [45]. The genetic peculiarities of the Siwa Berbers can be ascribed to the several human groups that have been settled in this oasis, which is a cross-road between Near East, eastern Africa, central Sahel and northern Africa. The populations analysed in this study did not include any Moroccan or Algerian Arab groups. However, we analysed both Egyptian and Libyan Arabs (Supplementary Table 1) and they did not show clear differences with the Berber groups. This finding was consistent with similar results obtained from genome-wide data [7] and from autosomal and Yfiler STR data [46,47], which showed a substantial lack of differentiation between Berbers and non-Berbers.

The effect of E-M81 distribution on the Y-STR diversity in northern Africa

Our results pointed to a low degree of intra-population differentiation and to a variable degree of inter-population diversity, with the observed differences following an east-to-west axis. For example, the Y-STR diversity decreased from east to west, as shown by the DC values, which showed the lowest figures in Morocco and the highest ones in Libya (Table 1). Also the MDS plot highlighted that the northern African groups are differentiated along an east-to-west direction (Fig. 4). These results are mirrored by the geographic distribution of the E-M81 haplogroup, whose frequencies decline from northwestern to northeastern Africa (Fig. 5) and whose Y-STR variability decreases in the opposite direction [20–22].

This distribution has been usually explained as a consequence of a westward expansion from the Near East and this event probably occurred in recent times, possibly about 2 kya [22]. We calculated the forensic indexes including only the E-M81 subjects and we could observe an even more enhanced clinal pattern (Moroccan DC = 0.87; Libyan DC = 0.96; the values for the Egyptian populations cannot be considered informative because our collection included only four E-M81 chromosomes from this country) (Supplementary Table 12).

We reconstructed the median joining network for the 220 Y chromosomes belonging to the E-M81 haplogroup, using both the Yfiler® and the Yfiler® Plus loci (Fig. 6).

The complete versions of the E-M81 network, represented in the insets of Fig. 6, were characterised by a clear star-like structure, as already observed for this haplogroup [22]. This topology is indicative of a rapid and recent expansion, that usually leads to a frequency increase of a lineage with a low degree of the internal genetic diversity. In order to check this scenario, we estimated the TMRCA of our E-M81 subjects. In the analysis, we used two different subsets of Y-STRs (22 loci and 17 loci, as described in Material and Methods) and the germ-line mutation rate, because it has been observed that this rate led to better time estimates for recent events [48,49]. Considering the problems linked to the time estimates obtained from the Y-STR data [49,50], we also estimated the TMRCA for the E-V65 haplogroup, which was precisely dated to 1.47 kya using a calibrated Y-SNP mutation rate [10] (Table 3).

Comparing the Y-SNP and the Y-STR coalescence ages for the E-V65 haplogroup, we noted that the use of the 17 Y-STR set, which exclude the RM Y-STRs, produced a closer estimate, although even the 22 Y-STR set led to a comparable figure. Our E-M81 time estimate based on 17

Y-STRs was of about 2 kya, a figure very similar to the TMRCA obtained for the E-M183 sub-clade (Table 3), consistently with the observation that E-M183 captures nearly all E-M81 Y-chromosomes [22]. The slight difference between the two time estimates and the reported older age of the E-M183 sub-haplogroup can be due to: i) a difference in the dating methods (Rho *vs.* ASD); ii) a difference in the class of used loci (Y-SNPs *vs.* Y-STRs); iii) a difference in the number of used loci (13 Y-STRs *vs.* 17 Y-STRs) [52]. However, the observed data are in line with a very recent expansion of the E-M81 (and E-V65) haplogroup in northern Africa.

The homogenizing effect of the E-M81 spread was testified also by the lack of geographic clusters in the networks, with subjects from different populations or different countries splitting from the same node. There was only one exception, represented by the Algerian Berbers, which formed a recognizable group including the large majority of the Y chromosomes from this population. This cluster was more pronounced in the Yfiler® network (Fig. 6, panel a) rather than in the Yfiler® Plus network (Fig. 6, panel b). In this context, it must be noted that the Algerian Mozabite Berbers experienced a long isolation [43,53], so their clustered distribution in both networks may reflect the effects of the genetic drift, which led to the reduction of their Y-STR diversity (average gene diversity over loci: 0.4194; Table 1).

The highest discrimination capacity of the Yfiler® Plus kit compared to the Yfiler® multiplex was highlighted in the reduced versions of the networks, where only the shared haplotypes are represented. Comparing the two plots, we could note that the E-M81 Yfiler® network contained more haplotypes shared by different samples, which often came from different populations and, in some cases, from different countries (Fig. 6, panel a). For example, there was a haplotype observed in three Moroccan groups (in different shades of purple), in the Algerian Mozabite Berbers (in orange) and in the Libyan Arabs (in dark green). On the contrary, this was not the case for the E-M81 Yfiler® Plus network, that was characterised by a better differentiation of the Y

chromosomes, with only few haplotypes shared between samples from the same population (Fig. 6, panel b).

Conclusions

Northern Africa showed complex genetic features, which cannot be approximated by the results obtained from only one or few ethnic groups, highlighting the importance of the analysis of several populations from different countries to make forensic inferences. Moreover, we found that the RM Y-STRs can be disproportionally informative to discriminate subjects coming from groups characterised by low intra-population variability because of a recent and rapid expansion. In particular, we observed that a large proportion of samples belonging to a recently expanded haplogroup such as E-M81 could not be adequately distinguished using the Yfiler® set of markers, which do not include any rapidly mutating loci. The inclusion of 7 RM Y-STRs in the Yfiler® Plus kit allowed the differentiation of most of these subjects, however the discrimination capacity is still not powerful enough to properly distinguish all the individuals. The inclusion of additional RM Y-STRs can be helpful to obtain a complete differentiation of the subjects, in particular for samples coming from populations characterised by isolation or organised in a patrilocal/patrilineal society.

Acknowledgements

We thank all the donors for providing DNA samples. We are grateful to all the researchers that helped us in the sample collection: A. Cambon-Thomsen, M.S. Issad, G. Hariti and M. Ziani for the collection of the Algerian sample; E. B. Colomb for the collection of the Egyptian Berbers samples; the National Laboratory for the Genetics of Israeli Populations for the Moroccan and Libyan Jews samples. The sampling of the Berbers was made within the framework of the Inserm "Réseau Nord/ Sud" No. 490NS1 (Mozabite Berbers), "The Origin of Man, Language and Languages", EUROCORES Programme and benefited from funding by the Région Midi-Pyrénées (Toulouse, France), the CNRS, and the E.C. Sixth Framework Programme under Contract ERASCT-

2003-980409. This work was supported by the Sapienza University of Rome (grant number C26A153PCN to FC) and by the National Geographic Society (grant number HJ-139-R17 to FC).

References

[1] E.A.A. Garcea, The spread of Aterian people in North Africa., in: E.A.A. Garcea (Ed.), South-Eastern Mediterr. Peoples between 130,000 10,000 Years Ago, Owbow Books, Oxford, 2010.

[2] C. Hunt, J. Davison, R. Inglis, L. Farr, T. Reynolds, D. Simpson, H. el-Rishi, G. Barker, Site formation processes in caves: The Holocene sediments of the Haua Fteah, Cyrenaica, Libya, J. Archaeol. Sci. 37 (2010) 1600–1611. doi:10.1016/j.jas.2010.01.021.

[3] R.N.E. Barton, A. Bouzouggar, J.T. Hogue, S. Lee, S.N. Collcutt, P. Ditchfield, Origins of the iberomaurusian in NW Africa: New AMS radiocarbon dating of the middle and later stone age deposits at taforalt cave, Morocco, J. Hum. Evol. 65 (2013) 266–281. doi:10.1016/j.jhevol.2013.06.003.

[4] E.M.L. Scerri, The Aterian and its place in the North African Middle Stone Age, Quat. Int. 300 (2013) 111–130. doi:10.1016/j.quaint.2012.09.008.

[5] J.L. Newman, The peopling of Africa : a geographic interpretation, Yale Univ. Press, 1995.

[6] B.M. Henn, L.R. Botigué, S. Gravel, W. Wang, A. Brisbin, J.K. Byrnes, K. Fadhlaoui-Zid, P.A. Zalloua, A. Moreno-Estrada, J. Bertranpetit, C.D. Bustamante, D. Comas, Genomic Ancestry of North Africans Supports Back-to-Africa Migrations, PLoS Genet. 8 (2012) e1002397. doi:10.1371/journal.pgen.1002397.

[7] L.R. Arauna, J. Mendoza-Revilla, A. Mas-Sandoval, H. Izaabel, A. Bekada, S. Benhamamouch, K. Fadhlaoui-Zid, P. Zalloua, G. Hellenthal, D. Comas, Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa., Mol. Biol. Evol. 34 (2017) 318–329. doi:10.1093/molbev/msw218.

[8] Q. Fu, H. Li, P. Moorjani, F. Jay, S.M. Slepchenko, A.A. Bondarev, P.L.F. Johnson, A. Aximu-

Petri, K. Prüfer, C. de Filippo, M. Meyer, N. Zwyns, D.C. Salazar-García, Y. V. Kuzmin, S.G. Keates, P.A. Kosintsev, D.I. Razhev, M.P. Richards, N. V. Peristov, M. Lachmann, K. Douka, T.F.G. Higham, M. Slatkin, J.-J. Hublin, D. Reich, J. Kelso, T.B. Viola, S. Pääbo, Genome sequence of a 45,000-year-old modern human from western Siberia, Nature. 514 (2014) 445–449. doi:10.1038/nature13810.

[9] M. Karmin, L. Saag, M. Vicente, M.A.W. Sayres, M. Järve, U.G. Talas, S. Rootsi, A.-M. Ilumäe, R. Mägi, M. Mitt, L. Pagani, T. Puurand, Z. Faltyskova, F. Clemente, A. Cardona, E. Metspalu, H. Sahakyan, B. Yunusbayev, G. Hudjashov, M. DeGiorgio, E.-L. Loogväli, C. Eichstaedt, M. Eelmets, G. Chaubey, K. Tambets, S. Litvinov, M. Mormina, Y. Xue, Q. Ayub, G. Zoraqi, T.S. Korneliussen, F. Akhatova, J. Lachance, S. Tishkoff, K. Momynaliev, F.-X. Ricaut, P. Kusuma, H. Razafindrazaka, D. Pierron, M.P. Cox, G.N.N. Sultana, R. Willerslev, C. Muller, M. Westaway, D. Lambert, V. Skaro, L. Kovačevic´, S. Turdikulova, D. Dalimova, R. Khusainova, N. Trofimova, V. Akhmetova, I. Khidiyatova, D. V. Lichman, J. Isakova, E. Pocheshkhova, Z. Sabitov, N.A. Barashkov, P. Nymadawa, E. Mihailov, J.W.T. Seng, I. Evseeva, A.B. Migliano, S. Abdullah, G. Andriadze, D. Primorac, L. Atramentova, O. Utevska, L. Yepiskoposyan, D. Marjanovic´, A. Kushniarevich, D.M. Behar, C. Gilissen, L. Vissers, J.A. Veltman, E. Balanovska, M. Derenko, B. Malyarchuk, A. Metspalu, S. Fedorova, A. Eriksson, A. Manica, F.L. Mendez, T.M. Karafet, K.R. Veeramah, N. Bradman, M.F. Hammer, L.P. Osipova, O. Balanovsky, E.K. Khusnutdinova, K. Johnsen, M. Remm, M.G. Thomas, C. Tyler-Smith, P.A. Underhill, E. Willerslev, R. Nielsen, M. Metspalu, R. Villems, T. Kivisild, A recent bottleneck of Y chromosome diversity coincides with a global change in culture, Genome Res. 25 (2015) 459–466. doi:10.1101/gr.186684.114.

[10] E. D'Atanasio, B. Trombetta, M. Bonito, A. Finocchio, G. Di Vito, M. Seghizzi, R. Romano, G. Russo, G.M. Paganotti, E. Watson, A. Coppa, P. Anagnostou, J.-M. Dugoujon, P. Moral, D.

Sellitto, A. Novelletto, F. Cruciani, The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages, Genome Biol. 19 (2018) 20. doi:10.1186/s13059-018-1393-5.

[11] M. Jobling, E. Hollox, M. Hurles, T. Kivisild, C. Tyler-Smith, Human Evolutionary Genetics, Second Edi, Garland Science, 2013.

[12] P. De Knijff, M. Kayser, A. Caglià, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, L. Roewer, Chromosome Y microsatellites: Population genetic and evolutionary aspects, Int. J. Legal Med. 110 (1997) 134–140. doi:10.1007/s004140050052.

[13] L. Roewer, Y chromosome STR typing in crime casework, Forensic Sci. Med. Pathol. 5 (2009) 77–84. doi:10.1007/s12024-009-9089-5.

[14] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. von Wurmb-Schwark, P. de Knijff, D. Labuda, H. Vézina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications, Am. J. Hum. Genet. 87 (2010) 341–353. doi:10.1016/j.ajhg.2010.08.006.

[15] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. De Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208–218. doi:10.1016/j.fsigen.2011.04.017.

[16] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, J. Balažic, J.

Ballantyne, D.J. Ballard, B. Berger, C. Bobillo, M. Bouabdellah, H. Burri, T. Capal, S. Caratti, J. Cárdenas, F. Cartault, E.F. Carvalho, M. Carvalho, B. Cheng, M.D. Coble, D. Comas, D. Corach, M.E. D'Amato, S. Davison, P. de Knijff, M.C.A. De Ungria, R. Decorte, T. Dobosz, B.M. Dupuy, S. Elmrghni, M. Gliwiński, S.C. Gomes, L. Grol, C. Haas, E. Hanson, J. Henke, L. Henke, F. Herrera-Rodríguez, C.R. Hill, G. Holmlund, K. Honda, U.D. Immel, S. Inokuchi, M.A. Jobling, M. Kaddura, J.S. Kim, S.H. Kim, W. Kim, T.E. King, E. Klausriegler, D. Kling, L. Kovačević, L. Kovatsi, P. Krajewski, S. Kravchenko, M.H.D. Larmuseau, E.Y. Lee, R. Lessig, L.A. Livshits, D. Marjanović, M. Minarik, N. Mizuno, H. Moreira, N. Morling, M. Mukherjee, P. Munier, J. Nagaraju, F. Neuhuber, S. Nie, P. Nilasitsataporn, T. Nishi, H.H. Oh, J. Olofsson, V. Onofri, J.U. Palo, H. Pamjav, W. Parson, M. Petlach, C. Phillips, R. Ploski, S.P.R. Prasad, D. Primorac, G.A. Purnomo, J. Purps, H. Rangel-Villalobos, K. Reogonekbała, B. Rerkamnuaychoke, D.R. Gonzalez, C. Robino, L. Roewer, A. Rosa, A. Sajantila, A. Sala, J.M. Salvador, P. Sanz, C. Schmitt, A.K. Sharma, D.A. Silva, K.J. Shin, T. Sijen, M. Sirker, D. Siváková, V. Škaro, C. Solano-Matamoros, L. Souto, V. Stenzl, H. Sudoyo, D. Syndercombe-Court, A. Tagliabracci, D. Taylor, A. Tillmar, I.S. Tsybovsky, C. Tyler-Smith, K.J. van der Gaag, D. Vanek, A. Völgyi, D. Ward, P. Willemse, E.P.H. Yap, R.Y.Y. Yong, I.Z. Pajnič, M. Kayser, Toward Male Individualization with Rapidly Mutating Y-Chromosomal Short Tandem Repeats, Hum. Mutat. 35 (2014) 1021–1032. doi:10.1002/humu.22599.

[17] J.K. Olofsson, H.S. Mogensen, A. Buchard, C. Børsting, N. Morling, Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® Plus PCR amplification kit, Forensic Sci. Int. Genet. 16 (2015) 232–236. doi:10.1016/j.fsigen.2015.02.006.

[18] G. Iacovacci, E. D'Atanasio, O. Marini, A. Coppa, D. Sellitto, B. Trombetta, A. Berti, F. Cruciani, Forensic data and microvariant sequence characterization of 27 Y-STR loci

analyzed in four Eastern African countries, Forensic Sci. Int. Genet. 27 (2017) 123–131. doi:10.1016/j.fsigen.2016.12.015.

[19] Y.M. Khubrani, J.H. Wetton, M.A. Jobling, Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs, Forensic Sci. Int. Genet. 33 (2018) 98–105. doi:10.1016/j.fsigen.2017.11.015.

[20] B. Arredi, E.S. Poloni, S. Paracchini, T. Zerjal, D.M. Fathallah, M. Makrelouf, V.L. Pascali, A. Novelletto, C. Tyler-Smith, A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa, Am. J. Hum. Genet. 75 (2004) 338–345. doi:10.1086/423147.

[21] O. Semino, C. Magri, G. Benuzzi, A.A. Lin, N. Al-Zahery, V. Battaglia, L. Maccioni, C. Triantaphyllidis, P. Shen, P.J. Oefner, L.A. Zhivotovsky, R. King, A. Torroni, L.L. Cavalli-Sforza, P.A. Underhill, A.S. Santachiara-Benerecetti, 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 (2004) 1023–1034. doi:10.1086/386295.

[22] N. Solé-Morata, C. García-Fernández, V. Urasin, A. Bekada, K. Fadhlaoui-Zid, P. Zalloua, D. Comas, F. Calafell, Whole Y-chromosome sequences reveal an extremely recent origin of the most common North African paternal lineage E-M183 (M81), Sci. Rep. 7 (2017) 15941. doi:10.1038/s41598-017-16271-y.

[23] F. Cruciani, P. Santolamazza, P. Shen, V. Macaulay, P. Moral, A. Olckers, D. Modiano, S. Holmes, G. Destro-Bisol, V. Coia, D.C. Wallace, P.J. Oefner, A. Torroni, L.L. Cavalli-Sforza, R. Scozzari, P.A. Underhill, A Back Migration from Asia to Sub-Saharan Africa Is Supported by High-Resolution Analysis of Human Y-Chromosome Haplotypes, Am. J. Hum. Genet. 70 (2002) 1197–1214. doi:10.1086/340257.

[24] F. Cruciani, R. La Fratta, P. Santolamazza, D. Sellitto, R. Pascone, P. Moral, E. Watson, V.

Guida, E.B. Colomb, B. Zaharova, J. Lavinha, G. Vona, R. Aman, F. Cal?, N. Akar, M. Richards, A. Torroni, A. Novelletto, R. Scozzari, Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out Of Africa, Am. J. Hum. Genet. 74 (2004) 1014–1022. doi:10.1086/386294.

[25] F. Cruciani, R. La Fratta, B. Trombetta, P. Santolamazza, D. Sellitto, E.B. Colomb, J.-M. Dugoujon, F. Crivellaro, T. Benincasa, R. Pascone, P. Moral, E. Watson, B. Melegh, G. Barbujani, S. Fuselli, G. Vona, B. Zagradisnik, G. Assum, R. Brdicka, A.I. Kozlov, G.D. Efremov, A. Coppa, A. Novelletto, R. Scozzari, 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. 24 (2007) 1300–1311. doi:10.1093/molbev/msm049.

[26] F. Cruciani, B. Trombetta, D. Sellitto, A. Massaia, G. Destro-Bisol, E. Watson, E. Beraud Colomb, J.-M. Dugoujon, P. Moral, R. Scozzari, Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages, Eur. J. Hum. Genet. 18 (2010) 800–807. doi:10.1038/ejhg.2009.231.

[27] F. Cruciani, B. Trombetta, A. Massaia, G. Destro-Bisol, D. Sellitto, R. Scozzari, A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa, Am. J. Hum. Genet. 88 (2011) 814–818. doi:10.1016/j.ajhg.2011.05.002.

[28] B. Trombetta, E. D'Atanasio, A. Massaia, M. Ippoliti, A. Coppa, F. Candilio, V. Coia, G. Russo, J.-M. Dugoujon, P. Moral, N. Akar, D. Sellitto, G. Valesini, A. Novelletto, R. Scozzari, F. Cruciani, Phylogeographic Refinement and Large Scale Genotyping of Human Y Chromosome Haplogroup E Provide New Insights into the Dispersal of Early Pastoralists in the African Continent, Genome Biol. Evol. 7 (2015) 1940–1950. doi:10.1093/gbe/evv118.

[29] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New

binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res. 18 (2008) 830–838. doi:10.1101/gr.7172008.

[30] B. Trombetta, F. Cruciani, D. Sellitto, R. Scozzari, A New Topology of the Human Y Chromosome Haplogroup E1b1 (E-P2) Revealed through the Use of Newly Characterized Binary Polymorphisms, PLoS One. 6 (2011) e16073. doi:10.1371/journal.pone.0016073.

[31] R. Scozzari, A. Massaia, E. D'Atanasio, N.M. Myres, U.A. Perego, B. Trombetta, F. Cruciani, Molecular Dissection of the Basal Clades in the Human Y Chromosome Phylogenetic Tree, PLoS One. 7 (2012) e49170. doi:10.1371/journal.pone.0049170.

[32] G.D. Poznik, B.M. Henn, M.C.M.-C. Yee, E. Sliwerska, G.M. Euskirchen, A.A. Lin, M. Snyder, L. Quintana-Murci, J.M. Kidd, P.A. Underhill, C.D. Bustamante, Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females, Science. 341 (2013) 562–565. doi:10.1126/science.1237619.

[33] P.A. Underhill, P. Shen, A.A. Lin, L. Jin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonné-Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J.R. Kidd, S.Q. Mehdi, M.T. Seielstad, R.S. Wells, A. Piazza, R.W. Davis, M.W. Feldman, L.L. Cavalli-Sforza, P.J. Oefner, others, Y chromosome sequence variation and the history of human populations, Nat Genet. 26 (2000) 358–361. doi:10.1038/81685.

[34] M. Nei, F. Tajima, DNA polymorphism detectable by restriction endonucleases, Genetics. 97 (1981).

[35] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567. doi:10.1111/j.1755-0998.2010.02847.x.

[36] H.J. Bandelt, P. Forster, A. Rohl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 16 (1999) 37–48. doi:10.1093/oxfordjournals.molbev.a026036.

[37] H.J. Bandelt, Y.-G. Yao, C.M. Bravi, A. Salas, T. Kivisild, Median network analysis of defectively sequenced entire mitochondrial genomes from early and contemporary disease studies, J. Hum. Genet. 54 (2009) 174–181. doi:10.1038/jhg.2009.9.

[38] N.M. Myres, J.E. Ekins, A.A. Lin, L.L. Cavalli-Sforza, S.R. Woodward, P.A. Underhill, Y-chromosome short tandem repeat DYS458.2 non-consensus alleles occur independently in both binary haplogroups J1-M267 and R1b3-M405., Croat. Med. J. 48 (2007) 450–459.

[39] S. Tofanelli, G. Ferri, K. Bulayeva, L. Caciagli, V. Onofri, L. Taglioli, O. Bulayev, I. Boschi, M. Alù, A. Berti, C. Rapone, G. Beduschi, D. Luiselli, A.M. Cadenas, K.D. Awadelkarim, R. Mariani-Costantini, N.E. Elwali, F. Verginelli, E. Pilli, R.J. Herrera, L. Gusmão, G. Paoli, C. Capelli, J1-M267 Y lineage marks climate-driven pre-historical human displacements., Eur. J. Hum. Genet. 17 (2009) 1520–1524. doi:10.1038/ejhg.2009.58.

[40] N.M. Myres, K.H. Ritchie, A. a Lin, R.H. Hughes, S.R. Woodward, P. a Underhill, Y-chromosome Short Tandem Repeat Intermediate Variant Alleles DYS392.2, DYS449.2, and DYS385.2 Delineate New Phylogenetic Substructure in Human Y-chromosome Haplogroup Tree, Croat. Med. J. 50 (2009) 239–249. doi:10.3325/cmj.2009.50.239.

[41] C. Rapone, E. D'Atanasio, A. Agostino, M. Mariano, M.T. Papaluca, F. Cruciani, A. Berti, Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler® Plus kit) in an Italian population sample, Forensic Sci. Int. Genet. 21 (2016) e1–e5. doi:10.1016/j.fsigen.2015.11.006.

[42] C.L. Campbell, P.F. Palamara, M. Dubrovsky, L.R. Botigue, M. Fellous, G. Atzmon, C. Oddoux, A. Pearlman, L. Hao, B.M. Henn, E. Burns, C.D. Bustamante, D. Comas, E. Friedman, I. Pe'er, H. Ostrer, North African Jewish and non-Jewish populations form distinctive, orthogonal clusters, Proc. Natl. Acad. Sci. 109 (2012) 13865–13870. doi:10.1073/pnas.1204840109.

[43] K. Fadhlaoui-Zid, M. Haber, B. Martínez-Cruz, P. Zalloua, A. Benammar Elgaaied, D. Comas, Genome-wide and paternal diversity reveal a recent origin of human populations in North

Africa., PLoS One. 8 (2013) e80293. doi:10.1371/journal.pone.0080293.

[44] L. Pagani, S. Schiffels, D. Gurdasani, P. Danecek, A. Scally, Y. Chen, Y. Xue, M. Haber, R. Ekong, T. Oljira, E. Mekonnen, D. Luiselli, N. Bradman, E. Bekele, P. Zalloua, R. Durbin, T. Kivisild, C. Tyler-Smith, Tracing the Route of Modern Humans out of Africa by Using 225 Human Genome Sequences from Ethiopians and Egyptians, Am. J. Hum. Genet. 96 (2015) 986–991. doi:10.1016/j.ajhg.2015.04.019.

[45] C. Coudray, A. Olivieri, A. Achilli, M. Pala, M. Melhaoui, M. Cherkaoui, F. El-Chennawi, M. Kossmann, A. Torroni, J.M. Dugoujon, The Complex and Diversified Mitochondrial Gene Pool of Berber Populations, Ann. Hum. Genet. 73 (2009) 196–214. doi:10.1111/j.1469-1809.2008.00493.x.

[46] C. Coudray, E. Guitard, C. Keyser-Tracqui, M. Melhaoui, M. Cherkaoui, G. Larrouy, J.M. Dugoujon, Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco, Forensic Sci. Int. 167 (2007) 81–86. doi:10.1016/j.forsciint.2006.01.003.

[47] L. Palet, C. Coudray, C. Galey, C. Keyser, M. Melhaoui, C. Gagnor, M. Sabatier, J.M. Dugoujon, Y-STR genetic diversity in Moroccans from the Figuig oasis, Forensic Sci. Int. Genet. 4 (2010) 139–141. doi:10.1016/j.fsigen.2010.02.006.

[48] L.A. Zhivotovsky, P.A. Underhill, M.W. Feldman, Difference between Evolutionarily Effective and Germ line Mutation Rate Due to Stochastically Varying Haplogroup Size, Mol. Biol. Evol. 23 (2006) 2268–2270. doi:10.1093/molbev/msl105.

[49] P. Hallast, C. Batini, D. Zadik, P. Maisano Delser, J.H. Wetton, E. Arroyo-Pardo, G.L. Cavalleri, P. de Knijff, G. Destro Bisol, B.M. Dupuy, H.A. Eriksen, L.B. Jorde, T.E. King, M.H. Larmuseau, A. Lopez de Munain, A.M. Lopez-Parra, A. Loutradis, J. Milasin, A. Novelletto, H. Pamjav, A. Sajantila, W. Schempp, M. Sears, A. Tolun, C. Tyler-Smith, A. Van Geystelen, S. Watkins, B. Winney, M.A. Jobling, The Y-Chromosome Tree Bursts into Leaf: 13,000 High-Confidence

SNPs Covering the Majority of Known Clades, Mol. Biol. Evol. 32 (2015) 661–673. doi:10.1093/molbev/msu327.

[50] O. Balanovsky, Toward a consensus on SNP and STR mutation rates on the human Y-chromosome, Hum. Genet. 136 (2017) 575–590. doi:10.1007/s00439-017-1805-8.

[51] Q. Fu, A. Mittnik, P.L.F. Johnson, K. Bos, M. Lari, R. Bollongino, C. Sun, L. Giemsch, R. Schmitz, J. Burger, et al., A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes, Curr Biol. 23 (2013) 553–559. doi:10.1016/j.cub.2013.02.044.

[52] P. Hallast, P. Balaresque, G.R. Bowden, S. Ballereau, M.A. Jobling, Recombination Dynamics of a Human Y-Chromosomal Palindrome: Rapid GC-Biased Gene Conversion, Multi-kilobase Conversion Tracts, and Rare Inversions, PLoS Genet. 9 (2013) e1003666. doi:10.1371/journal.pgen.1003666.

[53] A. Bekada, L.R. Arauna, T. Deba, F. Calafell, S. Benhamamouch, D. Comas, Genetic Heterogeneity in Algerian Human Populations, PLoS One. 10 (2015) e0138453. doi:10.1371/journal.pone.0138453.

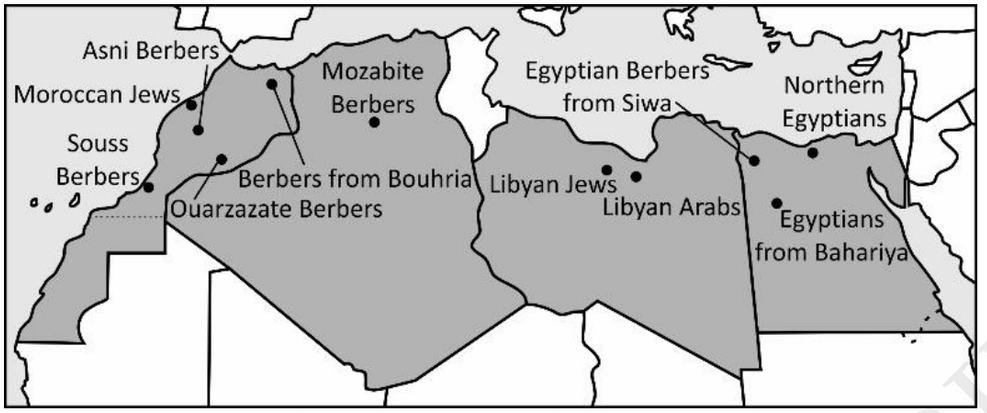


Figure 1: Geographic localization of the eleven northern African populations analysed.

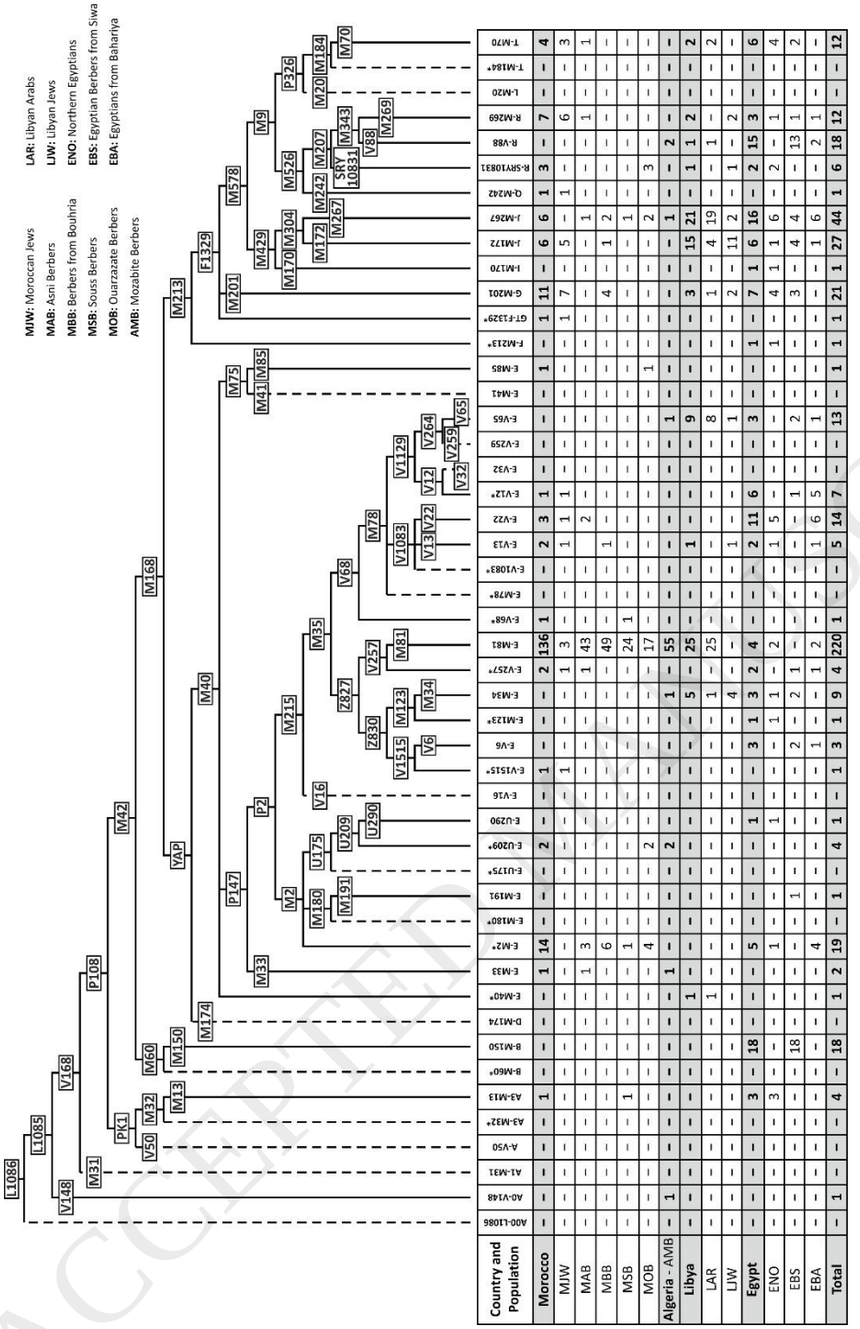


Figure 2: Low-resolution Y phylogenetic tree and geographic distribution of the Y-SNPs analysed in this study. Haplogroups/paragroups which are absent in our sample are represented by dashed lines.

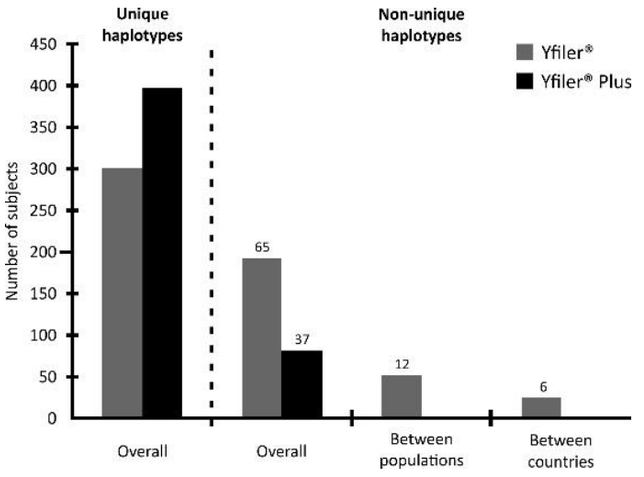


Figure 3: Comparison of the haplotype sharing proportion between Yfiler® and Yfiler® Plus multiplex. On the ordinate axis, the number of subjects carrying a non-unique haplotype is represented. Above each bar, we reported the number of haplotypes shared between two or more subjects.

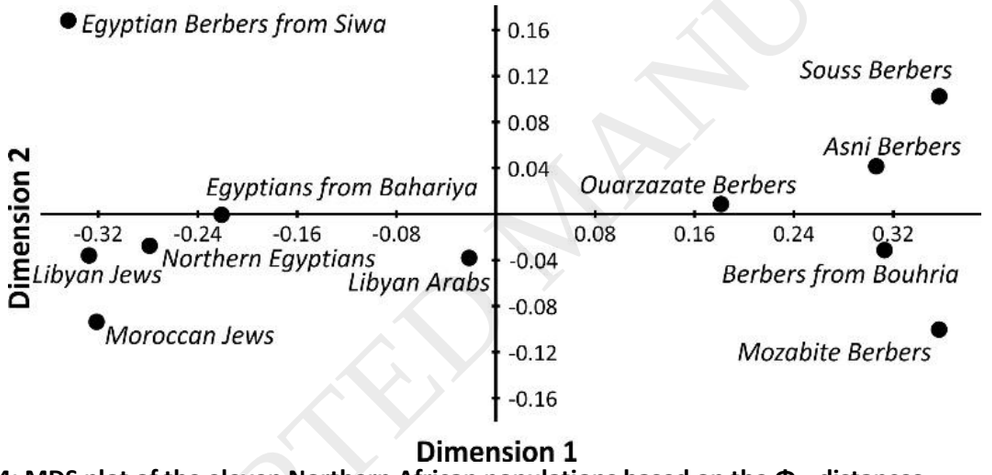


Figure 4: MDS plot of the eleven Northern African populations based on the Φ_{ST} distances.

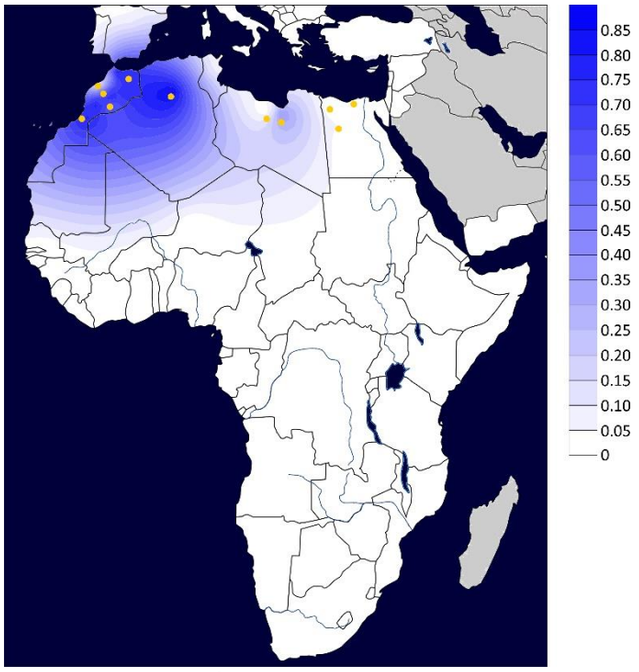


Figure 5. Frequency map of the E-M81 haplogroup. The geographic localization of the 11 populations is represented by the yellow dots.

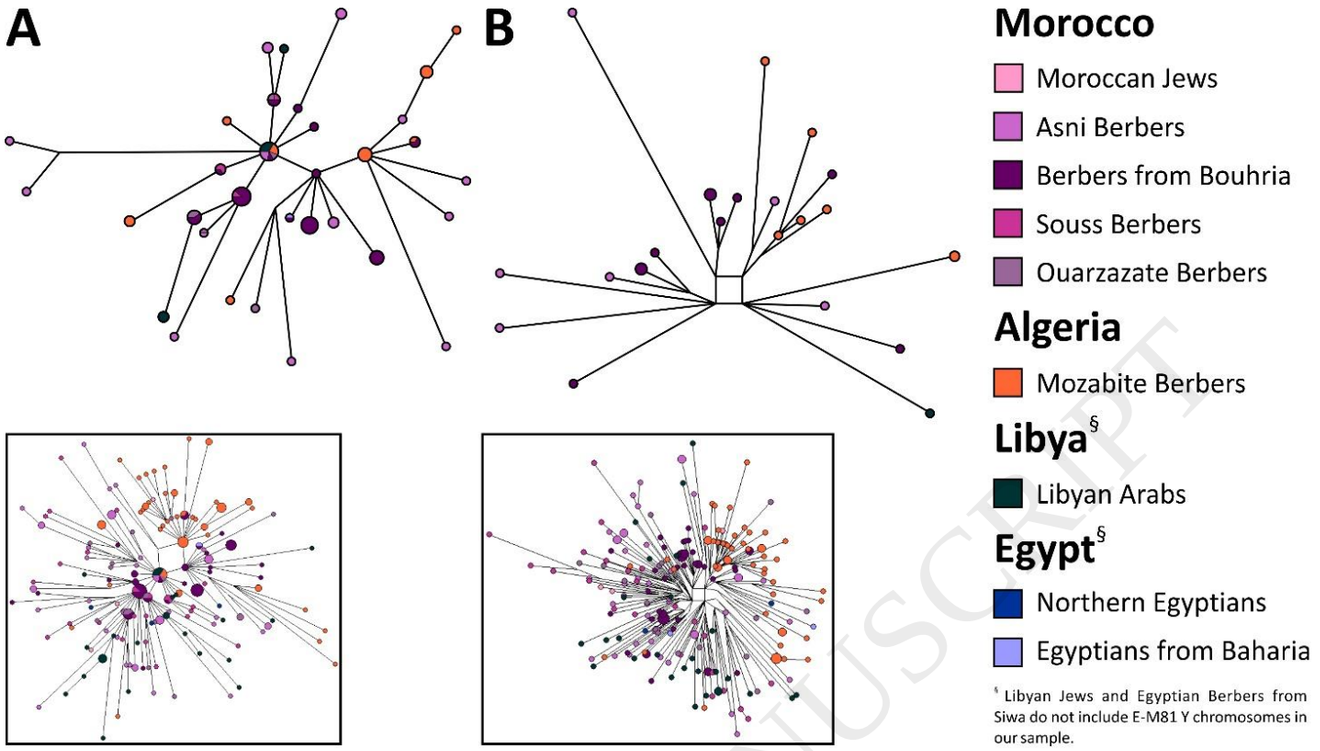


Figure 6: Median-joining networks of the E-M81 subjects. a) Network reconstructed using the Yfiler® loci, representing only the shared haplotypes. b) Network reconstructed using the Yfiler® Plus loci, representing only the shared haplotypes. Insets: complete versions of the E-M81 networks. Each circle is drawn proportionally to the number of subjects that share the same haplotype and is coloured according to the legend.

Country	Ethnic group	Sample size (*n*)	Number of haplotypes	HD	HMP	DC	Mean number of pairwise differences	Average gene diversity over loci
Morocco		***206***	***182***	***0.9985***	***0.0064***	***0.8835***	***14.1560***	***0.5662***
	Moroccan Jews	31	31	1.0000	0.0323	1.0000	18.1419	0.7257
	Asni Berbers	53	47	0.9956	0.0231	0.8868	12.2925	0.4917
	Berbers from Bouhria	66	49	0.9883	0.0266	0.7424	11.3161	0.4526
	Souss Berbers	28	28	1.0000	0.0357	1.0000	11.3254	0.4530
	Ouarzazate Berbers	28	27	0.9974	0.0383	0.9643	14.7593	0.5904
Algeria		***64***	***57***	***0.9960***	***0.0195***	***0.8906***	***10.4846***	***0.4194***
	Mozabite Berbers	64	57	0.9960	0.0195	0.8906	10.4846	0.4194
Libya		***87***	***85***	***0.9995***	***0.0120***	***0.9770***	***16.3101***	***0.6524***
	Libyan Arabs	63	62	0.9995	0.0164	0.9841	14.9805	0.5992
	Libyan Jews	24	23	0.9964	0.0451	0.9583	17.0145	0.6806
Egypt		***120***	***109***	***0.9985***	***0.0099***	***0.9083***	***17.7060***	***0.7082***
	Northern Egyptians	35	32	0.9950	0.0335	0.9143	18.0252	0.7210
	Egyptian Berbers from Siwa	54	48	0.9958	0.0226	0.8889	16.4668	0.6587
	Egyptians from Bahariya	31	29	0.9957	0.0364	0.9355	17.1441	0.6858
Overall		***477***	***433***	***0.9995***	***0.0026***	***0.9078***	***16.1034***	***0.6441***

Table 1: Forensic parameters for 11 northern African populations analysed with the Yfiler® Plus multiplex.

	Number of populations	Number of groups	Φ-statistics			F-statistics		
			Φ_{ST} (*p*)	Φ_{CT} (*p*)	Φ_{SC} (*p*)	F_{ST} (*p*)	F_{CT} (*p*)	F_{SC} (*p*)
All northern Africa	11	1	0.1968 *(0.0000)*			0.1032 *(0.0000)*		
Country level analysis								
Morocco	5	1	0.2116 *(0.0000)*			0.1003 *(0.0000)*		
Algeria	1	1	-			-		
Libya	2	1	0.1101 *(0.0010)*			0.1138 *(0.0000)*		
Egypt	3	1	0.0610 *(0.0000)*			0.0130 *(0.2111)*		
Overall[a]	11	4	0.2689 *(0.0000)*	0.1360 *(0.0489)*	0.1539 *(0.0000)*	0.1482 *(0.0000)*	0.0648 *(0.0271)*	0.0892 *(0.0000)*

Table 2: Analysis of molecular variance for the Y-STR haplotypes in northern Africa.

[a] Populations grouped into countries.

Haplogroup	Time estimates (22 Y-STRs) *(present study)*	Time estimates (17 Y-STRs) *(present study)*	Time estimates (Y-SNPs)[a,b] *(ref 22)*	Time estimates (13 Y-STRs)[a] *(ref 22)*
E-M81	1.66 (± 0.32)	1.93 (± 0.29)	2.27 (± 0.33)	2.01 (± 0.43)
E-V65	1.61 (± 0.31)	1.47 (± 0.22)	not present	not present

Table 3: TMRCA estimates, expressed in kya, of the E-M81 and the E-V65 subjects using two different sets of Y-STRs and comparison with the time estimates from ref. [22].

a. Time estimates of the E-M183 sub-haplogroup, obtained with the Rho method.

b. Time estimates calculated with the mutation rate of 0.617×10^{-9} substitutions/site/year [51].