Annals of Human Biology

ISSN: 0301-4460 (Print) 1464-5033 (Online) Journal homepage: https://www.tandfonline.com/loi/iahb20

Paternal lineage of the Berbers from Aurès in Algeria: Estimate of their genetic variation

Amine Abdeli & Traki Benhassine

To cite this article: Amine Abdeli & Traki Benhassine (2019): Paternal lineage of the Berbers from Aurès in Algeria: Estimate of their genetic variation, Annals of Human Biology, DOI: 10.1080/03014460.2019.1602166

To link to this article: https://doi.org/10.1080/03014460.2019.1602166

Accepted author version posted online: 02 Apr 2019.

Paternal lineage of the Berbers from Aurès in Algeria: Estimate of their genetic variation

Amine Abdeli[a,b] and Traki Benhassine[a*]

[a] *Laboratoire de Biologie Cellulaire et Moléculaire, Faculté des Sciences Biologiques, Université des Sciences et de la Technologie Houari Boumediene, Algiers, Algeria*

[b] *Institut National de Criminalistique et de Criminologie de la Gendarmerie Nationale, Algiers, Algeria*

* Corresponding author :

Traki Benhassine

Laboratoire de Biologie Cellulaire et Moléculaire, Faculté des Sciences Biologiques, USTHB, BP 32, El Alia, Bab Ezzouar, 16111 Alger, Algérie

e-mail: trakibenhassine@hotmail.com

Paternal lineage of the Berbers Aurès in Algeria

ABSTRACT

Background: Aurès is a vast territory in the east of Algeria, characterised by its traditional Berber settlement which has preserved its language and its rich history; its name goes back to antiquity and before the Roman conquest it was part of the territory of ancient Numidia. The Chaoui people in this region are one of Algeria's largest Berber groups.

Aim: Our aim was to investigate the level of genetic diversity of the Berbers of Aurès through the analysis of the paternal gene pool and to estimate the percentage of genetic variation among different geographical regions and linguistic groups from Algeria.

Subjects and methods: 23 Y-STR were genotyped in a sample of 218 unrelated males of the Berbers of Aurès. Algorithms were used to estimate the Y-chromosome haplogroups. Genetic distance, non-metric MDS and AMOVA were used to analyse the genetic relationships between sample groups.

Results: The paternal lineage of our sample of the Aurès region did not exhibit strong signals of differentiation with other samples from North-central, Northwest, and South Algeria. However, significant differences were found within our sample, demonstrating a high degree of heterogeneity.

Conclusion: Our results demonstrate that Aurès people are isolated and closed, but nevertheless have quite different genetic profiles.

Keywords: Y-chromosome, population genetics, Aurès region, Algeria, haplotypes.

Introduction

Algeria is a North African country on the Mediterranean coast, with an area of 2 381 741 km². It shares land borders to the north-east with Tunisia, to the east with Libya, to the south with Niger and Mali, to the south-west with Mauritania and Western Sahara, and to the west with Morocco. The population of Algeria reached 42.2 million inhabitants in 1 January 2018 (ONS 2018).

North Africa was a strategic crossroad and a region of choice for different civilisations, from the prehistory period with Aterian, Iberomaurusian and Capsian cultures, followed by the history period with Phoenicians, Romans and Vandals, Byzantines, Muslim Arabs, Spaniards, and Ottomans, to the European countries colonial period during the 18th and 19th centuries. After the Muslim Arab conquest in the 7th century, the ancient inhabitants of North Africa, the Berber speaking people which are likely autochthonous to the region (Camps 1974), were converted to Islam and Arabic (Dugoujon et al. 2009).

This historical past, marked by many invasions, conquests and migrations in North Africa, has left an important imprinting on the current genetic background and has also had a direct impact on the geographical distribution of the current Berber groups that are now spread throughout a vast territory ranging from Mauritania to Egypt (the oasis of Siwah) and the Saharan desert to the Atlas Mountains (Camps 1980).

Today in Algeria, the Berber language (Tamazight) is composed of several different dialects of which the most important are: Kabyle, Chaoui, Mozabite and Touareg, the most indisputable criterion for identifying these people being their language. Berbers represent 15 to 33% of the population in Algeria, as well as in Libya and Tunisia. The Chaoui people (also spelled Shawiya: Chaouïa) who reside in the

Aurès Mountains of Algeria, are one of Algeria's largest Berber groups and are often described as one of the three major groups of Berbers (Danver 2015).

The development of molecular biology and the study of the human genome with the advent of genome-wide genotyping and sequencing techniques offer important tools to analyse the level of human diversity. During the last decades, several molecular markers of nuclear and mitochondrial DNA have been used in worldwide human population genetic studies in order to investigate the genetic history of populations, study migration of modern human, trace the "most recent common ancestor", estimate human geographic origins (Kundu and Ghosh 2015) and establish autosomal Short Tandem Repeat (autosomal STR) allele frequencies and Y chromosome haplotype reference databases that can be used by forensic DNA laboratories in the interpretation of their results.

Many studies on the genetic diversity of Berber and Arab speaking groups were thus conducted in the Algerian population using different informative markers in human population genetics, based on classical protein variants (Bosch et al. 1997; Lefevre-Witier et al. 2006), autosomal microsatellites (STR) (Bosch et al. 2000, Amir et al. 2015), autosomal SNP (Henn et al. 2012; Bekada et al. 2015), Alu sequences (Comas et al. 2000; González-Pérez et al. 2010), others on Y chromosomes (Bosch et al. 2000; Arredi et al. 2004; Robino et al. 2008; Vermeulen et al. 2009; Bekada et al. 2013, 2015; Solé-Morata et al. 2017), mitochondrial DNA (mtDNA) (Côrte-Real et al. 1996; Ivanova et al. 1999; Macaulay et al. 1999; Plaza et al. 2003; Pereira et al. 2010; Bekada et al. 2013, 2015) and X chromosomes (Bekada et al. 2010).

While all these studies were conducted on Arab and Berber speaking groups located in the northwest, north-central region, and south of Algeria, the lack of DNA

data from east and northeast Algeria is apparent, therefore we have analysed a sample from a vast territory comprising a mountain range and plains in the east of Algeria (Aurès region), which is characterised by both its traditional Berber settlement which has preserved its language (the Berber-speaking group of Chaouis) and its rich history; the name of the Aurès region goes back to antiquity (Perrot d'Ablancourt 1667).

The population of Aurès is composed mostly of different Berber tribes. According to Slane (1856), translator of the books of Ibn Khaldun, the various tribes of Aurès are mainly Zenata, which were one of the largest Berber tribal confederations (de Slane 1856).

The aim of this work was to investigate the genetic diversity of Chaoui peoples, through analysis of the paternal lineage of their gene pool, in order to evaluate genetic variation among different geographical regions and linguistic groups from Algerian, Berber and Arab Algerian speaking groups, and among 28 other populations corresponding to twenty five (25) countries in North Africa, Europe, Sub-Saharan Africa, and the Middle East. The aim was also to predict Y chromosome haplogroups to estimate the geographic origin of the Berbers of Aurès, through an alternative method in the absence of Y-SNP genotypes, using software for Y-haplogroup prediction based on Y-STR haplotypes.

Another aim of this study was to establish Y chromosome haplotype reference databases, based on the loci included in the PowerPlex®Y23 System (PPY23, Promega Corporation, Madison, WI), in order that they can be used by forensic DNA laboratories in the interpretation of their results (e.g. statistical calculations in criminal caseworks and kinship testing).

Subjects and methods

Blood samples were collected from 218 unrelated male donors of the Berbers of Aurès, originating from three different localities namely Batna, Khenchela and Oum El Bouaghi. At the time of sample collection, a questionnaire was completed and written informed consent was acquired from each donor. Our study complies with the ethical rules of all the institutions involved, and has been approved by the Laboratoire de Biologie Cellulaire et Moléculaire at the Faculté des Sciences Biologiques of USTHB, Algiers, and the Institut National de Criminalistique et de Criminologie de la Gendarmerie Nationale of Algiers. All the samples correspond to autochthonous people with their parents and all grandparents originating in this region. In addition, the data of Y-STRs of other Berber and Arab speaking groups in Algeria that have previously been published were used in the present study for comparative purposes, namely Oran (*n* = 102) (Robino et al. 2008), Mozabites from Ghardaïa (*n* = 20) (Vermeulen et al. 2009), Algiers (*n* = 26), Oran (*n* = 78), Reguibate from Tindouf (*n* = 54), and Zenata from Gourara (Timimoun) (*n* = 34) (Bekada et al. 2015*)*. The samples are presented with their respective geographic locations in Figure 1.

DNA was extracted using the QIAamp DNA blood mini kit (Qiagen GmbH, Hilden, Germany) following manufacturer's recommendations and was quantified by Quantifiler™ Human DNA Kit on 7500 SDS Real-Time PCR System (Applied Biosystems). PCR amplification was performed using PowerPlex® Y23 System (PPY23, Promega Corporation, Madison, WI), containing 23 STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, GATAH4, DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643). Amplified products were

separated and detected on the 3130xl Genetic Analyzer (Life Technologies). Allele designations were automatically assigned by GeneMapper® ID v3.2 Software (Life Technologies). All samples were analysed at the Department of Biology, National Institute of Criminalistics and Criminology of the National Gendarmerie (Algeria). This Department was accredited according to the ISO/ IEC 17025:2005 by ANAB and has participated in international proficiency tests (CTS: Collaborative Testing Services Inc., USA).

To predict haplogroups from Y-STR markers, the haplotypes of 21 Y-STR markers of the 218 individuals (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635, GATAH4, DYS481, DYS533, DYS570 and DYS576) were analysed using two Y-chromosome haplogroup predictors. The first predictor was Whit Athey's haplogroup predictor (Athey 2005, 2006). A minimum "fitness score" of 40 and "Bayesian probability" of 90% were selected, and "Mediterranean" was selected under the "area selection" option considering the geographic location of the Aurès region. The second predictor was Y Predictor by Vadim Urasin 1.5.0 (Free program distributed by Vadim Urasin: Personal Communication).

Statistical analysis

Forensic parameters

Forensic parameters were calculated for all samples (n=218) and for all 23 markers of the PPY23 kit. Allele frequencies, haplotype and haplogroup frequencies were estimated by the counting method and confirmed with Arlequin software v.3.5.2.2 (Excoffier and Lischer 2010). Genetic diversity (GD) and Haplotype diversity (HD) were calculated as GD = $n(1-\Sigma P_i^2)/(n-1)$, according to Nei and Tajima (1981), where n

is the sample size and *Pi* is the frequency of the i^{th} haplotype. Match probability (MP) was calculated as the sum of squared haplotype frequencies. The discrimination capacity (DC), defined as the probability that two randomly chosen haplotypes are different in the population, was determined by dividing the number of observed haplotypes by the number of sampled individuals.

Population differences

The population genetic structure of Algerian samples, the sample from the present study, and other published samples previously reported, were investigated by means of analysis of molecular variance (AMOVA) (Excoffier et al. 1992) based on 14 Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS635 and GATAH4). AMOVA was performed using Arlequin software ver 3.5.2.2 (Excoffier and Lischer 2010) to estimate the genetic variation among groups (*Fct*), among populations within groups (*Fsc*), and among all populations non-grouped (*Fst*) by assigning Algerian samples to different groups according to their geographic regions and their spoken language (Arabic dialects versus Berber dialects). The statistical significance of *F* values was estimated by permutation analysis using 10 000 random permutations.

All samples carrying a null allele, a duplication at one or more markers, or an intermediate allele (an incomplete repeat unit) were excluded from the AMOVA. The DYS385ab marker was not included in the AMOVA.

Genetic relationships were assessed by pairwise genetic distances (*Rst*) (Reynolds et al. 1983; Slatkin 1995) using Arlequin ver 3.5.2.2 (Excoffier and Lischer 2010). Two comparative analyses were carried out, the first calculated between our sample and six (6) other Algerian samples (Arab and Berber groups) available in the

literature (Robino et al. 2008; Vermeulen et al. 2009; Bekada et al. 2015), the second between our sample and 28 other populations from neighbouring regions or countries from North Africa (El-Sibai et al. 2009; Aboukhalid et al. 2010; Laouina et al. 2011; Ottoni et al. 2011; Elmrghni et al. 2012; Fadhlaoui-Zid et al. 2012; Triki-Fendri et al. 2013), Sub-Saharan Africa (Purps et al. 2014; Larmuseau et al. 2015; Iacovacci et al. 2017), Middle East (El-Sibai et al. 2009; Purps et al. 2014; Taqi et al. 2015; Jones et al. 2017; Tokdemir and Tunçez 2017) and Europe (Purps et al. 2014; Ramos-Luis et al. 2014; Turrina et al. 2015; Martinez-Cadenas et al. 2016). The analyses were based on the same 14 Y-STR loci cited above for 532 males, including the 218 analysed in this study, from different regions of Algeria and for 8558 males from twenty-six (26) countries. Data on origins, sample sizes and references from the populations used for the comparative analyses are summarised in Supplementary Table 1. The pairwise population comparisons were tested at a significance level of 0.05 with 10 000 permutations and *p*-values were revised with the sequential Bonferroni correction for multiple tests of significance (Rice 1989).

All the population genetic analyses were performed using the Arlequin software package (Excoffier and Lischer 2010). To visualise differences in Y-STR genetic variation between populations based on the pairwise linearised values (*Rst*), two Multidimensional Scaling (MDS) were performed using PAST version 3.16 software (Hammer et al. 2001).

Results

The Y-chromosome haplotypes and haplogroups of 218 Chaoui individuals are shown in Supplementary Table 2. A total of 178 different haplotypes (among 218 samples) were found, from which 154 (86.5%) were unique and 24 (13.5%) were found more than once: eighteen haplotypes were detected twice, three haplotypes appeared three times and one haplotype appeared four, six and nine times (Table 1).

In our Aurès sample, genetic diversity values of all 23 Y-STR loci ranged from a minimum of 0.0632 for DYS392 to a maximum of 0.8363 for DYS385a/b (Supplementary Table 3). It exceeded 0.5 for 16 markers (69.6%), 0.6 for 9 (39.1%) and even 0.7 for 5 (21.7%) markers. The least polymorphic loci were DYS437 with 2 alleles and DYS389I, DYS391, DYS393, DYS438, DYS456 and DYS533 with 4 alleles.

We could also find duplications in two individuals, one showed two alleles, 22 and 23, for locus DYS570 and two alleles, 17 and 18, for locus DYS576, while the other showed alleles 11 and 12 for locus DYS439. The haplotype diversity (HD), match probability (MP) and discrimination capacity (DC) were 0.9964, 0.0081 and 0.8165, respectively.

Regarding the Y-chromosome haplogroups provided by the Y-STR haplotype based predictors, the results obtained from the two algorithms were almost similar (85.3% and 84.4%, according to Whit Athey's and Vadim Urasin's algorithms, respectively). The main paternal lineage in our Aurès sample was represented by haplogroup E1b1b, and the second was haplogroup J (5.5%), (J1: 4.6% and J2: 0.9%) followed by haplogroups E1b1a (4.6%) and R1a (2.3%). The frequencies of these haplogroups were similar according to the two haplogroup predictors. Haplogroups R1b, L and Q, assigned by Whit Athey's haplogroup predictors, and I2, L1b*, F, K and

T2, assigned by Vadim Urasin's haplogroup predictors, were present with a frequency less than 1.4% (Table 2).

Pairwise genetic distances *Rst* and associated *p*-values calculated between the paternal gene pool of the Aurès sample (northeast Algeria sample's) and six (6) other Algerian samples (Arab and Berber groups) are shown in Supplementary Table 4. Population pairwise comparisons showed significant genetic differences between the males of the Aurès and all the other male samples from Algeria. The largest coefficient of genetic differentiation (*Rst*) was noted with the Zenata samples from Gourara/Timimoun (*Rst* = 0.097) followed by the sample from Oran (Oran 2) (*Rst* = 0.095), whereas the smallest coefficient was that of the sample from Algiers (*Rst* = 0.053). When using the sequential Bonferroni correction for type I errors, the differences with the samples from Ghardaïa (Mozabite) and Algiers ($p = 0.04$, $p = 0.01$, respectively) became statistically insignificant ($p = 0.28$, $p = 0.12$). Also, no significant differences were observed between the samples of Zenata from Timimoun and the Mozabites from Ghardaïa.

To illustrate these genetic differences, we performed a multi-dimensional scaling (MDS), using pairwise *R*st values. Results are shown in Figure 2. The MDS based on data from fourteen (14) Y-STR loci displayed a clear separation between our sample and the other samples from Algeria, and to a lesser extent with the Mozabite sample from Ghardaïa. There was not a clear distinction between the Northwest (Oran 1 and Oran 2) and North-central (Algiers) samples, as no statistically significant differences could be observed. The Reguibate sample from Tindouf occupied a peripheral position in the plot, far from the others with the exception of the sample from Algiers.

In a second step, we initiated a comparative study based on pairwise genetic distances in order to position the Chaoui sample within 28 other populations from neighbouring regions or countries from North Africa, Sub-Saharan Africa, the Middle East and Europe. Pairwise genetic distances *Rst* and associated *p*-values are shown in Supplementary Table 5.

The smallest *Rst* values were observed for male populations from North Africa, namely Morroco2 ((*Rst* = 0.074), individuals resident in Casablanca), followed by Tunisia (*Rst* = 0.102) and Egypt (*Rst* = 0.105). The largest *Rst* value was observed for the sample of Nigeria (*Rst* = 0.692), which was also found to be distant from all the other populations. For Sub-Saharan Africa, the smallest *Rst* value (*Rst* = 0.107, *p* = 0.0000) was noticed for Ethiopia. The corresponding *p*-values showed that the differences between our sample and the 28 populations compared were all significant (*p* = 0.0000), even for the closest countries such as Spain (*Rst* = 0.311, *p* = 0.0000), France (*Rst* = 0.300, *p* = 0.0000) and Italy (*Rst* = 0.183, *p* = 0.0000). These findings were in accordance with the MDS (Figure 3).

MDS revealed that our Aurès sample and the North African populations compared are closer to those from the Middle East, while the Southwestern and East Libyan populations have more affinity with Sub-Saharan populations. Ethiopia (*Rst* = 0.107), Greece (*Rst* = 0.176) and Italy (*Rst* = 0.183) also cluster with the Middle Eastern populations.

In addition, MDS showed the remarkable heterogeneity of Sub-Saharan African populations, the similarity between male populations from the Middle East (Iran, Iraq, Jordan, Kuwait, Lebanon, Turkey, and United Arab Emirates) and little difference between the European populations (France, Germany, Greece, Italy, Portugal and

Spain). However, no statistically significant differences were observed between male populations from Portugal and France ($p = 0.06$).

AMOVA analysis based on Y-STR haplotype frequencies, when all the seven Algerian samples were considered as a single group (Table 3), statistically demonstrates that there is significant genetic differentiation between these populations (*Fst*=0.076, *p*-value=0.00). This result shows the importance of the degree of heterogeneity of the Y chromosome haplotypes in the Algerian human populations.

Table 3 shows another part of the results we obtained, where seven samples organised into four groups (Northeast, Northwest, North-central, and South) have been analysed. The statistical evaluation shows that there are no significant differences between these groups (*Fct*=0.013, *p*-value=0.31). However, it has been found that there are significant differences among these populations within the groups (*Fsc*=0.065, *p*-value=0.00).

In the same context, it is important to mention that the results obtained when performing the AMOVA analyses with two groups according to their geographic regions (Northeast versus Northwest, Northeast versus South), or according to their spoken languages (Arabic dialects versus Berber dialects), show that the variation among populations within groups (*Fsc*) was higher than that observed between groups (*Fct*) with the exception of the comparison of Northeast versus Northwest, but the results reveal no significant genetic differentiation among groups for all comparisons (*Fct* < 0.08, *p-value* > 0.31), suggesting no geographical or linguistic structuring for these populations.

Discussion

The results obtained in this study represent the first data reported on the genetic diversity of Berbers from the Aurès region. The paternal lineage of the sample taken from this Algerian region was characterised by two major haplogroups: E1b1b and J. These lineages are usually associated with North African and Middle Eastern populations, respectively. According to some studies, the first haplogroup is mainly found at high frequencies in the autochthonous North African lineages (Arredi et al. 2004; Cruciani et al. 2004; Fadhlaoui-Zid et al. 2013; Bekada et al. 2015; Solé-Morata et al. 2017). Regarding the second haplogroup, the reviewed literature shows that this has the highest concentration in the Middle East (Zalloua et al. 2008; El-Sibai et al. 2009; Chiaroni et al. 2010; Haber et al. 2011). The presence of both haplogroups in our sample could be attributed to many reasons, for example the Phoenician, Islamic and/or Ottoman Empire expansions.

Moreover, other haplogroups were found in the examined sample, namely the E1b1a (4.6%), which is more frequent in Western and Central Sub-Saharan African countries (Luis et al. 2004). The existence of the Sub-Saharan paternal lineages in the Chaouis of Aurès may be explained as a result of the trade in African slaves across the Sahara to North Africa (the Maghreb), and the Trans-Saharan exchanges between North and West Africa during the pre-Islamic period (Nixon 2013). Additionally, haplogroups which are predominant in European, Eastern European, Western and Central Asian populations were observed in our sample with low frequencies (less than 2.3%).

Regarding the comparative study carried out by pairwise genetic distance based on data from 14 Y-STR, it is worth noting that the results of this section show a similarity between the paternal gene pool of the samples taken from the Aurès and

Ghardaïa regions on one hand, and between Aurès and Algiers regions on the other hand. The similarity with the Ghardaïa sample might be explained by the presence of some tribes of Zenata in this region, as they had lived in the Aurès since a very remote period (de Slane 1856). With respect to the similarity with the sample of Algiers, it is important to mention that this city has undergone exponential demographic growth since the independence of the country, particularly because of migration waves of other regions of the country towards the city.

Our data also show significant differences between the paternal gene pool of the Aurès region sample and the other four Algerian samples that have previously been published (Bekada et al. 2015) taken from Oran (two samples), Gourara and Tindouf. These results could be explained by the fact that in the past the Aurès region was mainly inhabited by different tribes of Zenata, compared to Oran and Gourara (Timimoun). Thus, in addition to the existence of these tribes in Oran and Gourara, these regions are characterised by the presence of a mixture of Algerians from different parts of the country for Oran, and by the presence of a considerable percentage (46.5% in 1952) of South Saharan origin (Bisson 1999) for Gourara. In fact, their presence in Gourara has been highlighted by Bekada et al, where the E1ba1 (M2) haplogroup was found at high frequency (22,86%) in this region (Bekada et al. 2015).

Regarding the significant genetic difference noted with the sample of Tindouf (Reguibate), it has been well noted by many historians that the Reguibate is a tribe of Sanhadja-Berber origin, which does not belong to the Zenata confederation (Barbier 2003).

Estimates of levels of genetic variation among groups according to their geographic regions and linguistic features, based on data from 14 Y-STR, show no

significant differences between them. Several prehistoric and historical factors might be linked to this situation, particularly the propagation of Zenatian tribes over the whole of Africa including Central Maghreb, Maghreb El Acsa and even in the desert of Ifrikia.

In a broader comparative context, and based on haplotype data, *Rst* pairwise analyses and multidimensional scaling show that males of the Aurès region are closer to North African (namely: Morocco 2 and Tunisia) and Middle Eastern male populations than those from Europe and Sub Saharan Africa populations; cultural, geographical, and historical backgrounds may be the fundamental reasons for this.

Acknowledgements

We thank Dr. Tarik Hamadouche (Laboratory of Molecular Biology at the UMBB University of Boumerdes, Laboratory of Neurosciences at the University of Algiers) for his suggestions, comments and critical reading of this manuscript.

References

Aboukhalid R, Bouabdellah M, Abbassi M, Bentayebi K, Elmzibri M, Squalli D, Amzazi S. 2010. Haplotype frequencies for 17 Y-STR loci (AmpFlSTR® Y-filer™) in a Moroccan population sample. Forensic Sci Int Genet. 4(3):e73–e74.

Amir N, Sahnoune M, Chikhi L, Atmani D. 2015. STR-based genetic structure of the Berber population of Bejaia (Northern Algeria) and its relationships to various ethnic groups. Gene. 574(1):140–148.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C. 2004. A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. Am J Hum Genet. 75(2):338–345.

Athey TW. 2005. Haplogroup prediction from Y-STR values using an allele-frequency approach. J Genet Geneal. 1:1–7.

Athey TW. 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J Genet Geneal. 2:34–39.

Barbier M. 2003. Le conflit du Sahara occidental [The conflict of Western Sahara]: Réédition d'un livre paru en 1982 [Reissue of a book published in 1982]. Paris: L'Harmattan.

Bekada A, Benhamamouch S, Boudjema A, Fodil M, Menegon S, Torre C, Robino C. 2010. Analysis of 21 X-chromosomal STRs in an Algerian population sample. Int J Legal Med. 124(4):287–294.

Bekada A, Fregel R, Cabrera VM, Larruga JM, Pestano J, Benhamamouch S, González AM. 2013. Introducing the Algerian mitochondrial DNA and Y-chromosome profiles into the North African landscape. PLoS One. 8(2):e56775.

Bekada A, Arauna LR, Deba T, Calafell F, Benhamamouch S, Comas D. 2015. Genetic heterogeneity in Algerian human populations. PloS One. 10(9):e0138453.

Bisson J, 1999. Gourara, Encyclopédie berbère. Aix-en-Provence: Edisud. (Gland – Hadjarien; vol. 21).

Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E, Bertranpetit J. 1997. Population history of North Africa: evidence from classical genetic markers. Hum Biol. 69(3):295–311.

Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, Mateu E, Martínez-Arias R, Morera B, Brakez Z, Akhayat O, et al. 2000. Genetic structure of north-west Africa revealed by STR analysis. Eur J Hum Genet. 8(5):360–366.

Camps G. 1974. Les civilisations préhistoriques de l'Afrique du Nord et du Sahara. Paris: Doin.

Camps G. 1980. Berbères: aux marges de l'histoire. Toulouse: Éditions des Hespérides.

Chiaroni J, King RJ, Myres NM, Henn BM, Ducourneau A, Mitchell MJ, Boetsch G, Sheikha I, Lin AA, Nik-Ahd M, et al. 2010. The emergence of Y-chromosome haplogroup J1e among Arabic-speaking populations. Eur J Hum Genet. 18(3):348–353.

Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. Hum Genet. 107(4):312–319.

Côrte-Real HBSM, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum genet. 60(4):331–350.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, 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(5):1014–1022.

Danver SL. 2015. Native peoples of the world: an Encylopedia of groups, cultures and contemporary issues. London: Routledge.

de Slane WMG. 1856. Histoire des berbères et des dynasties musulmanes de l'afrique septentrionale. Vol. 3: a translation of Ibn Khaldoun. Algiers: Imprimerie du

Gouvernement.

Dugoujon JM, Coudray C, Torroni A, Cruciani F, Scozzari R, Moral P, Louali N, Kossmann M. 2009. The Berber and the Berbers: Genetic and linguistic diversities. In: d'Errico F, Hombert JM, editors. Becoming eloquent: Advances in the emergence of language, Human cognition, and modern cultures. Amsterdam: John Benjamins Publishing Company; p. 123–145.

El-Sibai M, Platt DE, Haber M, Xue Y, Youhanna SC, Wells RS, Izaabel H, Sanyoura MF, Harmanani H, Bonab MA, et al. 2009. Geographical structure of the Y-chromosomal genetic landscape of the levant: A coastal-inland contrast. Ann Hum Genet. 73(6):568–581.

Elmrghni S, Coulson-Thomas YM, Kaddura M, Dixon RA, Williams DR. 2012. Population genetic data for 17 Y STR markers from Benghazi (East Libya). Forensic Sci Int Genet. 6(2):224–227.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 131(2):479–491.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 10(3):564–567.

Fadhlaoui-Zid K, Chennakrishnaiah S, Zemni R, Grinberg S, Herrera RJ, Benammar-Elgaaied A. 2012. Sousse, Tunisia: Tumultuous history and high Y-STR diversity. Electrophoresis. 33(23):3555–3563.

Fadhlaoui-Zid K, Haber M, Martínez-Cruz B, Zalloua P, Benammar-Elgaaied A, Comas D. 2013. Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. PLoS One. 8(11):e80293.

González-Pérez E, Esteban E, Via M, Gayà-Vidal M, Athanasiadis G, Dugoujon JM, Luna F, Mesa MS, Fuster V, Kandil M, et al. 2010. Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems). Am J Phys Anthropol. 141(3):430–439.

Haber M, Platt DE, Badro DA, Xue Y, El-Sibai M, Bonab MA, Youhanna SC, Saade S, Soria-Hernanz DF, Royyuru A, et al. 2011. Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon. Eur J Hum Genet. 19(3):334–340.

Hammer Ø, Harper DAT, Ryan PD. 2001. Paleontological statistics software: package for education and data analysis. Palaeontol Electron. 4(1):9.

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. 8(1):e1002397.

Iacovacci G, D'Atanasio E, Marini O, Coppa A, Sellitto D, Trombetta B, Berti A, Cruciani F. 2017. Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries. Forensic Sci Int Genet. 27:123–131.

Ivanova R, Astrinidis A, Djoulah S, Lepage V, Wijnen E, Hors J, Charron D. 1999. Mitochondrial DNA polymorphisms of a west Algerian population (Oran region). Biomed Pharmacother. 53(8):386–392.

Jones RJ, Tay GK, Mawart A, Alsafar H. 2017. Y-Chromosome haplotypes reveal relationships between populations of the Arabian Peninsula, North Africa and South Asia. Ann Hum Biol. 44(8):738–746.

Kundu S, Ghosh SK. 2015. Trend of different molecular markers in the last decades for studying human migrations. Gene. 556(2):81–90.

Laouina A, El Houate B, Yahia H, Azeddoug H, Boulouiz R, Chbel F. 2011. Allele frequencies and population data for 17 Y-STR loci (The AmpFlSTR® Y-filer™) in Casablanca resident population. Forensic Sci Int Genet. 5(1):e1–3.

Larmuseau MHD, Vessi A, Jobling MA, Van Geystelen A, Primativo G, Biondi G, Martínez-Labarga C, Ottoni C, Decorte R, Rickards O. 2015. The paternal landscape along the Bight of Benin–Testing regional representativeness of West-African population samples using Y-chromosomal markers. PloS One. 10(11):e0141510.

Lefevre-Witier P, Aireche H, Benabadji M, Darlu P, Melvin K, Sevin A, Crawford MH. 2006. Genetic structure of Algerian populations. Am J Hum Biol. 18(4):492–501.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet. 74(3):532–544.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A, 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet. 64(1):232–249.

Martinez-Cadenas C, Blanco-Verea A, Hernando B, Busby GBJ, Brion M, Carracedo A, Salas A, Capelli C. 2016. The relationship between surname frequency and Y chromosome variation in Spain. Eur J Hum Genet. 24(1):120–128.

Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, et al. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur J Hum Genet. 19(1):95–101.

Nei M, Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. Genetics. 97(1):145–163.

Nixon S. 2013. Tadmekka. Archéologie d'une ville caravanière des premiers temps du commerce transsaharien. Afriques (4). DOI: https://doi.org/10.4000/afriques.1237.

ONS: Office National des Statistiques. 2018. Algiers: Office National des Statistiques; [accessed 2018 August 22]. http://www.ons.dz/-Demographie-.html.

Ottoni C, Larmuseau MHD, Vanderheyden N, Martínez-Labarga C, Primativo G, Biondi G, Decorte R, Rickards O. 2011. Deep into the roots of the Libyan Tuareg: a genetic survey of their paternal heritage. Am J Phys Anthropol. 145(1):118–124.

Pereira L, Silva NM, Franco-Duarte R, Fernandes V, Pereira JB, Costa MD, Martins H, Soares P, Behar DM, Richards MB, et al. 2010. Population expansion in the North African late Pleistocene signalled by mitochondrial DNA haplogroup U6. BMC Evol Biol. 10(1):390.

Perrot d'Ablancourt N. 1667. L'Afrique de Marmol. Vol. 2: a translation of Marmol y Carvajal, Luis del. Paris: Louis Billaine.

Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, Bertranpetit J, Comas D. 2003. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. Ann Hum Genet. 67(4):312–328.

Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, Angustia SMT, Santos LH, Anslinger K, Bayer B, et al. 2014. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet. 12:12–23.

Ramos-Luis E, Blanco-Verea A, Brion M, Van Huffel V, Sanchez-Diz P, Carracedo A. 2014. Y-chromosomal DNA analysis in French male lineages. Forensic Sci Int Genet. 9:162–168.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics. 105(3):767–779.

Rice WR. 1989. Analyzing tables of statistical tests. Evolution 43(1):223-225.

Robino C, Crobu F, Di Gaetano C, Bekada A, Benhamamouch S, Cerutti N, Piazza A, Inturri S, Torre C. 2008. Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. Int J Legal Med. 122(3):251–255.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 139(1):457–462.

Solé-Morata N, García-Fernández C, Urasin V, Bekada A, Fadhlaoui-Zid K, Zalloua P, Comas D, Calafell F. 2017. Whole Y-chromosome sequences reveal an extremely recent origin of the most common North African paternal lineage E-M183 (M81). Sci Rep. 7(1):15941.

Taqi Z, Alenizi M, Alenizi H, Ismael S, Dukhyil AAB, Nazir M., Sanqoor S, Al Harbi E, Al-Jaber J, Theyab J, et al. 2015. Population genetics of 23 Y-STR markers in Kuwaiti population. Forensic Sci Int Genet. 16:203–204.

Tokdemir M, Tunçez FT. 2017. Genetic polymorphisms of 17 Y-STR loci in Eastern Turkey population. Gene Rep. 6:15–18.

Triki-Fendri S, Sánchez-Diz P, Rey-González D, Ayadi I, Alfadhli S, Rebai A, Carracedo Á. 2013. Population genetics of 17 Y-STR markers in West Libya (Tripoli region). Forensic Sci Int Genet. 7(3):e59–e61.

Turrina S, Caratti S, Ferrian M, De Leo D. 2015. Haplotype data and mutation rates for the 23 Y-STR loci of PowerPlex® Y 23 System in a Northeast Italian population sample. Int J Legal Med. 129(4):725–728.

Vermeulen M, Wollstein A, van der Gaag K, Lao O, Xue Y, Wang Q, Roewer L, Knoblauch H, Tyler-Smith C, de Knijff P, et al. 2009. Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms. Forensic Sci Int Genet. 3(4):205–213.

Zalloua PA, Xue Y, Khalife J, Makhoul N, Debiane L, Platt DE, Royyuru AK, Herrera RJ, Hernanz DFS, Blue-Smith J, et al. 2008. Y-chromosomal diversity in Lebanon is structured by recent historical events. Am J Hum Genet. 82(4):873–882.

Table 1. Haplotype diversity and forensic parameter estimates for the PowerPlex® Y23 panel.

Number of haplotypes observed	Haplotype representation
n = 1	154
n = 2	18
n = 3	3
n = 4	1
n = 6	1
n = 9	1
Haplotype diversity	0.996448653
Match probability	0.008122
Distinct haplotypes	178
Discrimination capacity	0.816513761

Table 2. Y chromosome haplogroup frequencies distribution in the present study.

Most common	Whit Athey's haplogroup predictor			Vadim Urasin's haplogroup predictor		
	Predicted haplogroup	**Number of times observed**	**Frequency (%)**	**Predicted haplogroup**	**Number of times observed**	**Frequency (%)**
Sub-Saharan Africa (Western and Central region)	E1b1a	10	4.6	E1b1a	10	4.6
North Africa/East Africa	E1b1b	186	85.3	E1b1b1-M35	02	84.4
				E1b1b1a1*-M78 (xV12,V13,V22,V65)	46	
				E1b1b1a1-M78	02	
				E1b1b1a1a-V12	01	
				E1b1b1a1b-V13	02	
				E1b1b1a1c-V22	25	
				E1b1b1b-M81	106	
Middle East	J1	10	4.6	J1-M267	10	4.6
	J2b	2	0.9	J2-M172	2	0.9
Western Europe	R1b	1	0.5	/	/	/
	/	/	/	I2-M438	1	0.5
Eastern Europe Asia (west and central regions)	R1a	5	2.3	R1a1a1*	5	2.3
Asia	L	3	1.4	L1b*	1	0.5
	/	/	/	F-M89	1	0.5
	/	/	/	K-M9	1	0.5
Asia/America	Q	1	0.5	/	/	/
Middle East/East Africa/ Central Asia	/	/	/	T2-L162	3	1.4

Table 3. AMOVA analysis based on 14 Y-STR in the Algerian samples.

Groups	Among Groups		Among Populations Within Groups		Among all Populations	
	Fct	***p*-value**	***Fsc***	***p*-value**	***Fst***	***p*-value**
All populations (n=7)	/	/	/	/	0.076	0.0000
Present Study vs Algiers [Arab]	/	/	/	/	0.045	0.02
Geographical location groups						
Northeast vs Northwest vs North central vs South (All 4 groups)	0.013	0.31	0.065	0.00000	0.076	0.00000
Northeast vs Northwest	0.087	0.35	-0.004	0.75	0.083	0.00000
Northeast vs South	-0.099	1.00	0.167	0.00000	0.084	0.00000
Spoken language / dialects groups						
Berbers vs Arabs	0.002	0.373	0.074	0.00000	0.076	0.00000
Geographical location groups Group 1 (Northeast Algeria)= Present Study Group 2 (North central Algeria) = Algiers [Arab] Group 3 (Northwest Algeria)= Oran 1 [Arab] and Oran 2 [Arab] Group 4 (South Algeria)= Ghardaïa [Mozabite], Tindouf [Reguibate] and Gourera/Timimoun [Zenata] **Spoken language / dialects groups** Group 1 (Berber Dialects-speaking)= Present Study, Ghardaïa [Mozabite] and Gourera/Timimoun [Zenata] Group 2 (Arabic Dialects-speaking)= Oran 1, Oran 2, Algiers and Tindouf [Reguibate]						

LIST OF FIGURES

LIST OF SUPPLEMENTARY TABLES

Supplementary Table 1. List of populations used in the present study for comparative analyses.

Supplementary Table 2. Y-STR haplotypes and haplogroups in the Aurès region sample (N = 218), using haplogroup predictors.

Supplementary Table 3. Allele frequencies and gene diversity values for the 23 Y-STR loci in 218 samples from the Northeast of Algeria living in Aurès region.

Supplementary Table 4. Pairwise *Rst* genetic distances (below diagonal) and associated *p*-values (above diagonal) estimated using 14 Y-STR between the Aurès sample, analysed in the present study, and other Algerian samples previously published.

Supplementary Table 5. Pairwise *Rst* genetic distances (below diagonal) and associated *p*-values (above diagonal) estimated using 14 Y-STR between the Aurès sample, analysed in the present study, and previously published populations from North Africa, Sub-Saharan Africa, Middle East and Europe.