High-resolution characterization of genetic markers in the Arabian Peninsula and Near East

Verónica Cristina Neves da Nova Fernandes

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Biology
Faculty of Biological Sciences

November, 2013

I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. I confirm that appropriate credit has been given within the thesis where reference has been made to the work of others.

I performed the study of the complete mitochondrial DNA sequencing of the three minor West Eurasian mitochondrial haplogroups, N1, N2 and X, in a collaborative study between University of Leeds, IPATIMUP (the second Institution where I am doing the PhD) and other laboratories. This study is already published in The American Journal of Human Genetics in which I am the first author.[1] I performed the laboratory work and also the phylogeography and phylogenetic analyses. The follow authors: Alshamali F., Cherni L., Harich N. and Cerny V., provided the biological samples. Costa M. D., Pereira J. B., Soares P., helped me in the construction of an Excel database with control region (HVS-I and HVS-II) of the mitochondrial DNA sequences already published and deposited in GenBank or reported in the papers and Alves, M. constructed bioinformatic tools used in this study. Finally, Soares P., Richards M. B. and Pereira L., helped me in interpreting the results and drafting the manuscript.

Acknowledgements

Firstly I would like to express my sincere gratitude to my PhD supervisors, Martin Richards and Luisa Pereira for all the support that allowed this research to be undertaken. To Prof. Martin Richards I am truly grateful for his guidance, support and immense knowledge. I am really thankful for all Luisa Pereira's patience, motivation and enthusiasm. More than a supervisor she became a friend, being always present and guiding me in my research and helping me writing this thesis. I could not have imagined having a better supervisor and mentor.

I would like to acknowledge the Fundação para a Ciência e a Tecnologia, who financially supported my PhD project through the scholarship SFRH/BD/61342/2009 and the project PTDC/CS-ANT/113832/2009.

I would like to thank to my two host institutions: IPATIMUP, for the excellent laboratory conditions and staff and most important for the exceptional work environment that allowed me to work with truly friends; and the University of Leeds for all the experience that allowed me to grow not only intellectually but also socially, offering me new friendships and a lot of good moments.

I also would like to thank to all my co-workers in the IPATIMUP and in the University of Leeds. Particularly I am very grateful to Pedro Soares for all the scientific advices and help during the statistical analyses; to Petr Triska for helping me with the Genome wide analyses; to Nuno Silva, Marco Alves and Bruno Cavadas for the computer skills; and to Maria Pala for all the laboratory help. I would like to acknowledge Farida Alshamali and Victor Cerny for the Dubai and Yemen samples that without that this research would be scientifically poorer.

I am truly thankful to all my friends, for being always there cheering me up, and for all the unforgettable and wonderful moments.

At last, but not the least, I would like to thank to my family, for encouraging me and supporting me spiritually throughout my life.

Abstract

In this work, I analysed the maternally transmitted mtDNA and biparentally inherited genome-wide polymorphisms to shed light on successive migrations into and from the Arabian Peninsula, at the crossroads between Africa, Europe and Asia. I focused on three main issues:

1- The first descendants of the out-of-Africa migration: My phylogeographic analysis on 385 complete mtDNA N1, N2 and X sequences showed their common origin in the Gulf Oasis region at ~57-65 ka. Instead of isolation, I identified a continuous gene flow between Arabia and Near East. Genome-wide data supported the strong clustering of Arabian and Near Eastern populations.

2- *Major population expansions in the Arabian Peninsula*: My data on the rare N(xR) lineages supported a continuous settlement of the Peninsula, while my founder analysis of other N lineages suggested Near Eastern lineages arriving continuously from the Late Glacial (31%-46%; some U and N1 lineages), Younger Dryas (24-28%; R0a and HV), Neolithic (20-25%; J and T; supported by my new complete 44 sequences) and till recently (10-15%; derived lineages). These results again challenge the hypothesis of long-term isolation between these two regions.

3- Genetic exchanges across the Red Sea: Phylogeographic analysis of L4 and L6 mtDNA complete sequences and HVS-I founder analysis from Africa into Arabia/Near East indicated that the Arab maritime dominance and slave trade (0.5-2.5 ka) were the main contributors (~60-70%) to the African input, but the entrance began with the establishment of maritime networks in the Red Sea by 8 ka. Genome-wide analyses supported this recent introduction (ROLLOFF estimates of 30-40 generations) suggesting that Arabia has 6% eastern African and 3% western African input. The HVS-I founder analysis for the back-to-Africa migrations showed that the Late Glacial period dominated introductions into eastern Africa, while the Neolithic was more important for migrations towards North Africa.

Table of Contents

List of Tables

List of Figures

I. LIST OF ABBREVIATIONS

A- Adenine

AD- Anno Domini

ADP- Adenosine diphosphate

AMH- Anatomical modern human

AMS- Accelerator mass spectrometry

ATP- Adenosine triphosphate

bp- Base pairs

BSP- Bayesian skyline plot

C- Cytosine

Ca2+- Calcium

^{12}C, ^{13}C and ^{14}C - Carbon-12, Carbon-13 and Carbon-14

CNPs- Copy number polymorphisms

COI, COII, and COIII- Cytochrome c oxidases

CRS- Cambridge reference sequence

cytb- Cytochrome b

ddNTPs – dideoxynucleotide triphosphates

dHPLC- Denaturing high performance liquid chromatography

D-loop- Displacement-loop in mitochondrial DNA

dNTPs – deoxyribonucleotides tryphosphates

DNA – Deoxyribonucleic acid

ESR- Electron spin resonance

G- Guanine

Gb- Gigabases

GWAS- Genome-wide association studies

HGP- The Human Genome Project

HKY85- Kishino and Yano 1985 model

HSP1 and HSP2- Spaced promoters for the H-strand

H-strand- Heavy strand

HUGO- Human Genome Organisation

HVS-I, HVS-II and HVS-III- Hypervariable regions or segments I, II and III

IDW- Inverse distance weighted

IOM- Indian ocean monsoon

ka– Thousand years ago

kb- Kilobases

Km- Kilometre

LD- Linkage disequilibrium

LGM- Last glacial maximum

LHON- Leber's hereditary optic neuropathy

LSP- Spaced promoter for the L-strand,

L-strand- Light strand

m- Meter

MAF- Minor allele frequency

Mb- Mega base pairs

MCMC- Markov-chain Monte Carlo

MIS – Marine isotope stage

ML- Maximum likelihood

MP- Maximum parsimony

MRCA- Most recent common ancestor

mRNA- Messenger ribonucleic acid

MSA- Middle Stone Age

MSY- Male-specific region of the Y chromosome

mtDNA- Mitochondrial DNA

MTERF1, MTERF2 and MTERF3- Mitochondrial transcription termination factors

^{14}N- Nitrogen-14

NADH- Nicotinamide adenine dehydrogenase

ND1,ND2,ND3,ND4 and ND5- NADH-Ubiquinone oxidoreductases

nDNA- Nuclear DNA

NJ- Neighbour-joining

NUMT- Nuclear copies of mitochondrial DNA

OH- Mitochondrial origin of replication of the heavy strain

OIS- Oxygen isotope stage

OL- Mitochondrial origin of replication of the light strain

PAML - Phylogenetic Analysis by Maximum Likelihood

PCA- Principal components analysis

PCR- Polymerase chain reaction

Pi- Phosphate

POLRMT- Mitochondrial DNA-directed RNA polymerase

PPNB- Pre-pottery Neolithic B

rCRS- Revised Cambridge reference sequence

RFLP- Restriction fragment length polymorphism

RNA- Ribonucleic acid

ROS- Reactive oxygen species

rRNA- Ribosomal RNA

SNP- Single nucleotide polymorphism

STR- Short tandem-repeat polymorphism

T- Thymine

TFAM- Mitochondrial transcription factor A

TFB2M- Mitochondrial dimethyladenosine transferase 2

TL- Thermoluminescence

TMRCA - Time of the most recent common ancestor

tRNAs- Transfer RNA

U- Uracil

UAE- United Arab Emirates

UPGMA- Unweighted Pair-Group Method with Arithmetic Mean

UV- Ultraviolet light

YCC- Y-chromosome consortium

II. INTRODUCTION

1. The study of human evolution

Information about the human past and present, derived from its evolution as a species, can be obtained from several lines of evidence such as archaeological remains, fossils, climate record, linguistic diversity and also studies of modern human genetic variation. The disciplines that interpret this evidence are, respectively, archaeology, palaeontology, climatology, linguistics, and genetics. The combination and exchange of information between these different fields allows a very detailed perspective of the human history, in a way impossible to attain if working independently.

The Arabian Peninsula has been under the joint focus of these disciplines as it is thought to have been the first way-station of the successful modern human out of Africa migration, occurring around 70 thousand years ago (ka). By providing information from the maternally transmitted mitochondrial DNA and genome-wide diversity I will contribute to clarifying the role played by this fundamental geographical bridge between continents. Here was the crossroads where the ancestral out-of-Africa population began to split and move towards the Near East, Europe, Asia, America and Australia.

1.1. Archaeology

Archaeology provides contemporary interpretations of past societies, human development and settlements across the world based upon past artefacts, such as pottery, tools, ornaments, waste deposits, houses and landscapes.[2; 3]

Dating of the artefacts is of extreme importance in archaeology, usually relying on measuring the radioactive decay of some isotope present in the sample. For the modern human time-scale in which we are interested, radiocarbon dating is the method most often used. This method has been established since the 1940s and relies on the fact that the cosmic ray bombardment of ^{14}N in the upper atmosphere produces ^{14}C which is incorporated into biological materials during their lifetime, along with stable ^{12}C and ^{13}C. This incorporation ceases at death, and the ^{14}C then decays exponentially with a ‘half life’ (the time taken to the ^{14}C to be reduced by half) of 5,730 years. The time span since death can be inferred from the amount of ^{14}C remaining in the biological specimen, with good precision up to 50 ka with AMS (accelerator mass spectrometry).[4] As the amount of ^{14}C in the atmosphere has changed over time, a measured proportion of ^{14}C (^{14}C years) cannot be directly

converted into a date (calendar years), implying the need for a calibration against an independent standard, known as calibration curve.[5] These curves can be based on absolutely dated tree-ring chronologies (limited to the last 12,594 years), marine archives (corals and planktonic foraminifera), lake varves and a stalagmite that contained both thorium-230 (allowing measurement of the calendar age) and ^{14}C in the calcium carbonate (allowing measurement of the ^{14}C age).[2] There are other radiometric techniques, such as potassium-argon dating or uranium-lead dating, that can provide dates for older sites than the ones obtained by radiocarbon dating. Potassium-argon dating, based on potassium-40 decay into argon-40 and calcium-40, can theoretically date samples from 4.54 billion years to 100 ka, but is limited to volcanic material, while uranium-lead dating uses the properties of the radioactive half-life of uranium-238 and has a dating range between 1 million years and over 4.5 billion years. A biasing factor for radiometric dating is contamination with extraneous radiometric source during the sample collection.

The application of chronometric techniques as thermoluminescence (TL) and electron spin resonance (ESR) are other ways to date the artifacts by examining the effects of radioactive impurities on the crystal structure of minerals. The natural crystalline materials when exposed to sunlight or heat can be distorted, forming local humps and dips in the electric field. Thermoluminescence measures the time elapsed of the accumulated radiation formed by the electrons trapped since the material containing crystalline minerals was previously exposed to sunlight. Electron spin resonance is similar to thermoluminescence and measures the unpaired electrons by exciting the electron spins and without interfering with the sample properties. In the spin-trapping method the trapping electrons react with short-lived radicals, being transformed to long-lived radicals called spin-adducts which are late observed by the ESR spectra.[6] Thermoluminescence is mostly used in pottery and old ceramic while electron spin resonance are preferable used in teeth, shells and stalagmite calcite. In order to obtain good results, the best option is to collect as many dates as possible and apply several dating methods in order to have a more reliable timescale.

1.2. Palaeontology

Palaeontology is the study of the fossilized remains of living organisms[2], aiming to infer the evolutionary events of the ancient populations/species and locate them in time and space.[7] In the 19th century, the study of modern human variation was

centred on the study of the fossil record present in Europe, with few hominid fossil being found: Neanderthals, early modern humans and *Homo erectus.*[8] The increasing evidence that Africa may had played a major role on the human evolution led finally to the move of palaeontology research to Africa, in the 1950s and 1960s, and several studies were carried out on different species like *Australopithecus*, *Paranthropus*, and early members of the genus *Homo* including *H. habilis* and *H. rudolfensis.*[8]

The African studies revealed that modern human anatomy appeared firstly in Africa, at around 200 ka, with the oldest remains found in Ethiopia, in Kibish and Herto dated to 190-200 ka and 154-160 ka, respectively.[9-11] Modern humans remains were also found in South African sites (Blombos cave, Border cave and Klasies River Mouth) dating to 80 and more than 100 ka, conferring some uncertainty about the point of origin of the species.

Curiously, a genetic-driven hypothesis is recently promoting an intense archaeological survey in a region that was previously almost overlooked, the Arabian Peninsula[12], which is the geographical focus of the current thesis. We will come back to this issue in another section of the Introduction.

1.3. Climatology

Climate played an important role in the spread of humans, having interacted with biological evolution throughout the Earth's history.[2] In fact, extreme climatic changes alter the ecological structure and resource availability, leading to extinction, speciation and behavioural alterations, while positive climatic conditions can be the motor of expansions and exploration of new environments.[13]

The study of these past climate changes, named palaeoclimatology, is based upon information gathered from several natural sources such as rocks, lake sediments, ice sheets, tree rings, corals and shells. By the analysis of the oxygen isotope composition obtained from deep-sea core sediments, it is possible to use the oxygen isotopic composition of the global oceans to distinguish marine isotope stages (abbreviated as MIS; also known as OIS from oxygen isotope stage) and create a global stratigraphic framework for the marine sediment (Figure1).[14]

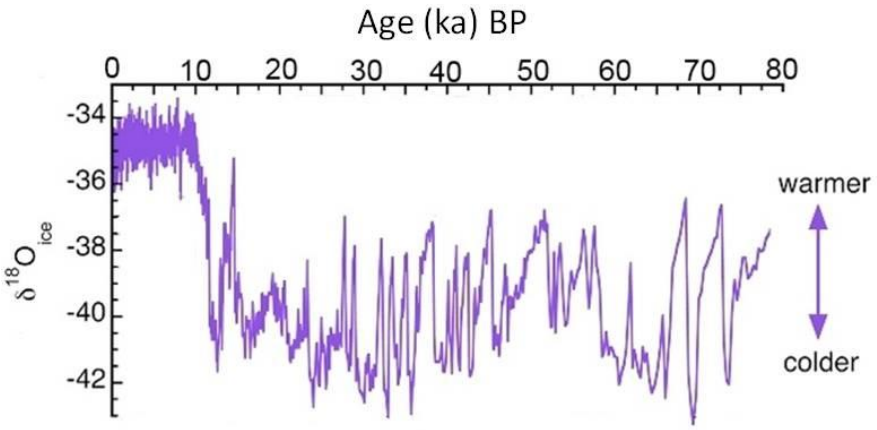

Figure 1 Oxygen isotope record ($\delta^{18}O_{ice}$) from the Greenland Ice Sheet Project II ice core, for the last 80,000 years (adapted from http://www.nature.com/scitable).[15] Lower negative $\delta^{18}O_{ice}$ values represent warmer air temperatures.

Changes in the positions of the continents, the amount of carbon dioxide in the atmosphere and the earth's orbit[16; 17] are responsible for more than 10ºC variation of the Earth's average temperature, oscillating between warm and cold conditions, over the last two million years. The low temperatures are responsible for the water being trapped in the polar ice caps (glaciers), with a corresponding decrease in the volume of the oceans. Although formation of glaciers rendered large regions uninhabitable, the low sea level made easier to cross water obstacles and allowed human dispersals.[18] The changes in the temperature were also accompanied by differences in precipitation and fluctuations in humidity, which affected the fauna and flora of the Earth and contributed to the expansion of desert areas.

Focusing on the time-scale important for this thesis, the last interglacial period occurred between 130-110 ka, with temperatures similar to the current ones. After 110 ka, the global climate gradually cooled, with temperatures being slightly lower than at present although some intense cold periods were observed in certain regions of the globe and severe aridity led to the reduction of the lake's water volume by at least 95% in Africa.[19] By the glacial maximum at 75-60 ka temperatures were significantly lower and this period could have been triggered by the large volcanic eruption of Mount Toba in Sumatra, Indonesia, which produced deposits of ash as far away as India, with a major impact on the environment and sea level across the world. After 70 ka, temperatures began to increase in a highly variable way, accompanied by periods of wetter conditions, although the sea level kept very low, about 70 m below the current levels; in this period expansion and migration of early modern human populations may have been stimulated.[19] The last glacial maximum (LGM) took place between 25-19 ka, with sea levels 120 m below

the present levels, extensive ice sheets covering northern Europe and large desert areas such as most of Australia and northern China. Most probably human populations survived during the glacial maxima at 75-60 and 25-19 ka through the contraction into refugial areas.[20] This climatic instability during the Pleistocene continued until 13 ka, when Europe returned to a full glacial condition, known as the Younger Dryas, between 13-11.5 ka, followed by a rapid warming in climate. The beginning of the Holocene epoch, also known as MIS 1, established the present warm and unusually stable climatic conditions.

1.4. Linguistics

Human language is both highly diverse and structurally complex, rendering it unique amongst animal communication systems.[21; 22] There are about 7,000 existing human languages with a variety of tones, clicks or manual signs.[21; 23] The language is culturally transmitted over generations and consequently the linguistic changes can be associated with the cultural evolution of human migrations.[2; 22]

It is agreed that the number of phonemes (the basic linguistic units) in a language is positively correlated with the size of its speaker population and that phonemes are more likely to be lost in small founder populations. Subsequently, in a scenario of population expansion, the phonemic diversity is reduced by increasing the distance from the point of origin.[24] It is also believed that the geographical spread of hunter-gatherer groups has been responsible for linguistically isolated populations which increase the diversity of languages.[23] Therefore by studying the worldwide languages it is possible to trace the ancestral languages (proto-languages) and to trace the human cultural and historical movements.[23] For example, some authors[24], through the analysis of phonemic geographic variation showed that outside Africa, the highest levels of phonemic diversity are found in language families of Southeast Asia indicating a population growth in this region experienced immediately after the African exodus.[24] Despite these positive results, the high variability and diversity of languages make it extremely difficult to date their evolution and can give unreliable rates leading to misinterpretations of the human history.[23]

1.5. Genetics

The use of genetics to investigate human origins began in the 20th century and several authors started to integrate palaeontology with genetics to establish the relationship between ancient and contemporary populations as well as to study the past population structure, changes in population size, and evolutionary relationships between taxa.[8; 25; 26] The considerable amount of population genetic data available worldwide, the genomic characterisation of some ancient samples (ancient DNA) and the development of new statistical methods is promising in the genetic-based approach to the study of the human past.

1.5.1. The DNA Molecule and the coding of information

The deoxyribonucleic acid (DNA) is a nucleic acid macromolecule encoding the genetic information as a sequence of monomeric subunits called nucleotides (Figure 2). These nucleotides are present in four different forms according to the type of base present: adenine, guanine, cytosine and thymine, abbreviated as A, G, C and T. Adenine and guanine are double-ringed molecules called purines, while cytosine and thymine are single-ringed molecules called pyrimidines. Each base is linked to a sugar molecule, deoxyribose, and each sugar has a phosphate group attached to it. The phosphate group of one nucleotide bond to the sugar of another, and so on, forming the DNA sequence. Within the cell, most DNA is present in double-stranded helices, spiralling around each other in a complementary double helix, where an A from one chain pairs strictly with a T of the other chain and a C with a G.[2] Each strand of the double helix acts as a template for the DNA replication during the cell division: the two strands separate and each new chain is produced based on the pairing of the complementary bases. In this way, two new DNA molecules are formed for each daughter cell.

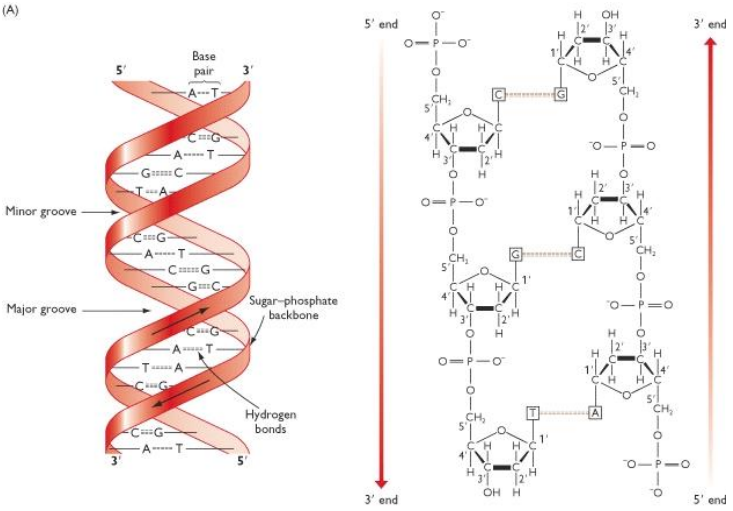

Figure 2 The double helix structure of DNA (adapted from Brown).[27]
Representation of the double helix structure of DNA (A) where the sugar-phosphate 'backbones' of each polynucleotide is represented in red and base pairs in black. In (B) is represented the chemical structure the base pairs and the base-pairing between two strands.

The DNA is packed in organised structures called chromosomes, but only a part of the molecule encodes functional or regulatory properties – this coding unit is called a gene. The genetic information contained in the genes must be transcribed into another type of nucleic acid called messenger ribonucleic acid (mRNA), which is a single-chain molecule containing the same nucleotides as the DNA, with the difference that the sugar is a ribose and the base thymine is replaced by uracil (U). The same complementary process between the bases (A complementing U) is used to produce the mRNA from a single DNA strand which acts as a template strand. This mRNA molecule will then be translated into protein in the ribosomes, according to the genetic code (Figure 3), which defines the correspondence between the four nucleotides and the 20 different amino acids constituting the proteins. Each triplet of bases in the mRNA, or codon, codes for a specific amino acid. Some codons can encode the same amino acid, and for this reason the genetic code is said to be redundant (there are some groups of four codons encoding the same amino acid, in which the base at the third position is synonymous; for example in Figure 3, CGU, CGC, CGA and CGG all code for the amino acid arginine). Stop codons do not encode any amino acid and act as a signal for the end of the protein; the translation of the protein always starts with a codon for methionine. The genetic code is almost universal for all organisms.

1st Letter		2nd Letter: U		2nd Letter: C		2nd Letter: A		2nd Letter: G		3rd Letter
	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
		UUC		UCC		UAC		UGC		C
		UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
		UUG		UCG		UAG	Stop	UGG	Trp	G
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
		CUC		CCC		CAC		CGC		C
		CUA		CCA		CAA	Gln	CGA		A
		CUG		CCG		CAG		CGG		G
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
		AUC		ACC		AAC		AGC		C
		AUA		ACA		AAA	Lys	AGA	Arg	A
		AUG	Met	ACG		AAG		AGG		G
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
		GUC		GCC		GAC		GGC		C
		GUA		GCA		GAA	Glu	GGA		A
		GUG		GCG		GAG		GGG		G

Figure 3 The nuclear universal genetic code.

Sometimes, during cell division, the sequence of bases is not properly copied, and a base can be replaced by another, or deleted or duplicated. These alterations in the genetic information can occur naturally at random, but also be caused by exogenous factors such as exposure to UV light or chemical substances. Usually, the cellular repair mechanisms are very efficient in repairing errors in the nucleotide sequence, but if not it causes a mutation or polymorphism. Some authors apply the term mutation only in cases where the error leads to a difference in the phenotype, using the term polymorphism to signify other differences relative to the original sequence. At the population level, a polymorphism must attain a frequency of 1% to be considered as such.

In the case of a single nucleotide polymorphism (abbreviated as SNP), in which a pyrimidine base changes for another pyrimidine (e.g., C for T), or a purine for another purine (e.g., A for G), the difference is called a transition, while a change from a purine to a pyrimidine, or *vice versa*, is called a transversion.[2] Transitions are much more common than transversions, due to the similar structure of the changed nucleotides. A base substitution in the coding region of the genome which leads to a change of the amino acid (usually mutations at the second position and some at the first position of the codon) is called nonsynonymous, missense or replacement substitution. On the other hand, a substitution that does not change the amino acid is a synonymous or silent substitution.[2] The deletions and duplications can be especially problematic in the coding region if not in multiples of three, as this cause a frameshift of reading of the protein, so that all amino acids

will be different from that point on, and a stop codon can be incorporated prematurely leading to a truncated protein.

By comparing the polymorphisms between the segments of DNA in different populations it is possible to establish patterns of nucleotide diversity within and among populations. This study is essential in order to understand the genome and the evolution of the species.

1.5.1.1. The autosomal nuclear genome

The human genome has 3.2 gigabases (Gb) of DNA stored in 23 chromosome which are present in pairs in diploid organisms, as in animals, located in the cell nucleus (Figure 4). Of the total chromosome pairs, 22 are autosomal, equally present in both females and males, and the remaining one sex-specific as two X-chromosomes present in females and one X-chromosome and one Y-chromosome in males.

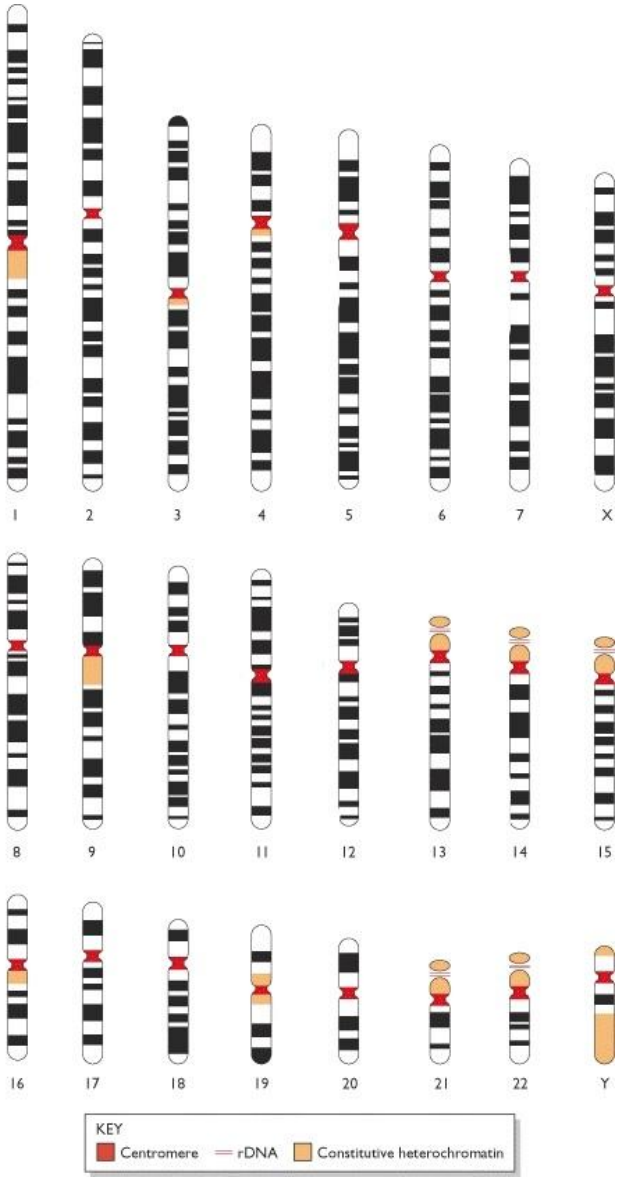

Figure 4 The human karyogram (adapted from Brown).[27]
Representation of the 22 pairs of autosomal chromosomes and one pair of sex chromosome with the centromere shown in red and the constitutive heterochomatin in light yellow. The chromosome numbers are given below each structure and the band number in the left of each chromosome.

Most of the current knowledge about the human genome derives from the international consortium launched in 1988, The Human Genome Project, abbreviated as HGP.[28] This aimed to: identify all the human genes; determine the sequences of the 3 Gb; store that information in databases; improve tools for data analysis; transfer related technologies to the private sector; and address ethical, legal and social issues. A substantially draft of the human genome was released in February 2001[29] (covering about 94% of the human genome), and in 2004[30] a high accuracy and nearly complete human genome sequence (covering approximately

99% of the euchromatic genome) was released. Although it represents the end of the sequencing *per se*, the analyses of results will continue for many years.

The results showed that the majority of the genome (~98%) does not code for any genetic information, that is, no genes were identified in those regions. Initially, it was believed that the non-coding regions did not have any role in cell metabolism, and were also known as junk DNA, but recently it was found that some of those portions are highly conserved, being under strong negative selection,[31] indicating that they play some functional role. In fact, the ENCODE[32] project claimed that ~80% of the human genome is functional, and that 76% is pervasively transcribed, including microRNAs which are involved in gene expression regulation but never used as intermediates in protein production. The number of human genes was initially estimated as being about 32,000, amounting only to twice as many in worms or flies, but being more complex, with more alternative splicing generating a larger number of proteins. But currently the number provided in the Ensembl database is lower, around 23,532 human genes.

The HGP[28], together with the SNP Consortium,[29] found that human SNPs amounted to more than 1.42 million, with an average spacing of 1.9 kb. This meant that there are about 15 SNPs for gene loci of average size, which could provide important insights in clinical genetics. Of course, the density of SNPs varies considerably across the genome, it is higher in portions were the local mutation rate is higher and also dependent on the “age” of the locus (defined as the average number of generations since the most recent common ancestor of two randomly chosen copies in the population) and selection events.

A key process in generating genetic diversity takes place during the pairing of the homologous chromosomes in the gametes' cell division, meiosis, and is known as recombination or crossing-over. The DNA strands of the paired maternal and paternal chromosomes break at the synapsis and there is DNA segment crossover or exchange between the pair of chromosomes, generating thus new combinations of alleles and increasing diversity. For generating new advantageous combinations of alleles and eliminating deleterious mutations from a population, recombination is an important evolutionary mechanism. Recombination thus destroys the haplotypic background (the combination of alleles/polymorphisms along the chromosome, descending from a common ancestor) inherited from each parent. SNPs that are closer in the genome have higher probability of being transmitted linked, as the occurrence of recombination between them will be lower; this pattern is known as linkage. Linkage is one of the factors influencing the widely used measure of linkage disequilibrium (LD), an estimate of the non-random association of alleles at

two or more loci. Other factors affecting LD are the rate of recombination, the rate of mutation, population structure, selection, non-random mating and genetic drift.

LD has been highly used as a tool for mapping disease genes and for probing population history. The collection of SNPs available from the HGP led to a clearer picture of the LD map across the genome, but a major effort was put together in a new international consortium designated HapMap.[33] Phase I[34] of this project identified 1,007,329 SNPs from many different sources, using various technologies, and then systematically genotyped them in 270 individuals (30 trios from Utah residents with northern and western European ancestry, 30 Nigerian Yoruban trios, 45 unrelated Japanese and 45 unrelated Chinese), yielding a genome-wide map of the haplotype structure in major human populations. Results allowed researchers to confirm that the regions with extensive LD over long spans were regions with low recombination intensity; sharply defined regions were found in which LD decayed very rapidly, consistent with local high levels of recombination. In theory, in an infinite population with only neutral variation, all LD would eventually decay, but in a finite population random genetic drift plays the main role in maintaining LD at equilibrium.

The SNP collection was later enlarged, based upon information from subsequently published complete genomes in several individuals, and Phase II of HapMap characterised more than 3.1 million SNPs (one SNP approximately every 1 kb)[35] in the same individuals, revealing that 10–30% of pairs of individuals within a population share at least one region of extended genetic identity arising from recent ancestry. In Phase III HapMap,[36] 1.6 million common SNPs were genotyped in 1,184 reference individuals from 11 global populations (the ones in Phase I and II, plus African-Americans, Chinese living in Denver, Gujarati Indians residing in Houston, Luhya and Maasai from Kenya, Mexicans living in Los Angeles, and Tuscans from Italy), and sequenced ten 100-kilobase regions in 692 of these individuals.

As expected, lower-frequency variation was less shared across populations, even closely related ones, highlighting the importance of sampling widely to achieve a comprehensive understanding of human variation. They also found that variants discovered through large-scale sequencing have longer haplotypes than more common variants.

1.5.1.2. The non-recombining segments of the genome

While the majority of our genome is biparentally inherited and thus undergoes recombination, there are two segments of the DNA which are non-recombining and uniparentally inherited: the maternally inherited mitochondrial DNA (mtDNA) genome; and the male specific portion of the non-recombining Y-chromosome (MSY) (Figure 5). The uniparental segments are revealing information about the history of human migrations. [37-39]

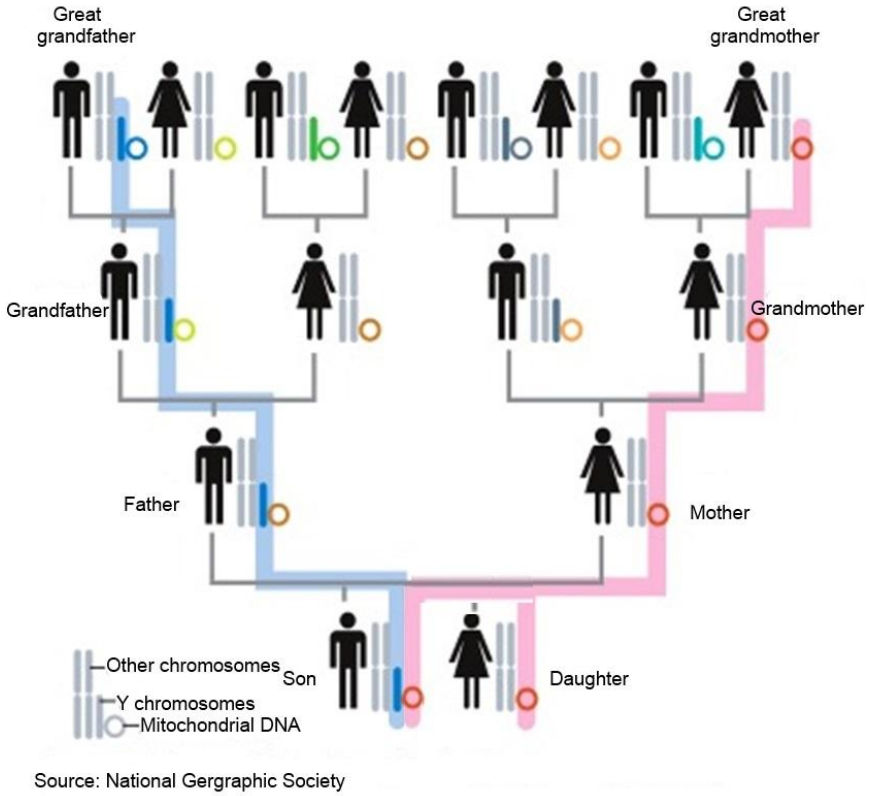

Figure 5 Uniparental inheritance (from the national geographic website https://genographic.nationalgeographic.com).[40]
The uniparental inheritance of the MSY and mtDNA reflects different aspects of human history. The MSY provides information passed from father to son and reflects the demographic history of males lineage (blue line); while the mtDNA presents the female to female transmitted lineage (pink line). Male children also inherit mtDNA but do not transmit it to their offspring

1.5.1.2.1. The mitochondrial DNA

The mitochondrion is a cytoplasmic organelle, present in almost all eukaryotic cells, and is highly dynamic, presenting a variety of shapes and containing a separate genome.[41-43] The number of mitochondria in a cell depends on the cell

type, with cells that require a high amount of energy, such as nerve or muscle cells, containing thousands of mitochondria, while other may contain just a few hundred.

The origin of mitochondria was addressed by Lynn Margulis in the 1960s, in a theory called endosymbiosis.[2] This proposes that mitochondria descend from an archaebacteria endosymbiont that was incorporated into eukaryotic cells about 1.5 billion years ago, in one of the symbiotic events that assembled the eukaryotic cell. The mtDNA maintained many features of the original prokaryotic DNA such as a circular rather than a linear genome, the absence of histones, discrete origins of replication and a different genetic code.[43; 44]

Mitochondria play several important functions in the human cellular activity, namely the production of cellular energy by actively synthesizing ATP from ADP and Pi through the mechanism of electron transport and oxidative phosphorylation.[42; 45] Mitochondria are also involved in apoptosis, the retrograde signalling pathways and Ca^{2+} signalling[45; 46], all important for the cellular activity regulation.[46]

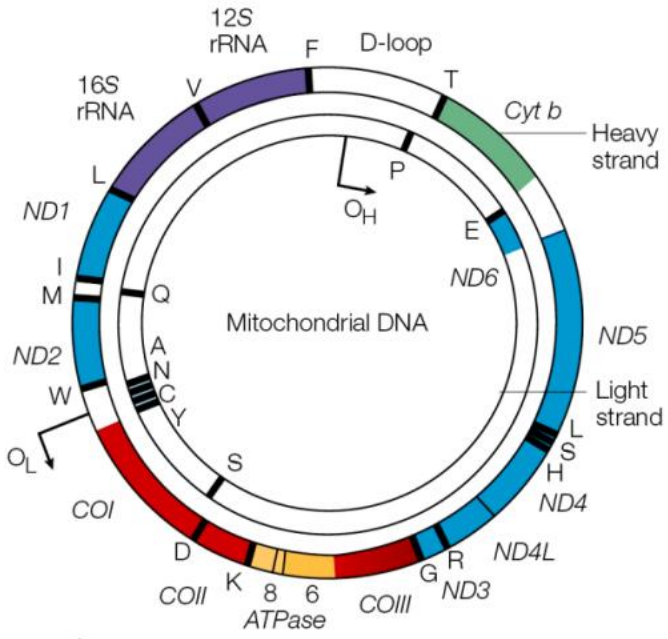

Figure 6 Mitochondrial DNA genome (adapted from http://www.nature.com/scitable).[15]
Both light and heavy strains are represented. NADH dehydrogenase subunits (blue), cytochrome oxidases (red), ATPase (yellow), rRNA (violet), tRNAs (black) and non-coding regions (white) are indicated.

Each mitochondrion contains a high copy number of its own genome (1,000-10,000 DNA molecules per cell), the mitochondrial DNA (mtDNA).[2; 47] As mtDNA does not undergo recombination, the genome is transmitted in a block from one

generation to the next, as a haplotype or lineage. Along the life of the individual, mutations can occur in some of mtDNA copies, so that the same mitochondrion/cell can bear these mutated mtDNA copies along with the wild type mtDNA genome, a condition known as heteroplasmy.

In animals the mtDNA genome (Figure 6) is organized in a small (15-20 kb), double stranded and circular molecule with a strikingly uniform structure.[43] In humans, the first complete mitochondrial genome sequencing was performed in a chimera of European placental tissue and in the African HeLa cell line and is known as the Cambridge Reference Sequence (CRS)[48], which was revised in 1991 (known as the revised or rCRS).[49] The two strands of the mtDNA are named as 'heavy' or 'light' according to their base content. The heavy (H) strand has a high concentration of G bases while the light (L) strand has a C-rich content,[2] and they can be easily separated by density gradient centrifugation. This molecule contains 37 genes (28 on the H-strand and nine on the L-strand) tightly packed within the 16.5 kb genome with no introns. These genes include two ribosomal genes, 22 tRNAs and 13 polypeptides of the oxidative phosphorylation system including ND1, ND2, ND3, ND4, ND4L, ND5, and ND6 of complex I (NADH dehydrogenase); cytochrome *b* (cytb) of complex III; COI, COII, and COIII of complex IV; and ATP6 and ATP8 of complex V.[50] Besides the long coding region, a significant stretch of the mtDNA, ~1.2 kb, forms a non-coding region, the control region or D-loop (standing for displacement-loop), extending from positions 16,024 to 576. This control region has some functionally important domains, containing regulatory elements essential for transcription and replication, and binding regions for DNA and RNA polymerases, as well as other regulatory and transcriptional factors, and is probably under selective pressure.[51] But two regions in the control region have a higher level of variation, known as hypervariable regions or segments I and II (HVS-I and HVS-II), ranging from positions 16,024-16,383 and 73-340, correspondingly.[52]

The mtDNA is replicated by a trimeric protein complex coded by the nuclear DNA,[53] although the event is otherwise totally independent of the cell phase or the nuclear DNA replication, taking place in the mitochondrial matrix (Figure 7). There is still some controversy about the replication process, with the most widely accepted asymmetric model suggesting that it involves two unidirectional and independent origins, the first at the O_H located in the control region leading to a daughter H-strand circle, and the second at the O_L located in the coding region producing a daughter L-strand circle.[45; 53] In the mtDNA transcription both strands seem to be completely transcribed from promoters situated in the D-loop region (a single promoter for the L-strand, LSP, and two closely-spaced promoters for the H-

strand, HSP1 and HSP2), and the two complete polycystronic transcripts are then cut into functional RNAs (tRNA, rRNA and mRNA) molecules.

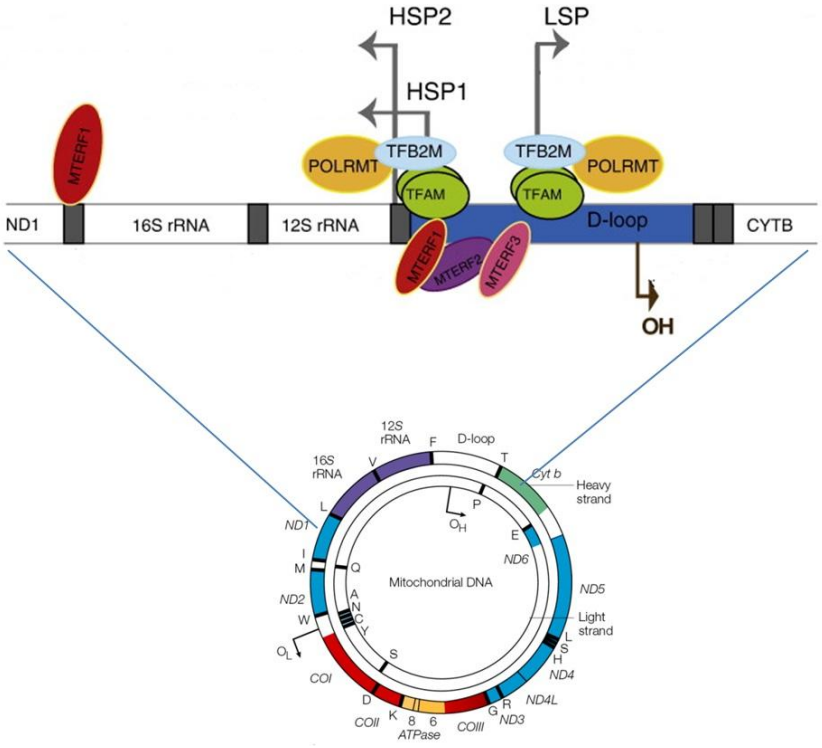

Figure 7 Model of mtDNA transcription initiation machinery and promoters (adapted from Peralta et al.[54] and http://www.nature.com/scitable).[15]
The bidirectional mitochondrial transcription model starts in the D-loop region where the promoters HSP1, HSP2 and LSP are located. The transcription initiation requires the cooperation of TFAM, TFB2M, POLRMT, MTERF1, MTERF2 and MTERF3. The OH indicates the origin of replication on the heavy strand.

The translation of the mtDNA encoded proteins follows a different genetic code to that of the nucleus.[42] The mitochondrial genetic code is different from the universal genetic code in: using AGA and AGG as stop codons rather than encoding arginine; UGA encoding tryptophan rather than being a stop codon; AUA codon encoding methionine instead of isoleucine.[53]

Other genes essential for mitochondrial function and for the synthesis of mitochondrial enzymes (such as the mtDNA polymerase) have been transferred to the nuclear genome,[2; 43] as well as more than 90% of the proteins involved in oxidative phosphorylation. A tight cross-talk between the mtDNA and nuclear DNA (nDNA) must operate in order to maintain the two-genome coded system, but it is so far totally obscure.

As mitochondria have extremely important cellular functions, it would be expected that the mtDNA would be highly selectively constrained,[53; 55] which was

shown to not be true.[56] The analysis of several DNA genomes in diverse species showed that mtDNA evolves 5 to 10 times faster than the nDNA.[55] The mtDNA mutation rate is also non-uniform along the molecule, with the more constrained coding region having a relatively lower mutation rate than the control region (an average of 5 times less variation) and, as already referred to, inside the control region, the two hypervariable regions or segments (HVS-I and HVS-II) are mutational hotspots.[2; 57] This higher mtDNA mutation rate may have possible several causes, as mtDNA is devoid of histones, a less effective mtDNA repair system than the one present in the nucleus,[53] a high concentration of the mutagenic reactive oxygen species (ROS) produced by the oxidative phosphorylation in the mitochondria matrix, and a higher turnover rate of the mtDNA than the nuclear DNA in the tissues, therefore requiring more replications per unit of time which can increase the possibility of generating more mtDNA errors.[55]

For many years it has been accepted that mtDNA is a non-recombining DNA, with a exclusively maternal inheritance.[47] The sperm cell mitochondria do enter the fertilized egg during fertilisation, but are vastly outnumbered (100,000 times or more) by the oocyte mitochondria, being selectively destroyed by a ubiquitin-dependent mechanism inside the oocyte cytoplasm and later subjected to proteolysis during the implantation process, they disappear in early embryogenesis.[58] The maternal inheritance was put in question by the description of a patient with a metabolic disorder leading to inability to do exercise, who displayed copies of the paternal mtDNA genome (bearing a two base pairs deletion in ND2 gene) in his muscle cells, amounting to 90% heteroplasmy.[59] Screenings of other datasets of mtDNA diseases did not reveal further cases of paternal inheritance and no healthy individual has ever been shown to have paternal inheritance of the mtDNA.[60; 61]

Another important issue is the possibility or not of recombination in the mtDNA. All phylogenetic statistical analyses based on the mtDNA diversity rely on the absence of recombination; if recombination occurred in the mtDNA, the phylogenetic tree reconstruction and dating of the time of the most recent common ancestor (TMRCA) might be inaccurate, leading to an overestimation of the mutation rate.[62-64] A study performed in the patient described above, having paternal mtDNA copies in the muscle cells, showed that recombination did occur in those heteroplasmic cells.[64] However heteroplasmy has been found mostly in somatic tissue and not in germ-line tissues in healthy individuals,[2; 65] and is an especially important issue when associated with disease phenotypes. Even if mtDNA might exceptionally undergo somatic recombination a number of events

would be needed in tandem for a recombinant to enter the gene pool. Carelli et al (2006) did not detect any existence of *in vivo* mtDNA recombination that could therefore affect the pedigree studies.[66]

A high number of human diseases caused by mitochondrial function abnormalities has been reported. These can be caused by mutations in the mitochondrial proteins either coded by the mtDNA or the nDNA, and can be both acquired along the life of the individual or inherited mutations. For the somatic mtDNA mutations, these have been reported mainly in adult post-mitotic cells and in cancer tissues, indicating that mtDNA variation has an important role in physiological aging and tumorigenesis.[67] Progressive neurodegenerative diseases such as Parkinson's disease and Alzheimer's have been associated with nuclear gene mutations that are involved in the mtDNA function.[68]

There are some mtDNA haplogroups that have been associated with mitochondrial disorders such as Leber's hereditary optic neuropathy (LHON). LHON was the first maternally inherited disease associated with mtDNA single mutations and is characterized by a degeneration of retinal ganglion cells that cause loss of central vision, a painless and sub-acute visual failure in young adult males. The clinical diagnosis is usually confirmed by molecular genetic analysis of three 'common' mtDNA mutations (3460G/A, 11778G/A and 14484T/C) in genes of the complex I.[53] After the detection of the LHON disease almost 200 different mtDNA mutations have been reported in humans and some of these mutation have been associated to a disease.[68] However, not all the inherited single base-pair variants present in the mtDNA are potentially pathogenic; most are neutral variants common in a population. Consequently, in order to identify mutations that are potentially pathogenic it is essential to carry out a deep study of the continent-specific mtDNA variants.[53; 69]

The discovery of the important role of mitochondria in some diseases led to the recent development of treatment research in mitochondria such as therapeutic interventions, transcriptional regulation, and genetic manipulation.[68]

1.5.1.2.2. The Y chromosome

Y chromosome is a male-specific haploid marker, were more than 90% of the length is encompassed by a non-recombining region, the largest non-recombining block in the human genome: 23 Mb of non-pseudoautosomal euchromatin coding for only 27 different proteins (1.2/Mb) (Figure 8).[2]

The Y chromosome variation is normally assessed by the study of two types of polymorphisms: microsatellites or short tandem-repeat polymorphisms (STRs) and bi-allelic markers (SNPs).[70; 71]

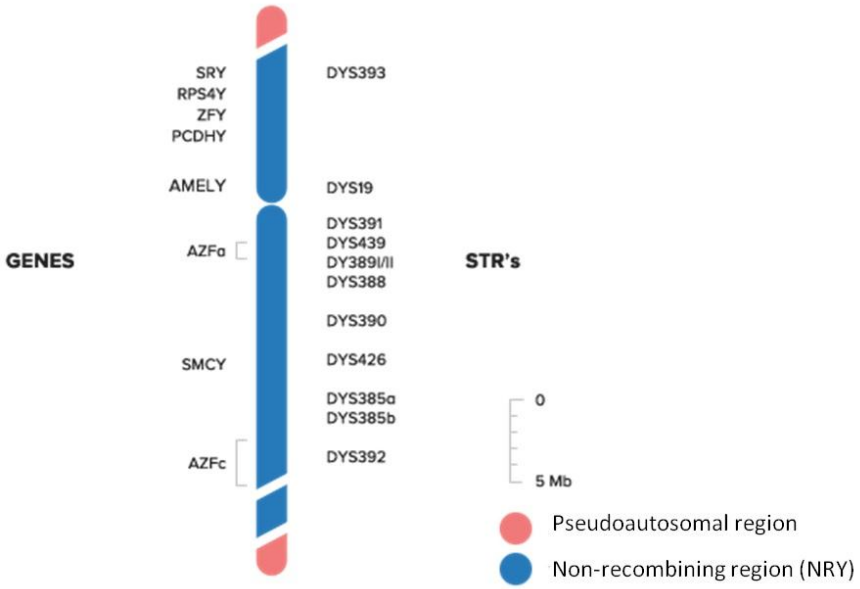

Figure 8 Human Y chromosome (adapted from the Genographic project https://genographic.nationalgeographic.com).[40]
Y chromosome with the information for the pseudoautosomal (pink) and non-recombining region (blue).

STRs consist of repetitions of a short motif (less than 10 bp), usually present in the non-coding region. The number of repetitions in each STR is highly variable (polymorphic) between individuals, and for that reason STRs are very useful in individual identification, being applied in forensic investigations and evolutionary studies.[72] Several commercial multiplex kits have been created with different numbers of Y-STRs, usually around 20.

A few STR mutation mechanisms have been proposed, including: unequal crossing-over in meiosis; retrotransposition; and replication slippage.[73] Authors agree that replication slippage is the main mechanism causing new mutations in microsatellites, so, a stepwise mutation model was developed, which assumes that a new allele is generated by the gain or loss of one repeat at a time, meaning that two alleles that differ by one repeat are more closely related than alleles differing by many repeats.[74] For Y-STRs, the mutation rate was estimated to be around 10^{-3},[75; 76] although locus-specific values can vary between $3x10^{-4}$ and $6.4x10^{-3}$ per generation. There are slight differences between pedigree and phylogenetic inferences of the mutation rate, with the latter three to four times slower.[77-80] The

discrepancy between the two kinds of mutation rate estimation can be explained by heterogeneity of the rate across the loci and the stochastic process of drift affecting phylogenies and not so much pedigrees.[80]

Thus, the advantage of the high mutation rate also implies a lower reliability in evolutionary inferences due to high recurrence (two individuals share a polymorphism by state but not by descent). For this reason, evolutionary studies based on the Y chromosome diversity combine microsatellite analyses with the study of binary markers or SNPs. SNPs usually mutate according to the infinite alleles model, in which each mutation event is unique and independent of all other SNPs.[81] The SNPs in the Y chromosome have a slow mutation rate similar to the autosomal SNPs, of 10^{-8} per base pair per generation.[82; 83] Due to their low mutation rate they can be considered as unique events in evolution. Therefore, they can be used to draw reliable phylogenetic inferences. Combinations of Y-SNPs are used to identify male lineages and to define haplogroups.[84]

2. Human population genetics

The genetic record of life contained in the DNA of all living individuals can be analysed to understand the human evolutionary history. The first human population studies were based on the characterisation of blood groups, followed by classical protein and immunological genetic markers.[85] A major result revealed by these studies was that humans are characterized by low intraspecies genetic variation, which occurs primarily within populations rather than between populations. Cavalli-Sforza was central in introducing the statistical evaluation of genetic data against models which incorporated inter-disciplinary information. By constructing models that approximate reality, it is possible to estimate parameters from the data and to test different hypotheses about the past.[86] These initial studies were focused in Europe and the Near East, for which a high amount of genetic information and well-documented archaeology data was available.

In the last 30 years, the great technological advances in genetics have extensively promoted the use of molecular genetics in the study of human history, leading to the emergence of a new discipline, "archaeogenetics".[87; 88] The study of the genetic diversity of living populations (using both haploid and diploid genetic markers), and moving backwards in time through generations till the most recent common ancestor, continues to be very informative. But technological improvements have also led to a second genetic approach to the study of the human past: by the analyses of ancient DNA from well preserved organic remains. Thus, not only the genetic diversity of successful ancestors can be inferred but also lineages that became extinct. By studying different parts of the genome, the ancestral histories of females, males or the general population can be ascertained.

2.1. Extant population analyses

2.1.1. Mitochondrial DNA diversity

The general attributes of the mtDNA render this genome very sensitive to demographic effects and therefore an excellent marker for analysis of population genetics and phylogenetic inferences throughout the world, from a female perspective.[89]

Various demographic parameters can be obtained through the coalescent process, which relates random samples of individuals, by moving backwards in time

throughout the generations, until the point of coalescence in the most recent common ancestor (MRCA)[2] (Figure 9a). The chronological information can be reconstructed by using a molecular clock, which assumes that any DNA sequence evolution is approximately constant over time.[90] This hypothesis has become an important tool in evolutionary biology, allowing timescales to be placed on evolutionary events and providing information about the ancient distribution of diversity (Figure 9b).[91; 92]

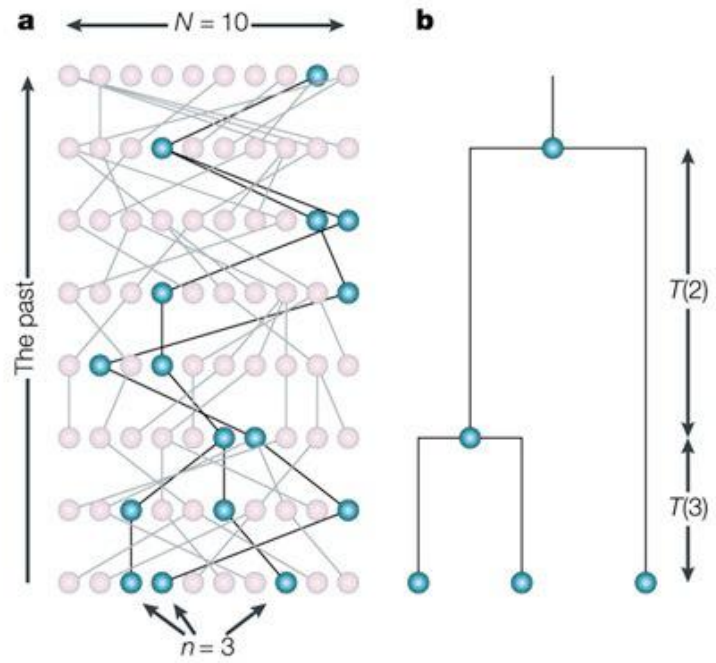

Figure 9 The basic principle behind the coalescent process (adapted from Rosenberg and Nordborg).[93]
Figure (a) shows the coalescent process of a given population, where by moving back in time the black lines trace the ancestries to a single common ancestor. Figure (b) represents the phylogeny of a hypothetical population undergoing expansion and illustrates the time between coalescence events (T2 and T3) and the tree topology.

The first mitochondrial DNA research applied to population genetics was conducted by Brown et al. (1980) in 21 samples from diverse populations, by digesting the entire molecule with 18 restriction enzymes.[56] The variation patterns of the restriction fragment length polymorphisms (RFLPs) were used to trace the human genetic history, indicating a common ancestor for all individuals with a coalescence age around 180 ka. This study was followed by another mtDNA-RFLP characterisation in 147 samples from several geographic regions.[26] The sequence divergence calculated for each population across seven functionally distinct regions of the human mtDNA showed that in general the African populations have the

highest diversity at a worldwide scale across all the functional regions. In addition, from the analyses of the tree of minimum length they also noticed that the primary branches of the tree led to African mtDNAs, showing therefore that the origin of modern humans was most probably in the African continent. The African MRCA of modern humans became known as the Mitochondrial Eve, dated to have lived around 200 ka.

The advent of the polymerase chain reaction (PCR) in the 1980's made sequencing segments of the genome in many samples more straightforward, which led to mtDNA studies being focused on HVS-I sequencing, sometime combined with RFLPs, resulting in a much more refined picture of the mtDNA world phylogeny.[94] However, as some of the control region positions have a high recurrence rate, the inferred phylogenies were reticulated, displaying homoplasy and conferring uncertainty in the branching pattern.[57; 95] Recently, a new phase of more comprehensive studies of the human mitochondrial molecule has begun with the sequencing of the entire molecule (16,568 bp), making possible the most resolved and precise inference for the mtDNA and providing new insights into the time and direction of human dispersals.[87; 90; 95]

In order to summarize the worldwide population diversity observed in the mtDNA trees, Torroni et al.[96] introduced an mtDNA nomenclature. The mtDNA haplotypes which are derived by descent from the same ancestor form a monophyletic group or clade designated as a haplogroup. Over time, diverse haplotypes affiliated within a certain haplogroup accumulate further diversity and form sub-haplogroups. The first four classified Native Americans haplogroups were named alphabetically as haplogroups A, B, C and D.[96] Later, the haplogroup structure of other continental populations was characterized[97] using always the same nomenclature, in which the major clades are named by single capital letters (e.g. haplogroup D), the sublineages within these clades are designated by the haplogroup letter followed by a numerical suffix (e.g. haplogroup D1), and further sub-lineages are named by alternating lower-case letters and numbers (e.g. D1a3b2).[2; 97] When a marker defined only a set of lineages but not all the known haplotypes for that haplogroup, indicating a paraphyletic group, the corresponding paragroup was named by using a * suffix (e.g. C*).

The human mtDNA tree is continental specific with branches typically African (L0, L1, L2, L4, L5 and L6), other observed in west Eurasian, North African and Central Asian populations (H, I, J, T, U/K, V, W and X) and branches embracing the majority of the lineages described in Asia, Oceania and Native Americans (A, B, F, O, P, S, T, M, Y and again X).[50; 89; 98] The tree structure reflects the history of

humans. Initially, a series of bifurcation events originated new haplogroups over time (L0, L1, L5, L2, L6 and L4), compatible with a low effective population size for most of the modern human time depth, from 200 to 70 ka.[87] The first multifurcation node gave rise to the haplogroup L3 (Figure 10), most likely reflecting local population growth within Africa (most of the L3 branches are largely restricted to this continent) and also leading to a successful colonization event responsible for all non-Africans haplogroups. Most probably this L3 expansion occurred in Eastern Africa, because the richest early branching around the L3 node is found in modern Eastern African samples.[99] Somewhere outside Africa (or in Eastern Africa), the basal L3 variation generated two major branches, the haplogroups M (mainly found in South Asian, East Asian and Oceania populations) and N (found at highest frequency in West Eurasian populations, but present throughout Eurasia, Oceania and Australasia). These two haplogroups cover the mtDNA pool of all non-Africans (excluding the descendants of migrations from Africa within the past few thousand years, such as those engendered by various slave trades).[87]

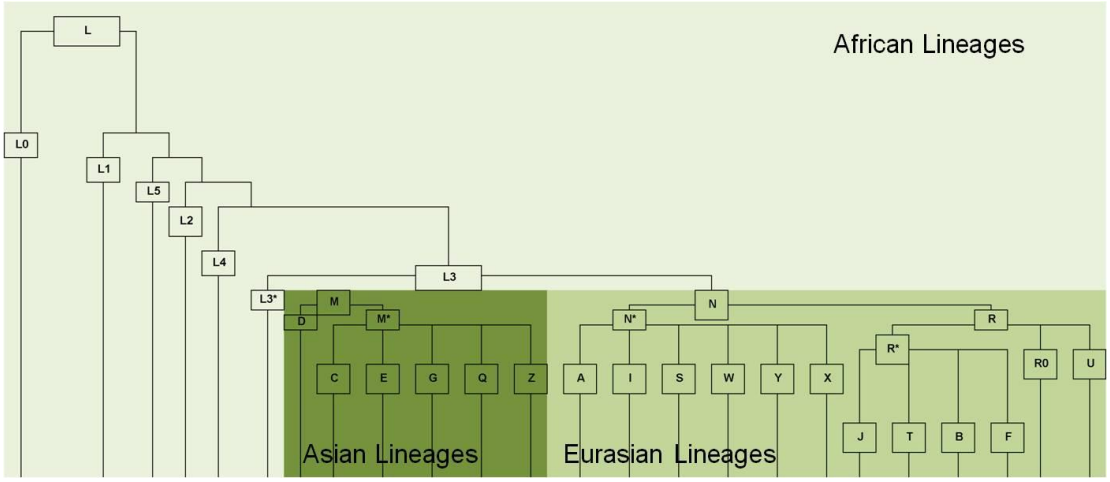

Figure 10 Phylogeny and geographic distribution of the worldwide human mtDNA diversity.

Haplogroup M is frequent in Siberia, Northern China, Japan, and South Asia and nearly absent from Southwest Asia. Haplogroup M is also found in Eastern Africa and the Near East but only as haplogroup M1, most likely the result of a back-migration into Eastern Africa. Besides M1, other haplogroups with a Eurasian origin are also present in Eastern Africa (e.g. haplogroups U6 or J and T) and also represent back-to-Africa migrations.[18; 100]

Haplogroup N almost immediately diverged further to give rise to haplogroup R (Figure 10). Haplogroup R is extremely widespread and encompasses almost all the European mtDNA lineages. It is formed by haplogroups R0 (previously known as pre-HV and including the haplogroups HV, R0a, H and V), R (including the

haplogroups, J and T) and U (including several haplogroups U and haplogroup K). Inside R, the JT branch, comprise the subclades J and T and arose ~58 ka ago, most likely in the Near East, between the time of the first settlement by modern humans and the LGM. Haplogroup J arose ~40 ka years ago and comprises almost 9% of the mtDNA pool in Europe and ~13% in the Near East. This haplogroup is divided into two main branches: the most frequent subhaplogroup, J1, defined by polymorphisms at positions 462 and 3010 and dating to ~33 ka ago; and the subhaplogroup J2 with an age estimation of 37 ka ago and defined by polymorphisms at positions 150, 152, 7476 and 15257. Both branches are quite diverse and widespread, being found mainly across Europe, the Near East, North Africa and the Arabian Peninsula.[101; 102] The haplogroup T is younger than its sister clade, having arisen ~29 ka ago, probably also in the Near East. This haplogroup is most frequent in Europe (~10%) and the Near East (~8%), but it is also found in North Africa (~4%) and the Arabian Peninsula (~6%).[101; 102] Like haplogroup J it also falls into two main subclades: T1, defined by a transversion 12633A and a transition at 16163, and T2, characterised by polymorphisms at positions 11812, 14233 and 16296.

Aside from haplogroup R, the deepest branches in West Eurasians are the non-R members of haplogroup N, which as a group can in shorthand be referred to as N(xR). These comprise the rare haplogroups N1 (including I), N2 (including W) and X. Other N haplogroups include N5 in South Asia, N9 (including Y) and A in East Asia;[103; 104] and O and S in Australasia.

The western Eurasian haplogroups N1 (including I), N2 (including W), and X are all very rare, with a relict distribution across Eurasia,[105] and have hitherto been relatively little studied. Inside haplogroup N1, N1a is extremely rare but found across Eastern Europe, Western Asia, North Africa and the Arabian Peninsula;[106; 107] and has been found at high frequency and high diversity in Neolithic skeletons from the Central European Linearbandkeramik culture (Germany, Austria and Hungary). It is characterized by the motif 16147G, 16172, 16248 in the control region.[107] The N1a frequencies in Arabia are higher than in surrounding areas, and the highest diversity and ancient nodes of this haplogroup are in the Arabian Peninsula[102]. Palanichamy et al,[107] studied N1a types from Eurasian and African populations and suggested that, although the origin for this haplogroup is Near Eastern, its presence in Europe was not the result of a Neolithic single migration event, and instead different sources contributed to the sequences such as Eastern Europe, Near East via southern Europe, and from a local central European source.[107]

Haplogroup N1b, with a transition at position 16145 and a transversion at 16176G, is also quite rare. It is very rare in Europe but appears at low frequencies in the Levant (~3%), Egypt and Arabia.[108; 109] Despite being a rare haplogroup, a specific subhaplogroup, N1b2, is present in the Ashkenazi Jewish population,[110] attaining a frequency of 9.6% in the Ashkenazim and 5.6% in Bulgarian Jews. A Near Eastern origin, and founder effect in Ashkenazi ancestors ~300-1500 years ago, has been attributed to this subclade.[111] Finally, haplogroup I, with transitions relative to the CRS at positions 16129 and 16223, is found at low frequencies in Western Eurasian (2.3%) and Central Asian populations, and is also found in Southern Siberia (at 0.5%–1.2%), Egypt and Arabian populations.[97; 106] From haplogroup N2, the haplogroup W, is characterized by the mutations at positions 16223 and 16292, is rare but distributed across Eurasia, attaining the highest frequencies in southern and northern Europe (1.8%).[97]

Finally, haplogroup X has a worldwide distribution, and is characterized by four transitions in the control region relative to the CRS at positions 153, 16189, 16223 and 16278.[56; 106] Haplogroup X has been divided into two primary branches, with the older X1 being observed mainly in Northeast Africa and the Near East, whilst the younger and more widespread X2 is observed in Europe, Central Asia, Siberia (very rarely) and in some northern groups of Native Americans.[112] The population expansion of X2 in Eurasia probably occurred around or after the LGM, but the lack of close ancestors in the Old World for the typical Native American X2a haplogroup renders unclear how and when this clade arrived in the Americas.[112] Recently, Shlush et al.[113] suggested that haplogroup X displays a relict distribution across Syria, Lebanon, Israel and Jordan, where it reaches both moderately high frequencies (~13%) and high diversities, pointing to an origin in the Near East, which would therefore represent a contemporary refugium or reservoir of ancient diversity.

The mitochondrial DNA has been the marker system most used to estimate the age of lineages and their dispersal times, as it allows for evolutionary reconstruction without the complexity imposed by recombination.[114-116] Evaluating the diversity and relating it to time requires many assumptions to be made and therefore a good estimation of the mtDNA mutation rate is extremely important.[2]

The first estimate of the mtDNA mutation rate was obtained for the study of the prehistoric migrations involved in the settlement of the American continent,[117] using the variation accumulated in the Eskimo and Na-Dene populations. The authors estimated a substitution rate of 1.80×10^{-7} transitions per nucleotide per year when considering only the HVS-I region between positions 16090-16365 in the rCRS.[117]

The advent of complete mtDNA sequences, allowed Mishmar et al (2003) to estimate a substitution rate of 1.26 x 10^{-8} base substitution per nucleotide per year[50] for the mtDNA protein-coding region (positions 577-16023 in rCRS), by calibrating with the human-chimpanzee split and assuming a linear relation between the accumulation of substitutions and time, independent of any selection. A major limitation of this mutation rate was the exclusion of the control region, which represents about a third of mtDNA variation and therefore can be very useful to increase the precision of the time estimation.[116]

Recent analyses of mtDNA sequences from a variety of taxa put in question the molecular clock hypothesis, arguing that the evolution of most of the mitochondrial genes has been nonneutral, most likely as a result of purifying selection.[91; 92; 114-116; 118; 119] In several of these studies the authors observed differences in the ratio of nonsynonymous to synonymous nucleotide changes along the mtDNA tree over time, indicating a possible mildly deleterious model of evolution.[114; 115] The younger branches of the mtDNA phylogeny trees show a higher proportion of nonsynonymous mutations in protein-coding genes and substitutions in RNA genes than the ancient branches.[116; 119] This evidence contradicts the assumption that the mutation and substitution rates should be equal under neutral evolution, indicating that purifying selection can act gradually over time on weakly deleterious characters which persist for some finite time in the population.[116]

The previous mtDNA substitution rate estimations ignored the effect of selection on the fixation rate of amino acid replacement mutations.[92] So, Kivsild et al.[119] proposed a new calibration for the mtDNA substitution rate based only on synonymous substitutions, which are not under the effect of selection, but neutral (3.5×10^{-8} substitutions per nucleotide per year).[119] This synonymous substitution rate reduces considerably the coalescence times when compared with estimates obtained by using the coding-region substitution rate.[50; 116; 119] Kivisild et al.[119] also tried to deal with criticisms to the use of the ancient human-chimpanzee split for the calibration, by using information from nuclear inserted mitochondrial fragments (NUMTs) and the Neanderthal mtDNA sequence.[119] Other authors addressed the calibration issue by employing internal calibration points based on archaeological episodes within the time frame of modern human evolution and by using Bayesian estimations with a relaxed molecular clock.[120] These calibrations allowed them to obtain posterior distributions of substitution rates and divergence times. However, this internal calibration strategy can be controversial once it can be not so straightforward to link archaeological episodes to mtDNA lineages. This is aggravated by the fact that usually the internal calibration points match recent time

depths, for which a higher impact of purifying selection can greatly bias estimations, rendering the substitution rates faster and age estimates younger than when using an outgroup.[116]

In an attempt to estimate a well-calibrated time-dependent mtDNA mutation rate and re-establish the credibility of mtDNA in evolution studies, Soares et al.[116] took into consideration the effect of purifying selection on the mtDNA coding region and proposed an improved molecular clock for dating the whole human mtDNA genome, incorporating both coding and control regions. Based on a reliable worldwide phylogeny of >2000 complete mtDNA genomes, calibrating the mtDNA molecular clock without any prior assumptions on intraspecific calibration points and against recent evidence for the time of the *Homo-Pan* split, they obtained an estimate of 1.665×10^{-8} ($\pm1.479\times10^{-9}$) substitutions per nucleotide per year for the complete-genome (one mutation every 3624 years).[116] They also estimated a synonymous mutation rate for comparison purposes, of one substitution every 7884 years. As expected, Soares et al.[116] confirmed the purifying selection on the mtDNA coding region and therefore the coalescent times in the younger branches were overestimated to some extent by methods that have assumed a linear coding-region clock, and they proposed a correction curve to allow for this. Nevertheless, the authors also claimed that a completely re-evaluation of the chronology of human mtDNA evolution is not warranted, as some earlier authors claimed, as the overestimation was less than has generally been asserted.[116]

2.1.2. Y chromosome diversity

Usually in population studies based on the Y chromosome diversity, the characterisation of the binary polymorphisms is performed in order to affiliate haplotypes into haplogroups and use STR diversity within the haplogroups to date the MRCA. The initial studies performed the SNP characterisation by RFLPs, but the discovery of more binary markers by Denaturing High Performance Liquid Chromatography (dHPLC[121]) led to the development of new technical methods for a faster and higher-throughput genotyping, namely the analyses of short amplicons by a multiplex SNaPshot assay.[84]

A big effort has been made to reconstruct the Y chromosome worldwide phylogenetic tree (Figure 11), by applying the principle of maximum parsimony, and to unify the several unrelated and nonsystematic nomenclatures in use.[2; 122] The Y Chromosome Consortium (YCC) published the first highly-resolved single most parsimonious tree based on 153 binary-defined haplogroups genotyped in a global

representative set of samples.[123] The major clades of the tree fell into 18 haplogroups, indicated by capital letters from A to R. The latest 'official' phylogenetic tree of the human Y chromosome published by the YCC in 2008 incorporated 599 binary markers, forming 20 major haplogroups (named from A to T).[124]

The primary split of the Y chromosome haplogroup tree leads to haplogroups A and the rest (B to T), followed by the separation of B from the rest (C to T). Haplogroup A, the most basal in the entire Y chromosome tree, was initially defined by two mutations (M91 and P97) and was most frequent in Khoisan, Ethiopians, and Sudanese populations, while haplogroup B was defined by four mutations (M60, M181, P85, and P90) and the highest frequency is found among Pygmies.[2; 89; 124]

The remainder of the deep structure of the phylogeny (the DECF branch in Figure 11) is characterized by subclusters that coalesce at the root of the node carrying the derived states M168, P9 and M294. They are basically non-African, except haplogroup E, which together with haplogroup D, is the first split from the remaining CF clade. Haplogroup E (defined by a long branch containing 18 mutations),[124] is highly frequent in African populations (including its subhaplogroups E1b1b1, found in northern and eastern Africa, as well as in the Mediterranean basin and Europe, and E1b1a, which is widespread in sub-Saharan Africa, probably reflecting the Bantu expansion),[125] moderately frequent in the Near East and southern Europe, and occasionally occurs in Central and South Asia. Its close relative haplogroup D is observed only in Asia, its most probable place of origin.

The remainder of the non-African Y-chromosome gene pool is characterized mostly by the CF branch (including haplogroups C to T), and the few haplogroups within the CF clade with a residual presence in African populations are associated with the recent admixture of back-to-Africa migrations.[89] An example is haplogroup R1b1, found at high frequency in the central-western region of the African continent.[126] Haplogroup C (the first to split) is widely distributed in Asian, Oceanic and Australian populations, and has not yet been detected in sub-Saharan African populations, suggesting an Asian origin; the sublineage C3b is restricted to Native American populations.[89; 124] Several branches derive from the root of haplogroup F: paragroup F*, and haplogroups F1 to F4 and H are restricted to Asia, from the Indian subcontinent (F*, H) to Sri Lanka (F1 and F4) and East Asia (F3); G is observed in the Near East, the Mediterranean and the Caucasus mountains; J lineages are found at high frequencies in the Near East, North Africa, Europe, Central Asia, Pakistan, and India, with J2 (defined by M172) the most common J

subclade in Europe while J1 (M267) predominates in the Near East, North Africa, and Ethiopia; haplogroup I is widespread in Europe and is virtually absent elsewhere and together with the haplogroup R, is the major haplogroup found in Europe;[124] and haplogroup K (from K to T).

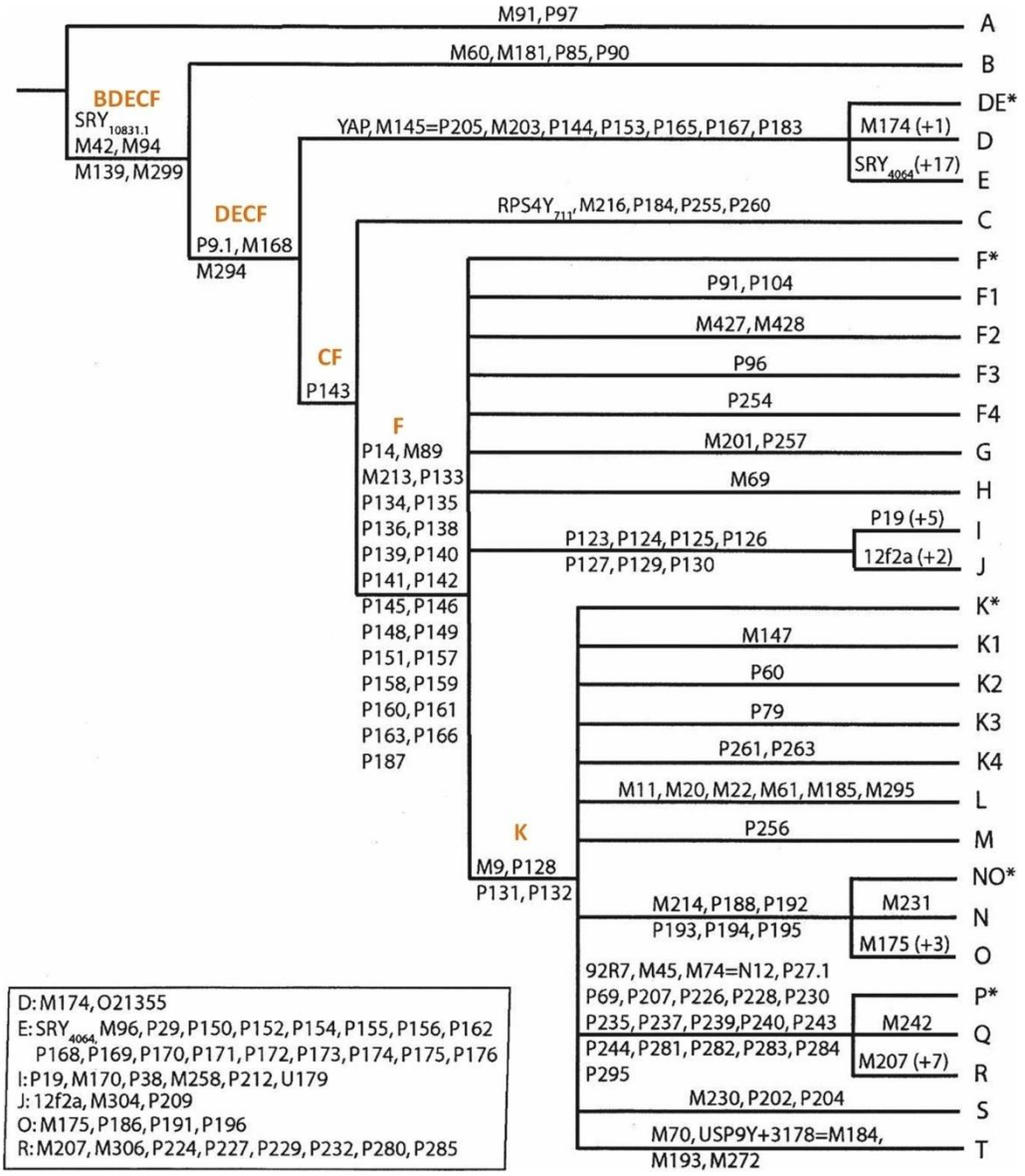

Figure 11 Schematic Y-chromosome tree (from Karafet et al.).[124]
Mutation names are given along the branches while the haplogroup affiliation is indicated in the tips of the branches. The length of each branch is not proportional to the number of mutations or the age of the mutation.

From the haplogroup K root, several lineages emerge: K* and K1'4 are found in India/Pakistan and Oceania, Indonesia and Australia; haplogroup L is mainly seen in the Indian Subcontinent, but also present in the Near East, Central Asia, North Africa and Europe; M is observed in Oceania and Eastern Indonesia; clade NO, for

which N is restricted to Northern Eurasia, while haplogroup O is the major one in East Asia, and also found at moderate or low frequencies in Central Asia and Oceania; haplogroup PQR, mostly dominated by two widely distributed lineages, Q which is found in Northeast Asia and in the American continent,[89] and R with an age estimation of 26.8 ka (19.9–34.3 ka), mainly found in Europe;[124] haplogroup S is predominant in Oceania and Indonesia; and T is present at low frequencies in the Near East, Africa and Europe.

The level of resolution of the MSY tree has been significantly increased in the last decade. Cruciani et al.[127] in an attempt to increase the robustness of the branching that separates the African-specific clade A from clade BDECF, performed MSY resequencing analysis of 205.9 kb for each representative of clade A, B and DECF. They obtained a most parsimonious tree, rooted with respect to either the orthologous chimp MSY or the paralogous human X chromosome sequence, with the deepest branching separating A1b from a monophyletic clade (A1a, A2, A3, B, C and R) defined by seven derived mutations. The next split is between chromosome A1a and the trifurcating clade formed by A2, A3 and BDECF. These results strengthened the evidence of an African origin of modern humans. They also obtained a much older time estimate of 140 ka for the inferred most recent common ancestor root (considering a germline MSY mutation rate of 1.0×10^{-9}).[127]

Until recently, dating of Y chromosome lineages has been performed by analysing the diversity of the fast-evolving STRs. As discussed before, the uncertainty in the Y-STR mutation rates is high, leading to considerable uncertainties when connecting Y chromosome genetic patterns to historical events. A recent example was the assignment of the most common European haplogroup, R1b1b2, to the Neolithic period as its dating was estimated as 6.5 ka,[128] challenging the previous idea of this haplogroup being Palaeolithic. Busby et al.[129] demonstrated that the STR-based TMRCA estimates proposed for that lineage are highly dependent on the microsatellite choice for age estimates. This, together with the lack of geographical trends in its diversity, in contradiction to expectations under the Neolithic hypothesis, led the authors to caution conclusions about the timing of its origin and dispersal, given the inadequate existing data and tools. The development of the next-generation sequencing will allow the complete sequencing of the Y chromosome at a population level, although its highly repetitive content is technologically challenging. One study[130] has already applied this approach to characterise 8.97 Mb in 36 diverse human Y-chromosomes from Africa, Europe, South Asia, East Asia, and the Americas, representing eight major haplogroups. By assuming a male mutation rate of 1×10^{-9} per base pair per year, the authors

identified/dated the following events: depth of the tree at 101-115 ka; lineages found outside Africa dated to 57-74 ka; striking Palaeolithic expansion at 41-52 ka; and the node of R1b at 4-13 ka, again supporting a Neolithic origin. Mendez et al.[131] sequenced 240 kb of the MSY of an African-American who displayed the ancestral state of all SNPs that defined the basal portion of the Y chromosome phylogenetic tree, and by using a likelihood-based method that uses mutation rates estimated by Kong et al[132] obtained a TMRCA for the Y tree of 338 ka (95% confidence interval; 237–581 ka). The old date obtained and the rarity of this lineage (named as A00 and observed at very low frequencies in central Africa) led the authors to hypothesise for archaic introgression of Y chromosomes into the modern human population. A high-resolution study of the MSY diversity in 1204 Sardinian males led to a estimate 180-200 ka for the of TMRCA of the Y chromosome human tree, more congruent with the analyses from the maternally inherited mtDNA.[133]

2.1.3. Genome wide-diversity

Recently, the ability to analyse the autosomal variation through genome-wide chip assays for thousands of markers and even whole-genome sequencing by the next-generation methods has dramatically increased our knowledge about human population structure and demographic history. The study of the presence of variation in levels of genetic similarity within a population as a consequence of factors such as geographical subdivision and finite population can give new insights about the population structure.[38; 134-136] The advantage of autosomal markers relative to the uniparental markers is that they represent a much higher proportion of the genome hence providing much more comprehensive information on ancient events of our species history[38; 39] for which both sexes have equally contributed genetic material. As gene trees are highly stochastic, the more loci analysed the better for making inferences on the population structure.[39]

Whole-genome sequencing is still expensive, but technological developments are promising to contribute to a rapid decay in prices. The 1000 Genomes Project was created to sequence the complete genome of around 2,000 individuals from worldwide populations, using a combination of low-coverage whole-genome and exome sequencing.[83] This combined strategy was able to capture up to 98% of accessible SNPs at a frequency of 1% in related populations. The first validated haplotype map for 1,092 individuals from 14 populations revealed 38 million SNPs, 1.4 million short insertions and deletions, and more than 14,000 larger deletions.

This substantial database confirmed some expectations, without ascertainment bias due to selection of markers, such as: individuals from different populations carry different profiles of rare and common variants; low-frequency variants show substantial geographic differentiation, which is further increased by the action of purifying selection; rare-variant load varies substantially across biological pathways; and each individual contains hundreds of rare non-coding variants at conserved sites, such as motif-disrupting changes in transcription-factor-binding sites. The second and third phases are now finished, totalling 2,535 samples completely sequenced in the 1000 genomes project, but information is still under analysis and not yet provided in the 1000 Genomes Ensemble browser (http://browser.1000genomes.org/index.html).[137]

Genome-wide sequencing in 78 Icelandic parent–offspring trios[132] indicated a mutation rate of 1.20 x 10^{-8} per nucleotide per generation. The authors also confirmed that this rate is highly dependent on the age of the father at conception of the child, with paternal mutations doubling every 16.5 years. This rate is quite in agreement with the one of 2.2 x 10^{-8} per nucleotide per generation obtained by Ebersberger et al.[138] when comparing ~1.9 million base pairs of the chimpanzee genome to corresponding human DNA sequences, with the average sequence difference between the two species low (1.24%). Both works also demonstrated that transitions are almost three times more frequent than transversions, while the ratio should be 1:2 transitions to transversions, if all mutations were equally probable. This deviation in the ratio of transitions to transversions was also observed in the mtDNA, where it is even stronger, as referred before.

Most of the population and clinical genetic studies have been based on high-throughput SNP arrays, including from hundreds of thousands to millions of SNPs.[38; 39] In the clinical field they have been used to map loci affecting complex disease traits, and are known as genome-wide association studies (GWAS). The development of these chips was made possible by the cataloguing performed by the HapMap and 1000 Genomes consortia, and the SNP selection from these discovery panels was performed so that SNPs have the minor allele frequency (MAF) higher than 5% and cover the common variation across the genome. These ascertainment biases can be problematic in some statistical analyses for population genetics purposes, as they assume a random subset of all variants are being surveyed.[39] Additionally, the rare variants (MAF<0.5%) have the potential to unravel much more recent gene flow patterns than the common variants, and this information is lost in the SNP selection for most arrays (not so much in the recent ones having millions of SNPs).

The amount of genetic information being gathered by the big consortia and population studies with SNPs arrays leads to pressing bioinformatics and biostatistical challenges.[39] Current statistical methods for analysis of genome-wide patterns of variation across individuals can be broadly classified into two categories:

(1) approaches based on discrete population admixture models such as STRUCTURE,[139] FRAPPE[140] or ADMIXTURE[140], which assume that each individual's genotype data is drawn from multiple clusters, each cluster being defined by genotypic frequencies; the methods estimate the proportion of an individual's alleles drawn from each cluster. The STRUCTURE program introduced the methodology, but is presently computationally ineffectual for the genome-wide SNP datasets, and largely replaced by FRAPPE and ADMIXTURE. More recent developments allow the estimation of local ancestry proportions along the chromosomes, site by site, as for example HapMix[141] and ROLLOFF[142]. These methods are mostly based on the decay of LD along the chromosomes and use information for putative ancestral populations.

(2) methods based on multidimensional statistics, such as principal components analysis (PCA). Principal component analysis does not impose a bifurcation pattern on evolution and address continuous spatial population structure into coordinates. In PCA, the raw data of allele frequencies are used to generate two- or three-dimensional graphical representations of distances between populations or individuals. The axes, denominated PCs or eigenvectors, are independent and encapsulate as much of the remaining variation as possible.

The application of these methods to the array SNP technology has already shed light on population events over the past. An influential study in this new era was performed by Novembre et al.[143] in 3,000 European individuals genotyped for over 0.5 million SNPs. The work confirmed a low average level of genetic differentiation among Europeans, but revealed a surprisingly close correspondence between genetic and geographic distances, so that a geographical map of Europe arises naturally as an efficient two-dimensional summary of genetic variation in Europeans (Figure 12). It was quite unexpected that 0.5 million SNPs would be enough to allow inference of an individual geographic origin within a few hundred kilometres. At the same time, this result reinforces the caution, when performing GWAS, of the spurious associations which can arise due to genetic structure. In fact, the clinical field has been responsible for the considerable genome-wide research performed in Europe, by far the best genotyped region in the globe.[39]

A first global preview was conducted by Li et al.[134] in 938 unrelated individuals from 51 populations of the Human Genome Diversity Panel for 650,000 SNPs. This

study confirmed that colonisation of the world by modern humans is consistent with the hypothesis of a serial founder effect with a single origin in sub-Saharan Africa. Further studies have been performed in the African continent, both South of the Sahara[144-146] and in North Africa[147], but new studies are still needed to fully characterize the high diversity of this region. A few studies have been performed in East Asia, but this region also needs a better screening; the HUGO Pan Asian SNP Consortium[148] proposed a single southern route during the initial colonisation of southeastern and eastern Asia, followed by northward expansions. The studies based on America have addressed two main issues: the diversity of the Amerindian populations and colonisation of the continent[149] and the level of admixture between Amerindian, Caucasian and African population groups.[150; 151]

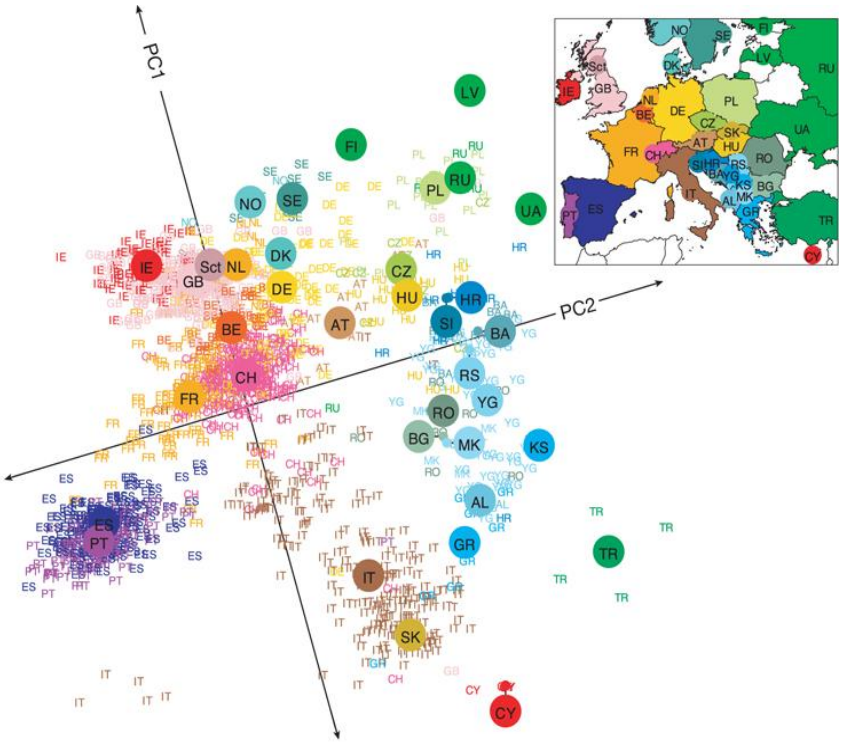

Figure 12 Population structure within Europe (from Novembre et al.).[143]
Principal components analysis with axis one (PC1) and axis two (PC2) of genetic data from 1,387 Europeans.

In conclusion, the results of recent SNP-based studies are giving new insights into the population structure, but further studies and new methods are still needed to obtain more reliable results and a finer detail of human history.[38]

2.2. Ancient DNA Analyses

Analysis of ancient DNA provides the ability to relate extinct lineages using molecular phylogenies and consequently to track changes in populations over time.[2; 152] After the death of an organism, DNA is usually rapidly degraded by the action of endonucleases, and several other factors like temperature, air/soil humidity, soil pH and microbial-mediated decay can interfere with the preservation of the DNA. The analysis of these factors can serve as a guide to estimate the recovery of ancient DNA from a sample.[153]

The first studies of ancient DNA were performed on a quagga (an extinct zebra subspecies) and an Egyptian mummy.[154; 155] But as the Egyptian sample was subsequently admitted to be contaminated and the tissues were too small to perform a second analysis, its authentication remains impossible.[152] In fact, one of the most important issues in ancient DNA is the contamination with ubiquitous modern DNA, both human during the handling of the samples (excavating team, archaeologists, museum curators and researchers) and also DNA of bacteria, fungi, and insects that feed on and degrade macromolecules.[2; 152] The introduction of the cloning PCR technique in ancient DNA study facilitated the identification of foreign sequences introduced by contamination.[152] More recently, the application of next-generation sequencing to the study of ancient DNA, in which the small fragments can be screened several times (depending on the coverage used) allowing quantification of the original DNA sample and possible contaminants, is promisingly contributing to a new boom in ancient DNA research.[156]

The mtDNA control region was under the focus of the first ancient DNA studies as it is more probable to recover this multicopy genome from old samples than the nuclear DNA retrieving results as old as 100 ka,[157-160] but nuclear genomic data are also becoming available for this species.[161-163] In general, these studies showed that Neanderthals and modern humans shared a most recent common ancestor approximately 550-700 ka, and although there is no evidence for gene flow from the mtDNA, the nuclear DNA suggests that around 2-4% gene flow seems to have occurred from Neanderthals into the ancestors of non-Africans, before the divergence of Eurasian groups from each other.[88; 162]

Ancient DNA studies in early modern human specimens are also becoming very common. The Tyrolean Iceman, a 5,300-year-old Copper Age individual from Tisenjoch Pass in the Italian part of the Ötztal Alps, was analysed for the mtDNA,[164] which allowed to affiliate it in a branch of haplogroup K1 that has not yet been identified in modern European populations. But results from the complete

genome[165] suggested a recent common ancestry between the Iceman and present-day inhabitants of the coast of the Tyrrhenian Sea, and also revealed considerable phenotype information such as brown eyes, blood group, lactose intolerance, an increased risk for coronary heart disease and infection with the pathogen for Lyme borreliosis. A 4,000-year-old human from the Saqqaq culture of Greenland was also sequenced first for the mtDNA[166] and later for the complete genome[167], evidencing a migration from Siberia into the New World some 5.5 ka, independent of that giving rise to the modern Native Americans and Inuit. Besides this individual characterisation of interesting specimens, some studies are already taking a population-based screening approach for complete mtDNA genome, such as the study of 39 haplogroup H individuals from ancient European human remains, from ~1.6-5.5 ka.[168]

Interestingly, ancient DNA has provided evidence to identify a new hominin species in Denisova Cave located in the Altai Mountains in southern Siberia.[169] The 48-30 ka fossil remains found consisted of only a distal manual phalanx and two molars, not enough to allow a morphological classification. This new species was called Denisovan, and the authors claimed, based on the analysis of the mtDNA, that this hominin shared a common mtDNA ancestor with anatomically modern humans and Neanderthals about 1.0 million years ago, indicating that it derives from a hominin migration out of Africa distinct from that of the ancestors of Neanderthals and of modern humans, although Denisovans lived close in time and space with Neanderthals and modern humans. Soon the nuclear genome of the Denisovans was characterised (at 1.9x and 30x coverage),[170; 171] and although it confirmed that Denisovans had an evolutionary history distinct from Neanderthals and modern humans, the data suggest that it contributed 4-6% of its genetic material to the genomes of present-day Melanesians.

3. The origin of modern humans

The main hypotheses for the evolution of modern humans agree that there was a *Homo erectus* spread from Africa around 2 million years ago.[90] However, these two hypotheses have different models for explaining the spread of modern human populations. The multiregional model postulates continuity over time between contemporaneous populations and the emergence of modern humans simultaneously in different worldwide regions. The out-of-Africa hypothesis argues that a single group of modern humans arose in Africa and spread throughout the world, replacing the local archaic human populations.[172] The genome data as a whole has led to abandoning the multiregional model, but for historical contextualisation I will describe it here in detail.

3.1. Multiregional model

The multiregional model, also known as regional continuity, hypothesizes that the transformation to anatomically modern humans occurred simultaneously in Africa, Europe and Asia and relies on substantial gene flow between local archaic humans and the emerging modern humans to maintain the different populations as a single species.[90; 153; 172] Wolpoff and colleagues[173] tried to explain the opposing evidence of some traits being shared between populations (such as the increased cranial capacity and the reduction of the face), while at the same time regional differences were retained (like the high prevalence of shovel-shaped incisors in Asian populations). The explanation was a balance between gene flow, genetic drift and selection.[174] This model was supported by fossil evidence of anatomical and behavioural continuity between archaic and modern humans.[90]

However the multiregional scenario is largely inconsistent with the genetic data. Gene flow at a global scale, as defined by this model, requires a continued high effective population size of the human lineage, which was definitely not the case.[7] Besides that, the genetic evidence showed a low genetic diversity in human populations, strongly reflecting a recent origin in a localized population, shown to be placed in Africa.[2]

3.2. Out-of-Africa model

The African replacement model is one form of what is referred as the out-of-Africa model.[174] According to it, anatomically modern humans originated from a single ancestral population in Africa, approximately 200 ka, followed by a later spread throughout the Old World, replacing, without admixture, archaic human species outside of Africa.[90; 153; 175]

Fossil evidence strongly indicates that the earliest transitional or anatomical modern human (AMH) fossils have been found in Africa. Several finds support an eastern African origin: the Omo 1 cranium from south-western Ethiopia and dating to ~190–200 ka;[11] crania from Herto (in northern Ethiopia), dating to ~154–160 ka;[9; 10] and remains from Sudan and Tanzania.[176] Other early modern *Homo sapiens* remains in Africa include the Jebel Irhoud fossils in Morocco, also dating to ~160 ka[177], albeit with wide confidence intervals and some disagreement about their status.[178] In southern Africa, the fossil from the Klasies River caves dates to ~65–105 ka[179], although again its status as AMH has been contested;[180] and there are more archaic remains (Florisbad, South Africa) dating to 190–330 ka[179], indicating that a southern origin cannot be yet discarded.

Human genetic diversity declines with distance from Africa, consistent with a longer period of evolution in Africa and therefore clearly supporting an out of Africa replacement model.[24] This decrease of diversity with distance from Africa is observed in the nonrecombining markers and also in the high-throughput SNP arrays studies.[181]

As already mentioned, the mtDNA evidence for an African mitochondrial Eve was the first genetic support for the out of Africa model. For the maternal genetic pool, the diversity among Africans is twice that among non-Africans and the 'star-like' phylogeny of the non-African sequences can be associated with the colonisation of Eurasia from Africa.[90] The L3 haplogroup originated in Eastern Africa around 70 ka and expanded throughout Africa, giving rise to the two major haplogroups (M and N) that cover the mtDNA pool of all non-Africans (excluding the descendants of migrations from Africa within the past few thousand years). Thus the emergence of L3 provides an upper bound for the out-of-Africa migration, refuting the possibility of a successful exit during the interglacial period MIS 5 at 130–75 ka, prior to the Toba super eruption in Sumatra at 73.5 ka. A complete mtDNA characterisation of several L3 lineages performed by Soares et al.[99] established an age for L3 between 59–70 ka, and therefore the out-of-Africa migration took place after that time.

3.2.1. Out-of-Africa routes

Since the agreement that the out-of-Africa model was the most probable model for the prehistoric human dispersal, research focused on which route did the ancestors followed. Two major routes were postulated: a northern route from the Nile Valley through the Levant corridor or a southern route from the Bab al Mandab Strait and crossing the south of the Arabian Peninsula.

3.2.1.1. Northern route

The traditional out-of-Africa model estimated the primary split between Africans and non-African populations to have occurred ~45 ka. A group of African populations crossed to the Levantine Corridor via the Nile Valley and Sinai Peninsula (Figure 13).[182] This northern route would be limited to narrow corridors and confined habitat refugia, which would be difficult to cross in the dry periods, but possible in the wetter climatic conditions of the interglacial periods.[183]

This northern route was supported by archaeological findings across the Nile Basin that resemble assemblages from the Levant, namely the Nubian Levallois reduction technology.[183] Another source of evidence comes from migrations of the mammalian fauna, with interglacial Levantine faunas coming to the African savannah and African faunas moving northwards from North Africa, both crossing the Sinai Peninsula.[184] If the mammals could follow this route, hominins could also cross it, and the evidence of this is documented by the skeletal remains found at Skhul and Qafzeh caves in the Levant during the Eemian interglacial (~90-120 ka).[185] It is known that the Skhul and Qafzeh specimens represent an unsuccessful attempt of modern humans to colonise regions outside Africa, as the upper levels of the site were occupied by Neanderthals.[186]

Figure 13 Two possible dispersal routes of anatomically and genetically modern human populations according to the out-of-Africa model (from Forster et al.)[187].
The northern route assumes a dispersal via the Nile Valley and Sinai Peninsula while the southern route assumes a route via through the Bab al Mandeb to Arabia and Australia.

There is no mtDNA genetic evidence supporting this route for the out of Africa migration, but instead the results seem to indicate that this route was used for a back-to-Africa migration of haplogroups U6 and M1.[94; 188] Macaulay et al.[94] described U6 as having evolved from a common ancestor of U in the Near East, approximately 50 ka, together with its sister clade U5, but while U5 spread along the northern Mediterranean coast with the European Early Upper Palaeolithic, U6 dispersed along the southern coast, as far as Cyrenaïca, alongside the Dabban industry, ~40-50 ka. Olivieri et al.[188] confirmed this earlier interpretation by performing complete mtDNA sequencing, and ascertained that the same migration may also have introduced haplogroup M1 to North Africa via the Levant, possibly during the Greenland Interstadial 12, from ~44-48 ka. This arrival to North Africa overlaps with the settlement of Europe, and most probably result from the same change in climate conditions. Although there has been some subsequently criticism of these conclusions claiming a younger chronology during and since the Neolithic and that more recent influx from Asia, possibly since the Last Glacial Maximum 20 ka, may better explain some of the major genetic and linguistic patterns in North Africa and adjacent areas.[189; 190]

3.2.1.2. Southern route

In 2005, two papers published in *Science* studied two putative Asian relict aboriginal populations. Macaulay et al.[191] studied the Orang Asli from Malaysia, while Thangaraj et al.[192] studied the Andaman Islanders.[191; 192] Both groups obtained an older age estimation of arrival of the relict populations when compared with their Asian neighbours[187]: the reconstructed mtDNA phylogeny obtained for the Orang Asli population provided a time depth of 44 to 63 ka;[191] M31 and M32 in the Onge and the Great Andamanese had age estimations surprisingly old, around 65 ka.[192] These two studies pointed to a single and rapid southern coastal route out of Africa, crossing the Red Sea via the Bab al Mandab strait, and rapidly dispersing into Arabia, Southern Asia and Australasia (Figure 13). In opposition to the recent time for the migration advocated by the northern route hypothesis, in this southern route model, the small subset of East African populations must have crossed the Red Sea sometime before 50 ka, with a rapid subsequently dispersal into New Guinea and Australia.[182]

The genetic-based hypothesis of a single older coastal migration through the Horn of Africa led to an intense archaeological survey in a region that was previously almost overlooked, the Arabian Peninsula,[193] as well as genetic and climate studies. Archaeological evidence for the presence of modern humans was described by about 125 ka in Faya Jebel, a mountain in the Gulf of Oman.[12] The site contains small hand axes, foliates, foliate preforms, end scrapers, sidescrapers, and denticulates from a possible Palaeolithic period.[12] No human remains were found, which undoubtedly could prove that AMH were the producers of these tools. If indeed, AMH were in South Arabia before 100 ka, these were also most probably the descendants of brief or unsuccessful human migrations in this region, in a similar way to the ones detected in the Levant. No artefacts or human remains contemporaneous of the 60 ka out-of-Africa migration have been found yet in South Arabia, but AMH populations were present at the Niah cave in Sarawak, Malaysia by at least 41 ka and between 45-40 ka at Lake Mungo in New South Wales.[182]

The Red Sea was originated as a terrestrial depression from a combination of rifting and volcanism and extends currently for 2,000 km in a north-south direction, spanning only 18 km wide but 137 m depth in the Strait of Bab al Mandab. Even nowadays, the sea level can be highly affected by the alteration of water masses on the continental shelf due to the narrow and shallow characteristics of the Strait of Bab al Mandab.[194] In the past, the abrupt climate changes occurring between 70 and 25 ka were responsible for sea-level changes at up to 35 m, as demonstrated

by the analysis of oxygen-isotope values of the calcite tests of foraminifera from Red Sea sediment cores.[194] During the LGM, the lowest level of depth was attained, of only 15 m, that made possible the sea crossing by swimming or simple rafting.[195; 196] Besides the low sea level, the passageway could have been facilitated by favourable winds, currents and good quantities of bamboo for the floatable material. The climatic conditions may have created attractive conditions in both coastal sides of the Red Sea channel, such as water supplies, food resources, animal resources and raw materials for stone-tool manufacture, that may have triggered the need for early humans to spread out of Africa.[196]

Through the analyses of stone tool technology, Mellars[182] also postulated that technological or cognitive capacities were also a cause for the postglacial populations expansions. However the expansion of mtDNA haplogroup L3 studied by Soares et al.[99] suggests that the first successful dispersal out of Africa was most likely caused by palaeoenvironmental forces than by symbolically mediated behaviour.[99]

4. The Arabian Peninsula

The Arabian Peninsula measures 2.3 million km^2 and is bounded to the west by the Red Sea and the Gulf of Aqaba, to the south by the Gulf of Aden and the Arabian Sea and to the east by the Gulf of Oman and the Arabian Gulf.[197; 198] The Peninsula comprises eight separate geopolitical entities (Figure 14): the Kingdom of Saudi Arabia, the Kingdom of Jordan, the Republic of Yemen, the Sultanate of Oman, the United Arab Emirates, the State of Qatar, the Kingdom of Bahrain, and the State of Kuwait.[199; 200]

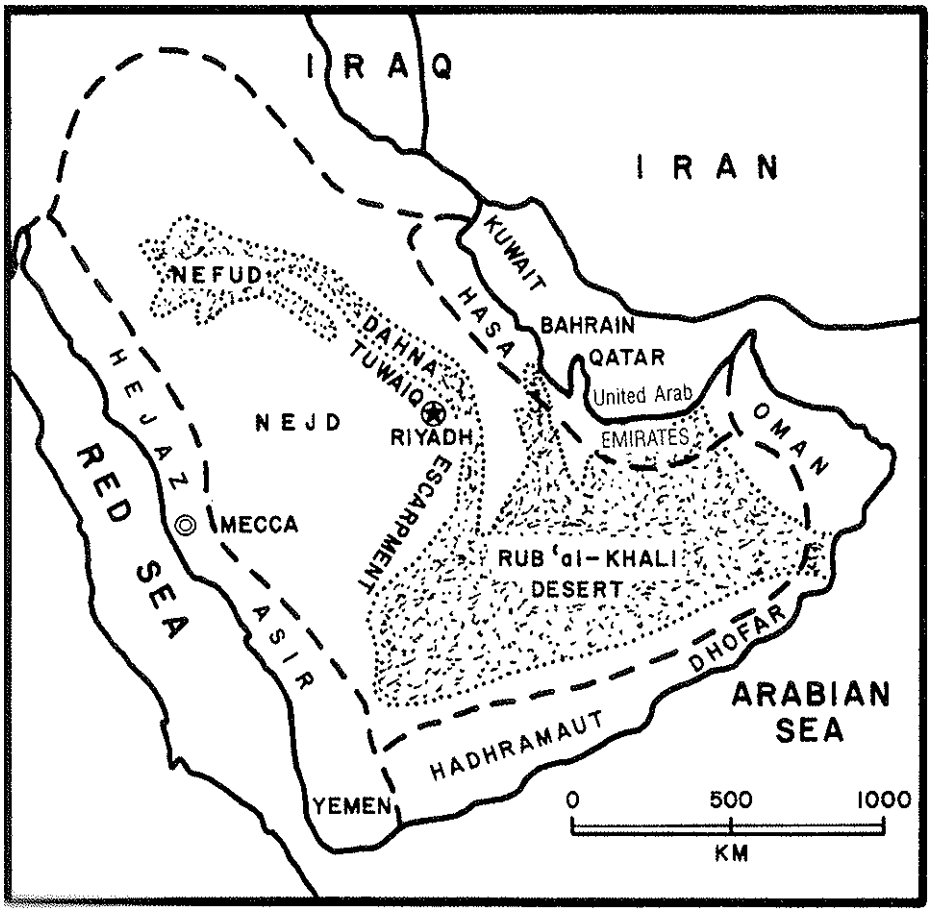

Figure 14 Arabian Peninsula (from Prehistory in Northeastern Arabia - the problem of interregional interaction).[200]

Islam has been the main religion in Arabian Peninsula, since the 630s AD, when Muhammad moved to Mecca and unified much of Arabia. During the European maritime exploration, some coastal regions of the Arabian Peninsula were forced to adopt the Christian religion, although this was soon abandoned as the European aims were primarily commercial.[201]

4.1. Geology

The Arabian landscape is the result of a combination of several factors, such as the impact of plate tectonics, sea-level changes, aridity and wetness.[202]

The Peninsula is divided in several geological provinces (Figure 15). The Arabian Shield, located in the western half, is characterized by various igneous and metamorphic rocks, which are the oldest in Arabia. Series of sedimentary rock are located to the east, becoming progressively younger towards the Arabian Gulf. Sand, coastal sediments, and extensive basalt sheets cover wide areas of these two main provinces. In opposition, Oman has a band of old ocean bed (ophiolite) in its surface rocks, resulting from a rare phenomenon worldwide of the denser oceanic crust overriding the continental crust.[203]

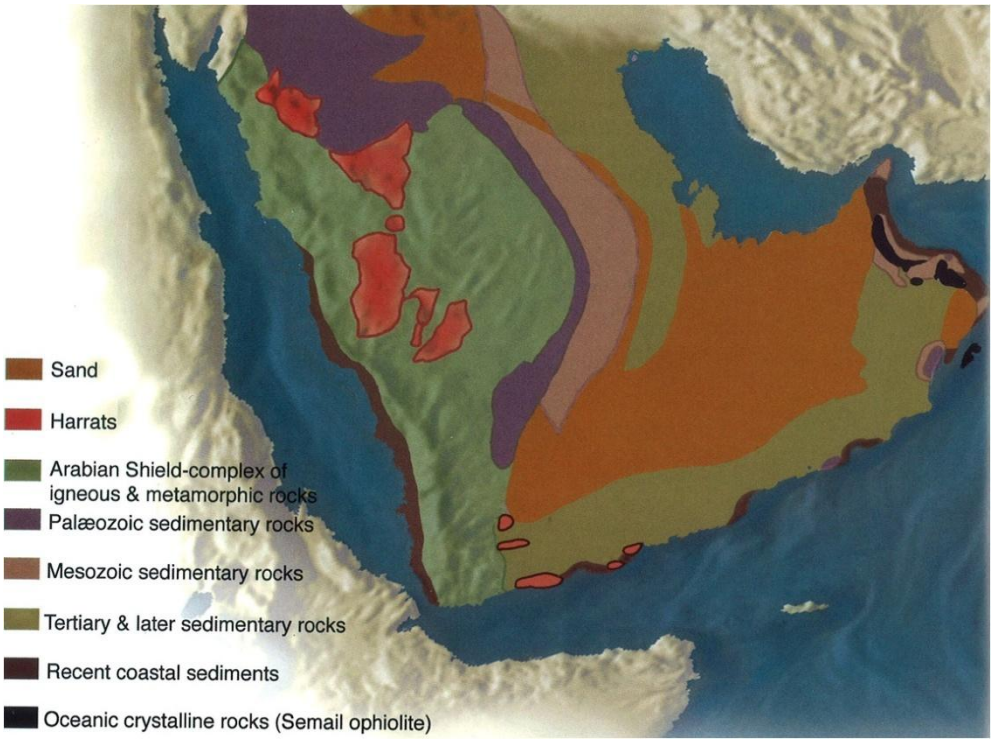

Figure 15 Surface geology of the Arabian Peninsula (from Thompson).[203]

About 60 million years ago, the Red Sea was opened due to the thinning and stretching of the crust underneath western Arabia conducting to a series of faults along a line; eventually the land along this line slumped and a chain of lakes was formed, becoming a continuous seaway during the Palaeocene epoch with the Mediterranean, followed by several episodes of widening along time and eventual closing up in the Isthmus of Suez and sinking of the Gulf of Aden and the Strait of Bab al Mandab with opening to the Indian Ocean by 5 million years ago. By this

time, there was a renewal rise of the escarpment and adjacent land, pushing the Highlands of Yemen and the Asir mountains in Saudi Arabia to its present heights, with highest elevations of over 3,000 m. The south of Yemen also contains steep escarpments of a broad plateau in the Hadhramaut region, averaging 1,370 m.[196; 199]

The last major tectonic movement occurred in the middle of the Pliocene age (3 million years ago), and the compressing forces between the Arabian and the Eurasian plates were responsible for the formation of the Zagros mountains along the north and northeast edges of the Arabian plate, and the Taurus mountains in Turkey. Also, the tectonic movements towards southwest and northeast resulted in depressing the east frame below sea level (tilting up in southwest and down in northeast), creating the Arabian Gulf and also sedimentary strata that nowadays represent the traps for the world's most prolific oilfields.[203]

Combined to these geological events, Arabian Peninsula suffered a high number of climate changes. For two million years (3-1 million years), the peninsula endured a high rainfall, and the great and powerful rivers in the Yemen mountains along the Red Sea transported a huge quantity of rock, sand and gravel towards east, to fill the Rub al Khali basin, while draining in the southeast Yemen plateaus created the Wadi Hadhramaut region.[203] As since 1 million ago, the climate in Arabia changed to aridity, although in cycles of glacial and interglacial periods, the Arabian Peninsula surface became eventually formed by extensive dune fields (Figure 16) along those riverbeds then dried, known as wadis, and covering more than one third of the Peninsula. This sand is mainly concentrated in three major sand seas (Figure 14) namely the Rub' al Khali (the Empty Quarter; 600,000 km^2), the Great Nafud (72,000 km^2) and Wahiba Sands (12,500 km^2).[199; 202]

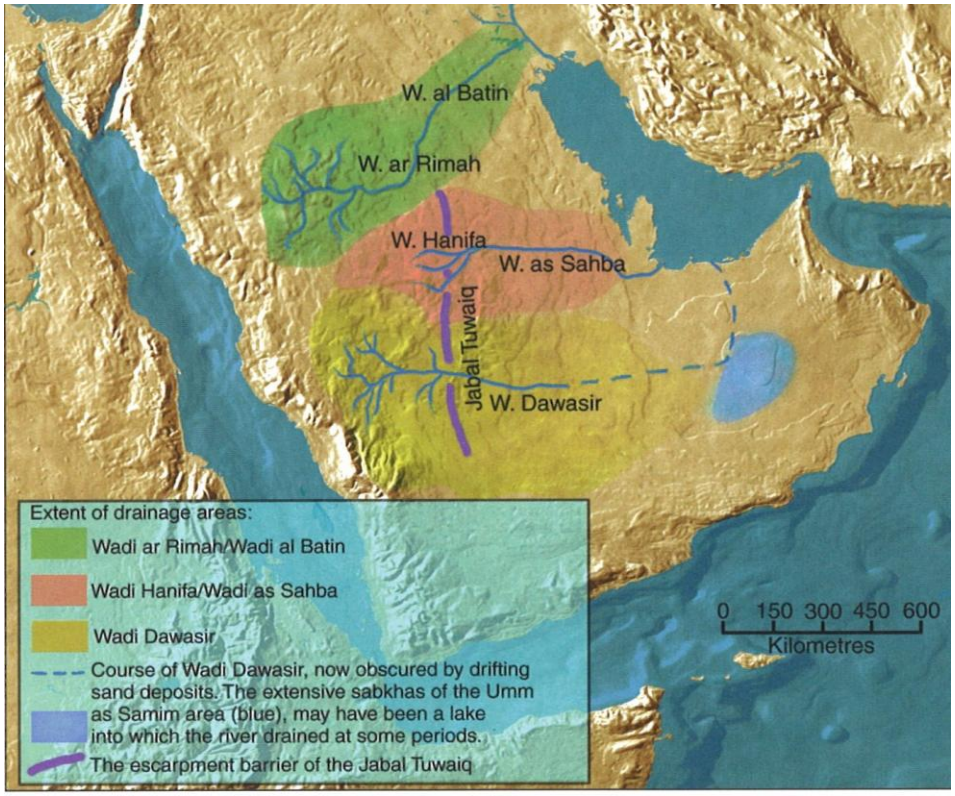

Figure 16 Wadis in the Arabian Peninsula (from Thompson).[203]

Besides these main sand seas, the Arabian Peninsula is also comprised by a few lush sub-tropical forests, deflated gravel plains, and as mentioned before, mountains and plateaus (Figure 17). But unquestionably, the main feature of the Arabian landscape is the scarceness of water resources. Arabia contains the most rainless countries on earth, with monsoon rains and westerly winds limited to the mountains and plateaus of the Republic of Yemen.[204] The main water resources dispersed in the Peninsula, are the porous strata of sedimentary rocks, or aquifers. Arabian Peninsula has nine principal aquifers and numerous smaller ones, holding a colossal quantity of water. Interestingly, radiocarbon dating of the water has shown that much of it is dated from 15-25 ka, and some older; the current arid climatic conditions of the region renders the recharge of the aquifers very limited.[203] Besides these rock aquifers, the already mentioned wadis, floored with thick layers of sand, gravel and soil accumulated over time, also accumulate water, but they are more loose and porous, having been more important in the past (even a recent past before mechanic drilling and pumping).

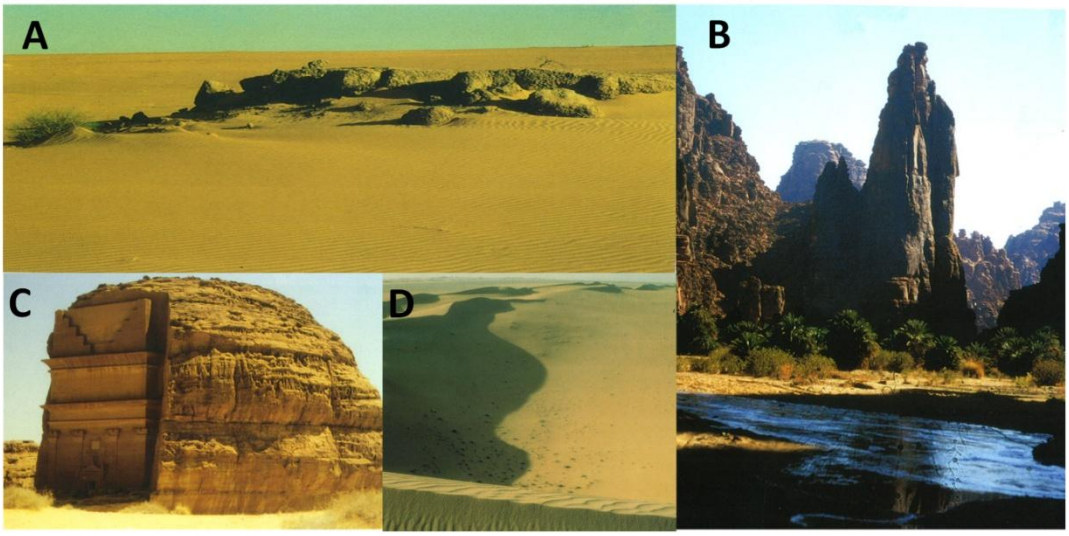

Figure 17 Examples of the landscape found in the Arabian Peninsula (adapted from Thompson).[203]
The figure shows in A dunes at Dahna, in B the spires and tower rocks in Wasi Qaraquir south of Tabuk; in C sandstone at Medain Saleh and in D sand dunes at the Tuwaiq.

4.2. Climate

At the present time, the Arabian Peninsula displays a semi-arid climate. In the north, the maximum daily temperatures range between 20ºC and 35ºC and in the south between 29ºC and 40ºC. Most of the precipitation in the Arabian Peninsula is due to the Southwest Indian Ocean Monsoon system (IOM).[202] Rainfall rarely exceeds 180 mm per year and the higher precipitation is present at the higher altitudes (e.g. the highlands of Yemen and the Asir mountains of Saudi Arabia). In contrast the Atlantic late-winter that moves over the Mediterranean sea is responsible for less winds and light precipitation in northern Arabia.[196]

The Arabian Peninsula has been exposed to several climate changes over time, with fluctuations between arid and humid phases, which can be important clue to understanding human history.[193; 202] On a global scale, glacial periods are associated to arid phases and interglacials to wetter conditions.

Considering the climate variation in Arabia in the Upper Pleistocene[202] the onset of the last interglacial period (MIS 5e), around 130 ka, was accompanied by a drastic increase in the rainfall over South Arabia, lasting 10 ka, and followed by a second peak in MIS 5a (82-74 ka). These pluvial conditions are well attested in speleothem (or cave deposit) growth and lake sediments, throughout the Peninsula.

By 75-60 ka (MIS 4), arid conditions were established, which some authors suggested was caused by the Toba eruption at ~74 ka,[205] the Sumatran supervolcano controversially argued to have led to human genetic bottlenecks.[206]

The period between 60 and 20 ka (MIS 3) has been thought of as another wet period, although some recent evidence suggests that this stage was complex and comprised a series of fluctuations between arid and wet periods (summed up in Parker).[202] Despite the variation during this stage, for the period between 35 and 20 ka the reports of lake deposits in Jordan, Wadi Muqat, Jafr, as well as thousands of small lakes spread across Yemen testify to the existence of a wet phase across the Peninsula.[202]

However, since the LGM (19-25 ka) the climate was marked by a very intense arid phase that continued until 15 ka, with a series of dune formations. The analyses of the dune formations in the Rub' al Khali in Nafud and the Wahiba sands show that signal of a major aridity phase with aeolian accumulation during the LGM.[202] This phase was followed by another wet phase between 15 and 13 ka that may coincide with the Bölling-Allerød (BA) interstadial occurring in the northern hemisphere. The onset of the Younger Dryas and early Holocene at 13.5 ka in central and northern Arabia was characterised by dune emplacement, peaking at 11.5 ka, and showing that monsoon activity was low during this period.

In the Holocene (11.5 ka till present), a wet phase was established again, between 9 and 5 ka. This wet phase began earlier in southern Arabia than in central desert regions and the Persian Gulf. By 5 ka, the present climate regime of the Arabian Peninsula was established - a new dry phase.[202] This alternation between dry and wet periods in Arabia conditioned its role as a bridge connecting Africa with Eurasia. During arid phases the bridge was discontinuous, restricting or preventing eastward migrations. During pluvial periods, range expansions of populations would be possible across the peninsula.

4.3. Population Distribution and Composition

The extreme climate changes over time, in addition to the distinctive features of the surface of Arabia, are responsible for the restriction of settled life to particular areas of the Peninsula. An area of 700,000 km^2 is covered by sand with only a few cultivated regions centred on the southern coastal tracts and few small oases across the Peninsula.[203] For these reasons, the population is mostly concentrated in the southwest part of the Arabian Peninsula (~22 million people in the southwest

Republic of Yemen according to the Central Statistical Organization (http://cso-yemen.org/index.php). Other populated places are recent urban and industrial growth, such as al-Hasa/Hofuf, Riyadh, the United Arab Emirates and Bahrain. The highest population cluster in Yemen is found at elevations greater than 1.500 specifically on the plateaus and mountains around Ibb and Ta'iz, and where the water shed by these mountains can be harnessed for irrigation.

Researchers believe that the current population distribution probably reflects the distribution in most past periods. Wet periods would lead to population expansions from the coasts to hinterland, along river valleys, with contractions to refugia during arid periods. Based on the climatic evidence, already mentioned, and the distribution of archaeological sites, which will be described in detail below, authors advance the existence of three main refugia in the Arabian Peninsula: the Red Sea coastal plain; the Dhofar Mountains and adjacent littoral zone in Yemen and Oman; and the emerged floodplain within the Persian Gulf basin (Figure 18).[199]

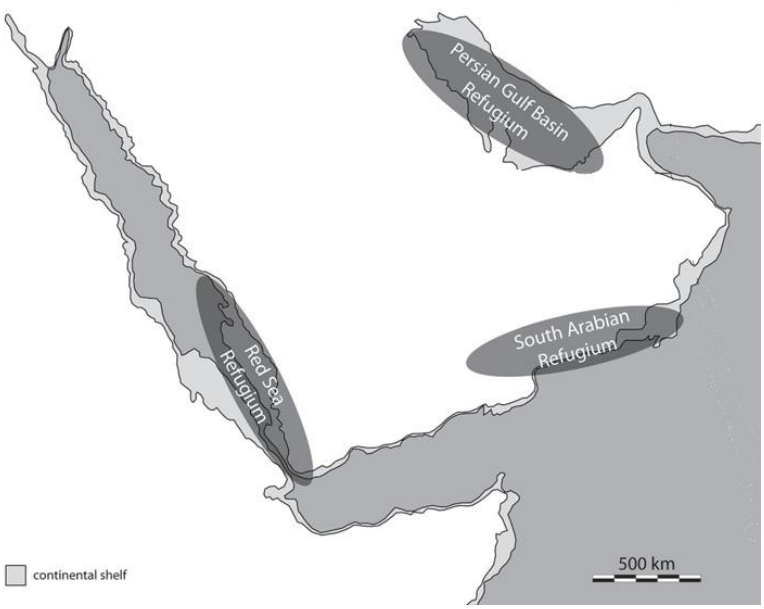

Figure 18 Three main environmental refugia in Arabian Peninsula during pluvial events (adapted from Rose and Petraglia).[199]

The refugia distribution reflect the importance of the three water-surrounded sites in the Arabian Peninsula, in detriment of the vast central desert area separating west from east coasts and the south from the north. Effective land exchanges with the 'Fertile Crescent' should have been difficult most of the time since the AMH settlement of the Arabia. Thus, it is not surprising that Arabia developed a remarkable maritime trade system with Africa and India, which must have contributed to commercial, cultural, linguistic and genetic exchanges. For

instance, the development of the incense trade, in the fourth millennium, contributed to the growth of some oasis towns at lower altitudes with irrigation systems (Figure 19), which were connected with the highly populated highland cities. These power relationships where probably responsible for the development of several early kingdoms such as the kingdoms of Qataban, Saba and Himyar.[204]

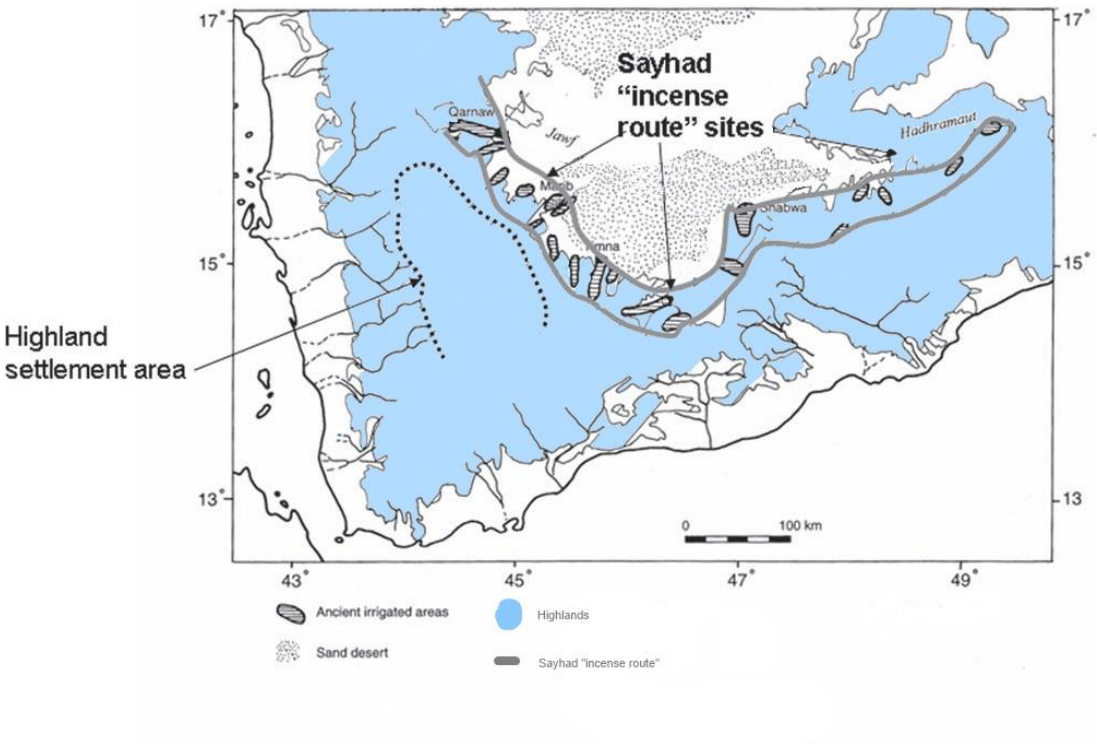

Figure 19 Sayhad incense trade settlement neighbouring the populated highland Yemen area (adapted from Wilkinson).[204]

Being at the crossroads between Africa, Europe and Asia, the Arabian population is necessarily a melting pot, enriched with successive migrations and gene flows. For prehistoric times, debates are centred around the earliest successful settlement of the region by AMH, discontinuity *versus* continuity of the Arabian population in the transition from Pleistocene to Holocene and contribution from the Levant in the Neolithic. In historic times, two population expansions/migrations dominate the region: the Islamic establishment in 630s A.D, followed by the unification of Arabia, until then made up of scattered and nomadic tribes, with the conquest of North Africa and Persia; and the movement of African people to the Near East, Arabian Peninsula and even India through the Arab slave trade established between the 7th and 19th centuries.[207] In Eastern Africa, the Arab slave trade operated from Nubia to Zanzibar (present-day Tanzania), and sea routes were established in the Red Sea and Indian Ocean (Figure 20). Ships from

Zanzibar stopped at Socotra or Aden before moving to the Persian Gulf and India, with slaves being taken as far as China. Estimates indicate that 2,400,000 African people were enslaved in the Red Sea and Indian Ocean routes.[208] The Arab slave trade was mainly focused on female slaves (a ratio of two to one for females to males), who became domestic servants, entertainers and/or concubines.[209]

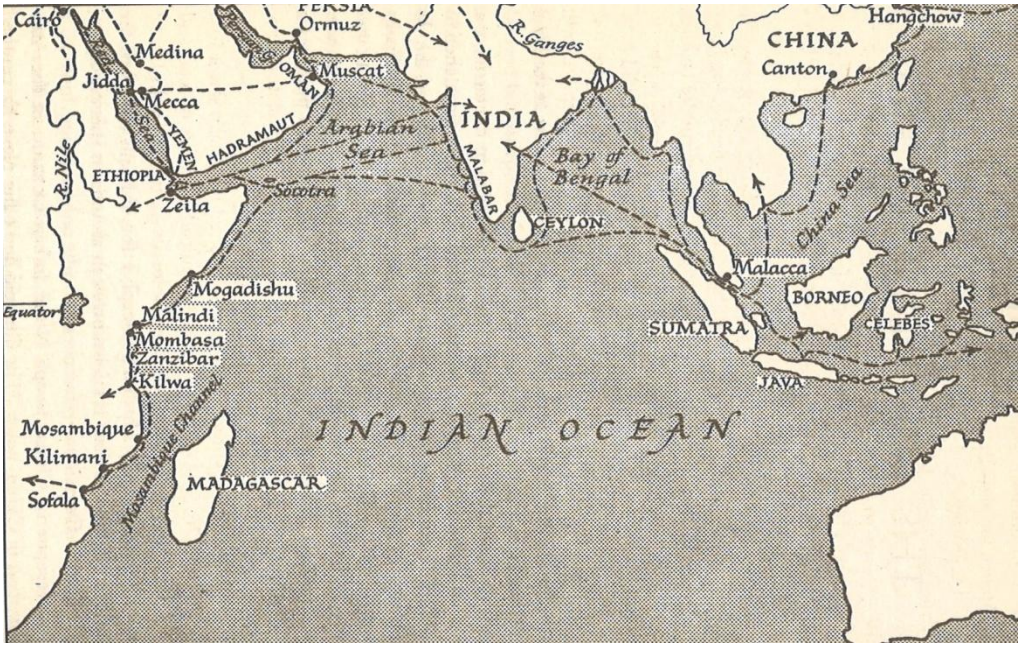

Figure 20 Slave trade routes of the Indian Ocean the fifteenth century (from Segal).[209]

4.4. Linguistics and Geography

The linguistic geography of the Arabian Peninsula derived from an Afroasiatic branch named Semitic languages, possibly originated at around 5.8 ka (95% highest probability density: 4.4–7.4 ka) from the basal lineage of the ancient Mesopotamian Akkadian language.[210] The Semitic languages probably first expanded to Arabia from the Near East, being also spoken in the Horn of Africa and in North Africa. They are of particular importance due to their association to the earliest civilisations in Mesopotamia, the Levant and the Horn of Africa, following the rise of the three religions of Judaism, Christianity and Islam.

The traditional approach to the Semitic languages, based on cultural and geographical features, established the following main groups (Figure 21):

- East Semitic: Akkadian; Eblaite
- West Semitic:
  . Northwest Semitic : Cannanite; Aramaic;

. South Semitic: Arabic; Southeast Semitic [Modern South Arabian; Ethiosemitic/Ethio-Sabean (Old South Arabian; Ethiopic)]

Hetzron, based on linguistic features, introduced the modification of grouping Arabic and Northwest Semitic together under the label of Central Semitic.[211]

Applying methods of genetic statistical analysis (Bayesian phylogenetic analysis, relaxed clock, prior sampling dates for the extinct languages to calibrate the clock) to a list of 96 words for 25 extant and extinct Semitic languages, Kitchen et al.[212] arrived at major conclusions on the origins and dispersals of the Semitic languages. The dated phylogeny of the Semitic language family estimates the origin of Semitic at 4.4-7.4 ka, with East Semitic (represented by Akkadian) as the deepest branch and appropriate root of the tree. As there were no non-Semitic languages to serve as outgroup, the authors could not ascertain when and where the ancestor of all Semitic languages (so far identified as having been present in the Near East, most probably in northeast Levant, not earlier than 7.4 ka) diverged from the other Afroasiatic languages, although phylogenetic analysis of Semitic languages and comparative analyses suggest that this must have occurred in Africa.

First divergence, occurring at 5.3 ka, in the northeast Levant led to Central Semitic spreading westwards throughout the Levant and South Semitic spreading southward from the Levant, eventually reaching southern Arabia. Early Bronze Age expansions were responsible for wider distributions of Central Semitic in the Levant and the South Arabian lineage ancestral to modern South Arabian on the southern coasts and coastal hinterlands of the Arabian Peninsula. The Arabic languages, the main extant Semitic languages, originated in northern Arabia and expanded along with Islam in the seventh century AD, in a range from Morocco to Iran, as supported by the date of divergence of 0.4-1.35 ka between Moroccan and Ogaden (spoken in the territory comprising the southeastern portion of the Somali Regional State in Ethiopia). Interestingly, the authors were able to infer that Ethiosemitic had a single, non-African origin at around 2.85 ka (and not earlier than 3.8 ka), and a rapid process of diversification and expansion occurred upon arrival in Ethiopia, contemporaneous with the development of autochthonous complex societies (Aksumite or pre-Aksumite) and the South Arabian influence in northern Ethiopia (from 2.4-2.7 ka). They suggest that this must have been a language diffusion with low gene flow from the Arabian Peninsula, given the evidence of genetic similarity between Ethiosemitic-speaking populations with Cushitic-speaking populations within Eritrea and Ethiopia.[213] Boivin et al.[214], nevertheless, associates this Arabian

migration into Ethiopia, carrying on the Ethiosemitic language, with a significant population migration from Arabia, presumably in the region of modern-day Eritrea, which transformed the economy of highland Ethiopia; these migrants also introduced the glume wheat emmer (*Triticum diococcum*) crop and chickpea, grass pea and flax, as well as sheep and goat into the Ethiopian highlands, where they have been the basis of a plowing based agricultural system throughout history.

Besides Arabic, a set of archaic and rather diverse South Semitic languages are still spoken in the coast of Hadhramaut (Yemen), in Oman and in Socotra island, indicating that they were probably more widespread before the Islamisation and Arabisation of most of the Peninsula.[214]

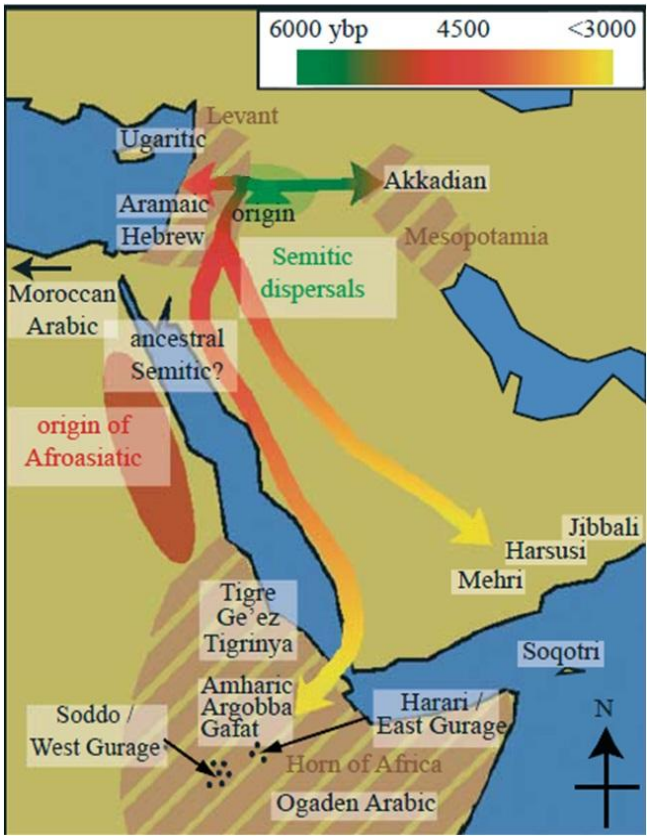

Figure 21 Distribution of Semitic languages and inferred dispersals (from Kitchen et al.).[212]
Semitic dispersals are identified by arrows coloured according to the estimated time of divergence.

4.5. Archaeology

In the Arabian Peninsula no hominin fossils have been discovered so far.[215] However archaeologists have identified stone tools sites that can give insights about the humans spread across Arabia.

4.5.1. Lower Palaeolithic (2 million - 200 ka)

Acheulean sites, containing the typical tool types of this period, described as handaxes, cleavers, picks, bifaces, discoids, polyhedrons, and spheroids (Figure 22), have been found in southwest Arabia, revealing occupation during the Lower Palaeolithic and providing evidence of African dispersals to Arabia during that period. Nearly 200 Acheulean sites recently discovered are now being studied in detail. These sites have been identified in several regions of Saudi Arabia, such as Shuwayhitiyah in the northeast, and Najran and Tathlith in the southwest.[216] Two other sites in South Arabia have been found at the elevated spots of Wadi Fatimah[217; 218] and Dawādmi,[219] near springs or stream channels. In Yemen, sites were found in the southern highlands south of Ta'izz,[220] in Al-Guza Cave in the Wadi Hadhramaut,[221] and at Humayd al'Ayn on a plateau in the vicinity of Ma'rib.[222; 223]

Most of these Palaeolithic materials were found in deflated (loose or fine-grained particles removed) or eroded surface scatters. Very few occur within a well-developed environmental or stratigraphic context, such as caves, so that it was impossible to establish a datable stratigraphy and make demographic estimates.[198; 204] But given the spatial distribution of these sites along the coasts and river valleys, it was possible to hypothesise that hominins migrated first along the southwest coast of Arabia and then travelled into the hinterland following the river systems during the wet periods, contracting back to refugia in the arid periods.[224] This significant occupation in southwest Arabia during the Paleolithic, in a location opposite the Horn of Africa, supports a migration pathway between Africa and Eurasia.[198]

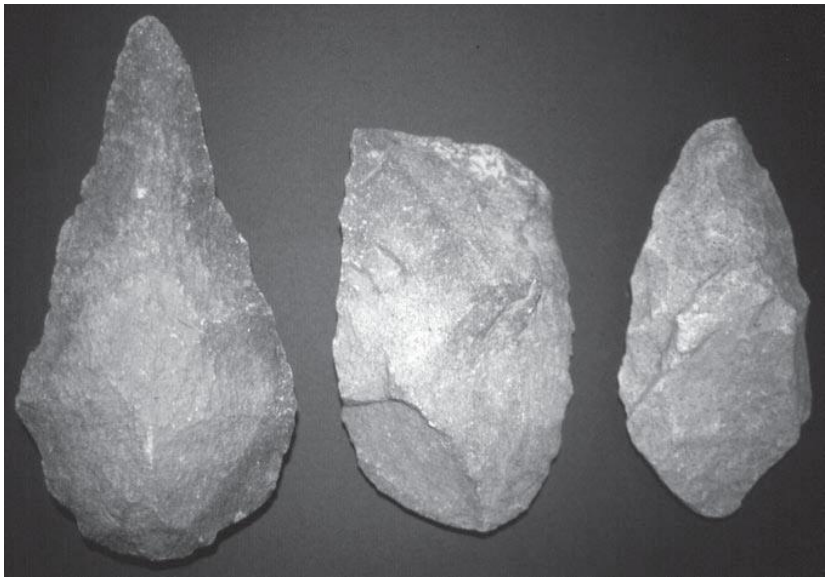

Figure 22 Acheulean assemblage present in the Dawādmi site (from Zarins et al.).[219]

4.5.2. Middle and Upper Palaeolithic (200 – 11.5 ka)

Armitage et al.[12] studied the archaeological site at Jebel Faya located in the United Arab Emirates (Emirate of Sharjah). The archaeological site is a rock-shelter and revealed archaeological levels and artifacts that include Iron and Bronze Ages, Neolithic, and Palaeolithic artifacts. Focusing on the Palaeolithic assemblage C (figure 23), small hand axes, foliates, foliate preforms, end scrapers, sidescrapers, and denticulates showed Levallois assemblages dating to 127 ± 16,123 ± 10, and 95 ±13 ka, and showing great similarities with the late Middle Stone Age (MSA) of northeast Africa. The authors conclude that technological innovation was not necessary to facilitate migration into Arabia, but that instead this was made possible by the low eustatic sea level and increased rainfall during the transition between MIS 6 and 5.[12] This evidence implies that AMH may have been present in South Asia before the Toba eruption but it is not certain that the tools can be undoubtedly associated with AMH.

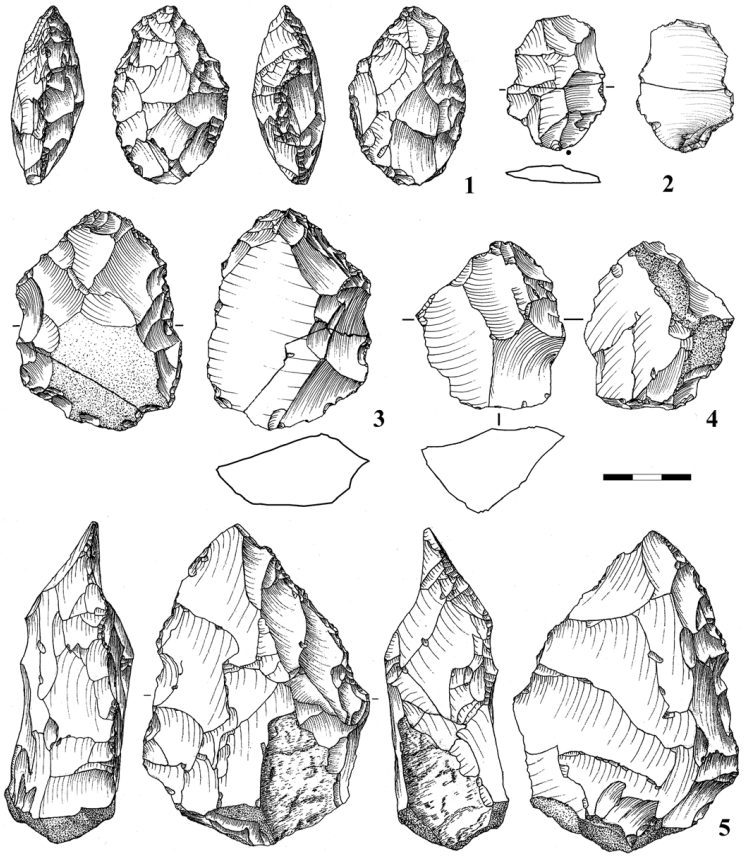

Figure 23 Lithics assemblage C tools present at the Jebel Faya site (from Armitage et al.).[12]
1. bifacial foliate; 2. Levallois flake; 3. bifacial preform; 4. radial core; 5. handax preform.

The Armitage study is just one example contributing to a change in the idea that the Levallois technology was not well-represented in Arabia. This specific reduction strategy is defined by the production of blanks showing predetermined dimensions and shapes, and this predetermination is achieved by several variations in core volume preparation.[215] Levallois assemblages were present for more than 400 ka and distributed across Europe, Levant and Africa. They were probably spread throughout Eurasia in the Middle Palaeolithic (from 300 ka) during the Mousterian period (300–30 ka). The presence of Levallois debitage in Arabia may represent diffusion between Africa, the Levant and Asia by the people that used this technology.[215; 225]

The studies on Arabic sites displaying Levallois assemblages began by performing only typo-technological comparisons[215], with no stratigraphic dating. For instance, the Levallois industries found in Wadi Jirdan, Wadi al-Gabre and Wadi Hadjar located in Hadhramaut showed, again, technical similarities with some

Mousterian industries from the Levant, and less relatedness with the African MSA or the Nubian Mousterian.[215]

More recently, dated Levallois occurrences have been evaluated in various parts of the Arabian Peninsula, namely southwestern Yemen (Shi'bat Dihya 1 in Wadi Surdud, dated to 55 ka),[226] in the already mentioned Emirate of Sharjah,[12] southern Oman (Aybut Al Auwal site in Dhofar dated at 106 ka, with tools technologically homologous to the Late Nubian Industry found in Africa, while others may represent a local industry derived from classic Nubian Levallois technology),[227; 228] northern Saudi Arabia (Nefud Desert associated with stratified deposits dated to 75 ka)[229; 230] and in southern Saudi Arabia (Mundafan in the Rub' al-Khali, associated with the wet pluvials of MIS 5: MIS 5e (ca. 125 ka), MIS 5c (ca. 100 ka) and MIS 5a (80 ka)).[231] In particular, the site in Nefud Desert, located in close proximity to a substantial relict lake, showed that Middle Palaeolithic hominins penetrated deeply into the Arabian Peninsula to inhabit landscapes vegetated by grasses and some trees. This supports the hypothesis of range expansion by Middle Paleolithic populations into Arabia during the final humid phase of MIS 5.[230] The younger Shi'bat Dihya 1 site bears tools similar to those documented from a number of nearly contemporaneous assemblages (from southern Arabia, the Levant, the Horn of Africa and North Africa) but implying also the development of local Middle Palaeolithic traditions in the Arabian Peninsula. This suggest complex settlement dynamics and possible population interactions.[226]

4.5.3. Neolithic and Seafaring

The transition from the Pleistocene to the Holocene is characterised by the initiation of the Neolithic period, defined by the development of agriculture and domestication in the Fertile Crescent, which afterwards spread to other regions of the world, included the Arabian Peninsula. A significant number of archaeological sites have been identified throughout the Arabian Peninsula dating from 9-8 ka. A variety of tools have been found in the different sites. For example, on the Red Sea coast shell middens tools have been found;[232-236] while along wadis and down to Tihama and in Hadhramaut several lithic scatters stones have been found.[236-239]

The Pleistocene/Holocene transition in Arabia is dominated by the issue of population discontinuity or continuity, associated with the origins of the Levant Pre-Pottery Neolithic B (PPNB)-related industry and its producers. One school of thought emphasises influence from abroad, mainly the Levant, while another

favours indigenous developments.[240; 241] Uerpmann et al. advanced three hypotheses for this issue in the context of eastern Arabia, favouring the second on the basis of linking the expansion with a period of environmental amelioration:[241]

1. The peopling of eastern Arabia by PPNB-related settlers was the result of widespread climatic deterioration to the north of the Arabian Peninsula around 8.2 ka.

2. The peopling of eastern Arabia by PPNB-related settlers was the result of widespread population dispersal during the Early Holocene.

3. The earliest settlement in southeastern Arabia reflects repopulation from South Arabia and/or northeastern Africa.

On the western side, Fedele,[242] when examining an Early Holocene occupation in the Yemeni Highlands (Wadi at-Tayyilah and Wadi Khamar), attested to an Early Holocene 'Pre-Neolithic' habitation throughout the eastern Yemen Plateau. The author concluded that the Pre-Neolithic and Neolithic of this region belong to a single continuum, and the features of this highland industry display hints of similarities with East Africa rather than the Fertile Crescent. Nonetheless, a recent study[243] of the Jebel Qattar 101 site, at Jubbah in the southern part of the Nefud Desert (northern Saudi Arabia; adjacent to an Early Holocene palaeolake) revealed a large collection of stone tools, similar to those recorded in Pre-Pottery Neolithic A and Pre-Pottery Neolithic B assemblages in the Fertile Crescent. As there were signs of a unique strategy to manufacture the final forms, it is possible that the producers migrated from the Levant or that they represent the acculturation of mobile communities in Arabia.

Further evidence comes from analyses of Early Holocene pastoralist societies in southern Arabia. Domesticated animals, including cattle, sheep and goat, arrived for the first time in Arabia by the late eighth millennium, roughly at the same time in eastern and western Arabia, most probably from the Levant.[244] At the same time, sheep/goat/cattle arrived in Egypt, suggesting parallel processes of dispersal.[214] McCorriston and Martin[244] support the idea of multiple human waves of expansion into Arabia introducing cattle and pastoralist strategies at different times – these expansions were, nevertheless of local hunters pioneers, and not of people culturally or temporally related to the Levantine PPNB, based on the author's interpretation of evidences in the Manayzah site (Wady Sana, Hadhramaut, Yemen).

An interesting point on the introduction of domesticates into the Arabian Peninsula is the possibility of maritime transportation alongside the initial overland arrival.[214] Livestock spread initially in the absence of plant-based agriculture in

Arabia,[241] as in Saharan and East Sudanic Africa,[245] as well as parts of savannah India.[246] The earliest field crops in Arabia appeared in the sixth millennium, consisting of cereals (wheat and barley), accompanied by pulses (pea and lentil), all originating in the Near East.[247] Curiously, eastern and western Arabia differ in the main kind of cereal[214], reflecting differential geographic interactions: free-threshing wheat, likely bread wheat (*Triticum aestivum*) on the eastern siding, common with the Indus region; and glume wheat emmer (*Triticum diococcum*) at Yemen sites, similar to ancient Egypt, Nubia, Ethiopia and Eritrea, most probably representing a maritime migration to Africa which also introduced the Ethiosemitic languages.

The evidence increasingly suggests that the Red Sea, the Persian Gulf and the Arabian Sea which surround the Peninsula witnessed some of the world's earliest seafaring and maritime exchange activities (Figure 24). The earliest evidence for maritime activity in Arabia consists of shell middens, which appear almost simultaneously at several sites around the littoral in the ninth millennium.[214] These were the 'Ichthyophagi' or 'Fish-Eaters' communities, although their sites bear sheep/goat and cattle bones, from the earliest strata[248], indicating an economic system based on a mixture of hunting, herding and shellfish collection, which was fairly stable, enduring until the later sixth millennium. The first evidence for seafaring activity appears some time after, in the eighth millennium, simultaneously in the Gulf and Red Sea, and in the form of movement of material objects across the sea: Ubaid pottery from Mesopotamia in the Gulf; obsidian artefacts (natural volcanic glass of rhyolite) from the Horn of Africa (mostly from the Eritrean/Ethiopian highlands) in the Red Sea.

The maritime traffic was intensified in mid-sixth millennium, with the appearance of the first major state-level civilisation of the Old World (the Predynastic Egyptian period), in a number of regions bordering the Peninsula. In the Red Sea, Egyptian rulers dominated long-distance trade, having introduced many innovations: in the middle of the sixth millennium they changed from reed or papyrus to wooden boats; by 5.1 ka they introduced the sail; in the fifth millennium improvements in wood sources, building techniques and sail rigging led to larger boats, increasingly well-adapted to the open sea. This led them to explore the maritime trade to the Land of Punt, trading a variety of products including myrrh, electrum, staves, exotic animals and probably slaves across the Red Sea. Punt was a land reported as the main source of incense and was a relevant pattern of maritime trade and exchange in the Red Sea region. Its location is still uncertain however it was probably situated in eastern Sudan and northern Eritrea.[249] In the Gulf, by this time, trade was established between communities in present-day Bahrain, the Oman peninsula, the

Indus valley and Gujarat, representing an intensification of trade and a moving-out into the wider Arabian Sea.[250] The maritime trade continued to increase over time, from then on.

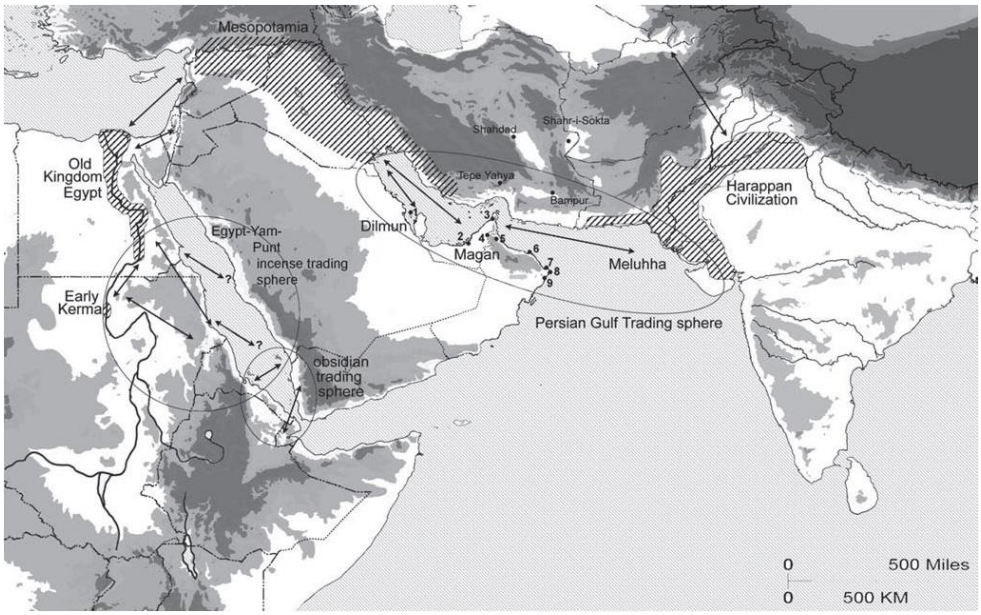

Figure 24 Arabian Peninsula seafaring and trade activities (from Boivin et al.).[214]

4.6. Genetics

As already mentioned, the Arabian Peninsula played an important role as a crossroads between Africa and Eurasia, contributing not only to the human diaspora out of Africa, but also harbouring some refugium areas during adverse climate conditions. The Arabian Peninsula has also linked China and India to communities of the Mediterranean and beyond.[251] Its strategic position made it an essential place for trade and cultural exchanges between different civilizations. The genetic information provided by the uniparental markers is contributing to an improved characterization of the Arabian Peninsula genetic diversity.[102; 197; 207; 251-256]

4.6.1. MtDNA

The Eurasian macrohaplogroup N (Figure 25), including its branches X, I, W, N1a, N1b and R (with all its subclades), is dominant in most of the Arabian Peninsula, attaining a frequency of: 83% in Saudi Arabia;[197] 66% overall in Yemen[252] (76% in Al-Mahra, eastern Yemen[257], 63% in the region close to Bab al Mandab[257] and 66% in Hadhramaut[108]); 79% in Dubai;[253] 85% in Dhofar, Oman.[257]

When considering the incidence of L lineages in the Arabian Peninsula this differs between the different regions: 10% in Saudi Arabia;[102] 16.3% overall in Yemen[252] (24% in Al-Mahra, eastern Yemen[257], 54% in the region close to Bab al Mandab[257] and 34% in Hadhramaut[108]); 22% in Dubai;[253] 12% in Dhofar, Oman.[257] These results probably reflect the close ties between Africa and southwestern Arabia. A few instances of M haplogroups (other than M1, which is most frequent in Africa) are observed in the Arabian Peninsula: 3% in Saudi Arabia;[102] 5% overall in Yemen[252] (0% in Al-Mahra, eastern Yemen[257], 5% in the region close to Bab al Mandab[257]); 9% in Dubai;[253] 2% in Dhofar, Oman.[257] Considering the incidence of M1 observed in the Arabian Peninsula, it is: 3.7% in Saudi Arabia;[102] 0.9% overall in Yemen[252] (0% in Al-Mahra, eastern Yemen[257], 3% in the region close to Bab al Mandab[257] and 4% in Hadhramaut[108]); 0.4% in Dubai;[253] 1.0% in Dhofar, Oman.[257]

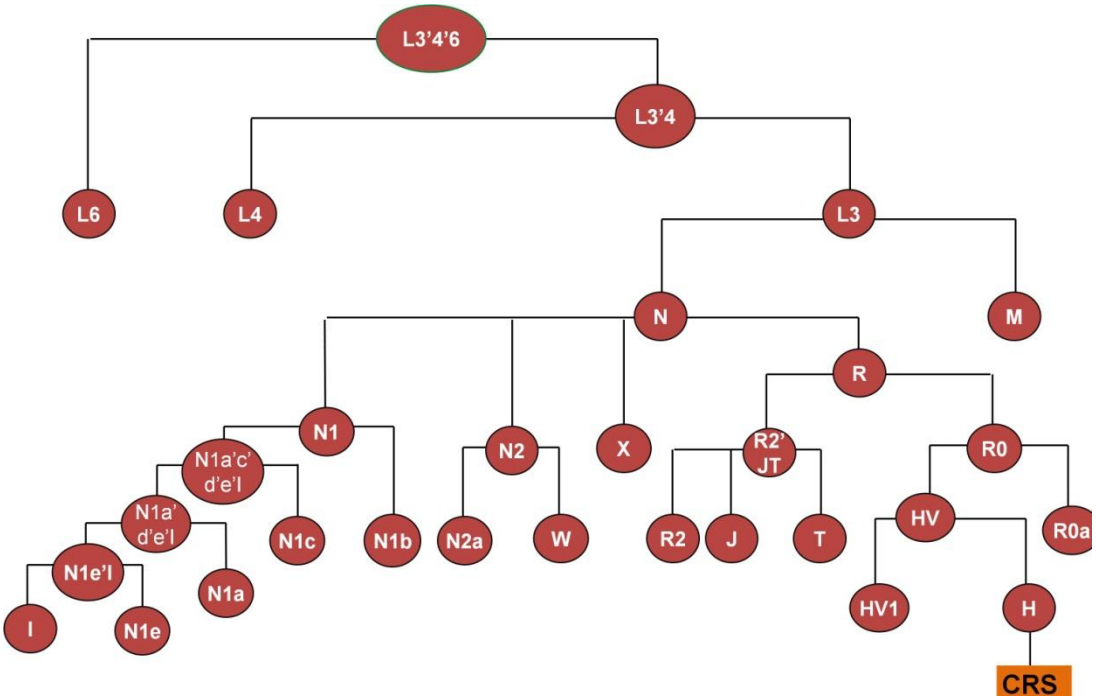

Figure 25 Schematic tree of the human mtDNA haplogroups most frequent in Arabian Peninsula.

Abu-Amero et al.[102] concluded that the eastern and western gene pools are distinguishable, due to differential geographical inputs. U, R2 and M clades, more frequent in Iran than the Near East, had a stronger impact in eastern and southern areas of the Arabian Peninsula, while the bulk of Arabian N and R lineages came from the Near East, mainly represented by R0a (more frequent in the North and Central Saudi region and in Oman) and J1b (more frequent in the peripheral West and Southeast Saudi regions) lineages. The authors, performing dating based on

HVS-1 diversity (and with the limitation that they used the TMRCA as an upper bound for a cluster radiation), concluded that there was an older Palaeolithic implantation in Saudi Arabia of the J1b lineage, constituting the primitive population, together with other N and L lineages. Later R0a subclade radiations, accompanied by other clades, penetrated from the Near East, using internal routes, and having secondary spreads in Central Arabia, diluting J1b lineages and causing its peripheral distributions.

Specifically haplogroup R0a is quite abundant in certain parts of Arabia: 18% in Saudi Arabia;[102] 13% overall in Yemen[252] (26% in Al-Mahra, eastern Yemen[257], 6% in Bab al Mandab[257], 7% in Hadhramaut[108] and 38% in Socotra island[254]); 5% in Dubai;[253] 22% in Dhofar, Oman.[257] Cerny et al.[258] performed complete mtDNA sequencing of 71 R0a samples from southern Arabia (Yemen and Socotra), North Africa (Tunisia and Morocco) and the Horn of Africa (Somalia, Ethiopia, Sudan), providing strong evidence that haplogroup R0 was firstly introduced to Arabia from the Near East, where the oldest lineages for the R clade are observed,[259] although the data did not allow them to discard the hypothesis of Arabia having been an additional centre for R0 emergence. The authors identified several founders of R0a in southern Arabia, with TMRCA estimates suggesting population continuity between the terminal Pleistocene and Holocene, and these lineages further participated in human demographic expansions occurring within South Arabia during the last 20 ka, when the climate was rapidly improving; some of these mtDNAs also spread to North and East Africa. Two periods of demographic expansion closely match two wet climatic periods: the end of the 35–20 ka wet period and the brief wet phase between 15 and 13 ka. As R0a lineages (as R0a1a1a and R0a2f1) reach the highest frequencies in the broad regions between Yemen and Oman, in Hadhramaut and Al Mahara regions, and even Socotra island, this is the first genetic evidence for the role that this region played in post-LGM demographic expansions, as an important refugium in Arabia.[199] The characterisation of another R lineage, R2, in Dhofar, west Oman and Al-Mahra, east Yemen[257], together with R0a and HV1, allowed them to confirm that the southern Arabian population underwent a large expansion some 12 ka, and that this expansion can be largely attributed to demographic input from the Near East. These results thus support the spread of a population coming from the north, but at a significantly earlier date than presently considered by archaeologists, and also that some of the mtDNA lineages found in southern Arabia have persisted in the region since the end of the Last Ice Age, favouring continuity of the population at the Pleistocene/Holocene transition. Haplogroup R0 and HV1[260] also support

extensive maritime trade in the Red Sea and the introduction and expansion of Socotran R0a clades date to between 6 and 3 ka. The detailed phylogeography of HV1 sequences shows that HV1a3 and HV1b1 passed from South Arabia to East Africa during the periods of intense commercial networks, while Egyptian HV1 haplotypes seem to be more similar to the Near Eastern ones.

Despite evidence for a southern out-of-Africa route, until now no precursors of M and N clades were found in the Arabian Peninsula.[261] This may suggest that it was a fast corridor for the first migrants, instead of a local place of expansion[102], indicating strong genetic drift at the time of the out-of-Africa migration, due to small effective size of the migrant groups. In fact, Macaulay et al.[191] calculated that a small effective size of 620 females is inferred from their estimated time of 20 ka of divergence between L3 and M/N roots; the authors estimated that a plausible stronger bottleneck would imply an effective size of between 100-400 females.

When focusing on the current African L lineages observed in Arabia, Richards et al.[108] provided genetic evidence of gene flow between sub-Saharan Africa and Arab populations occurring mainly during the Arab-conducted slave trade (evidence pointed to an input in the last 2.5 ka), through the maternal side. These sub-Saharan female lineages are almost absent in non-Arab Near Eastern populations. Kivisild et al.[207], based on mtDNA haplotype matching, also reported that the Arab slave trade was responsible for the presence of sub-Saharan lineages in Yemen, with a frequency of 23% matching southeastern African haplotypes (most derived from the Bantu influence, which did not influence Ethiopia; also limiting the time frame to the last 3-4 ka) and 9% sharing with Ethiopians. The authors also detected a high frequency (12%) of a low diversity of haplogroup L6 in Yemeni populations, which is observed marginally in Ethiopia and almost absent elsewhere in Africa. This led them to hypothesise that it could have originated from the successful out-of-Africa migration. Later on, the characterisation of other Arabian populations, including Yemen[102; 197; 252-254; 257], did not reproduce this high frequency of the mtDNA haplogroup L6.

4.6.2. Y chromosome

The Y chromosome genetic structure of the Arabian Peninsula has been estimated to approximate 10% from sub-Saharan Africa and 22% from Iran (by assuming that Iran was the entrance door for haplogroups typical from central, southern, and southeastern Asia); the remaining 68% could be considered of direct

or indirect attribution from Levant.[261] Haplogroups E and J are the most abundant ones in the Arabian Peninsula, especially J.

Haplogroup J is defined by three mutations (12f2a, M304, and P209) and found at high frequencies in the Near East, North Africa, Europe, Central Asia, Pakistan, and India.[124; 262] The main J subclades (J1-M267 and J2-M172), display opposite latitudinal gradients from the Near East to the Arabian Peninsula[261]: J1-M267 is more abundant in the southern areas, reaching a frequency around 73% in Yemen; whereas J2-M172 is more common in the Levant. Abu-Amero et al.[261] argue that the Near East was the most probable place of origin of the earliest dispersals of both subclades, but further subclades of J1-M172 are due to more recent Bronze Age expansions from Turkey and the Balkans. The age of J1 in Saudi Arabia and Yemen is significantly older than in UAE and Qatar, pointing to a terrestrial migration instead of a maritime colonization.[261] The geographic pattern, most probable origin and times of divergence for Y-chromosome haplogroup J in Arabia faithfully mirror those found for the most prevalent J and R0a mtDNA haplogroups, and seem to indicate the occurrence of migrations in the Neolithic.

Haplogroup E probably descended from the East African population that generated the out-of-Africa expansion and is defined by 18 mutations. It is found at high frequencies in Africa and at moderate frequencies in the Near East and southern Europe.[122; 181] The E3a-M2 subclade is present at high frequencies in the Arabian Peninsula (7.4% in Oman, 5.5% in UAE, 3.2% in Yemen and 2.8% in Qatar).[251] Abu-Amero et al.[261] suggest that the sub-Saharan Y chromosomes observed in Arabia result from a migration prior to the slave trade, as they did not find the strong sexual bias (low frequency of male compared to female sub-Saharan input) proposed by the authors who attribute this fact to the recent slave trade.[108] The sex biased input of sub-Saharan lineages in Arabia is concordant with historical data, which have shown that while females were normally introduced into the Arabian populations, males were often excluded from reproductive opportunities.[251]

By characterising Y-STRs, Alshamali et al.[253] have also presented evidence on local population structure in the Arabian Peninsula, demonstrating that geography played an important role in shaping the genetic structure of the region, with Dubai and Oman sharing genetic affinities with other Near Eastern populations, while Saudi Arabia and Yemen show a relatively distinctive isolated background.

Despite the mtDNA evidence for a southern out of Africa route, Y chromosome genetic studies in the Arabian Peninsula did not identify basal clades supporting the southern exit route of modern humans.[261] Instead, it seems that this area was a

place of gene flow from the surrounding African and Asian areas, but mainly since the LGM onwards.

4.6.3. Genome-wide markers

So far, only one genome-wide study has focused on the Arabian Peninsula, performing the genotyping of the Affymetrix 500k SNP array in 168 self-reported Qatari nationals sampled from Doha.[263] This study aimed to evaluate evidence for the recent founder effect on the formation of Qatar (just a small number of families from three tribes of the Arabian Peninsula, Persia, and Oman, with indications of African admixture), and the effect of inbreeding on the customary first cousin marriages (which amount to half of total marriages). The results revealed three clear clusters of genotypes, consistent with an Arabian origin, a more eastern or Persian origin, and individuals with African admixture. The levels of homozygosity in some individuals reflected substantial consanguinity, but the variance was higher than expected for such a population.

Two other screens included populations from Arabia and its neighbours from the Levant and Pakistan, but one was focused on worldwide diversity (including 45 Bedouins, 42 Druze, 46 Palestinians and many Pakistan groups)[134] and the other on analysing the population structure of Jewish communities (including 20 Saudi Arabians, 10 Yemeni, 20 Jordanians, 7 Lebanese, 3 Samaritans, 16 Syrians, 20 Iranians and 19 Turks).[136]

III. OBJECTIVES

In this work I aim to contribute genetic evidence to shed light on the main role played over time by the Arabian Peninsula as a corridor for important migrations at the crossroads between Africa, Europe and Asia, since the first successful out-of-Africa migration at ~60–70 ka. This follows the genetic and archaeological focus centred on the Arabian Peninsula, since the hypothesis of the southern route for the out-of-Africa migration highlighted this geographical region as the most probable initial staging-post. I will address three main moments of the Arabian Peninsula prehistory/history:

(1) *The first descendants of the out-of-Africa migration:* A key question related to this essential episode of modern human evolution is where L3 evolved into the two non-African haplogroups N and M. I will address this issue by performing phylogeographic analyses on 385 complete mtDNA sequences (85 characterised by me and 300 already published) belonging to the three basal clades, haplogroups N1, N2, and X, observed in western Eurasians. These are the relatively rare N(xR) haplogroups, as opposed to the widespread haplogroup R. Other N(xR) haplogroups dispersed towards the east of Eurasia, including N5 in South Asia, N9 and A in East Asia, N21 and N22 in Southeast Asia, and O and S in Australasia, and have been described as having separated from the western Eurasian pool somewhere between the Indus Valley and Southwest Asia.

Thus, the West Eurasian haplogroups N1 (including I), N2 (including W), and X, all very rare, presenting patchy distributions across Eurasia, and having hitherto been little studied, seem promising to address this issue. Strictly speaking, haplogroups N and R could have arisen anywhere in the region between Southwest Asia and Australasia. However, given L3's genesis in eastern Africa, the most parsimonious location for their origin is in the vicinity of the Arabian Peninsula, as modern humans arrived further north in the Near East only ~45–50 ka ago, most likely delayed by the desert barrier until climatic improvement ~50 ka ago.

(2) *Major population expansions in the Arabian Peninsula*: A major issue in archaeological research in Arabia is population continuity *versus* replacement since the first settlement. Was Arabia a continuously settled region or just a passage-way for the first out-of-Africa migration? Recent archaeological and palaeoclimatological investigations have revealed the importance of the terminal Pleistocene and Holocene in the population history of the Arabian Peninsula, with some indications of the arrival of material cultures from other locations. I will address this question by performing phylogeographic analyses of the mtDNA sister haplogroups J and T in

Arabia (44 new complete sequences), which together with haplogroups R0 and HV (previously analysed by our group) constitute between 14.9-44.8% of the Arabian pool.

(3) *Genetic exchanges across the Red Sea*: The route linking eastern Africa with the Arabian Peninsula, through the Bab al Mandab strait, played an important role not only in the out-of-Africa migration but also in more recent processes such as the seafaring activity across the Red Sea, during the early and middle Holocene period, and the Arab slave trade established between the 7th and 19th centuries. African lineages now extant in the Arabian Peninsula can provide insights into the movement from eastern Africa to Arabia, while Eurasian genes found in eastern Africa can inform us about movements in the opposite direction. I will investigate these movements through the phylogeographic analysis of mtDNA complete sequences (L4 and L6 newly sequenced), founder analysis in an enlarged mtDNA HVS-I database (for all L(xMN) haplogroups and Eurasian ones such as R0, HV, J, T, M1 and U6), and ADMIXTURE, ROLLOFF and PCA analyses of published genome-wide data.

IV. MATERIAL AND METHODS

5. Mitochondrial DNA

5.1. Biological samples

The samples consisted of buccal swabs or peripheral blood and belonged to Luisa Pereira (IPATIMUP) and Martin Richards (University of Leeds). Appropriate informed consent to anonymously use the data was obtained from all individuals.

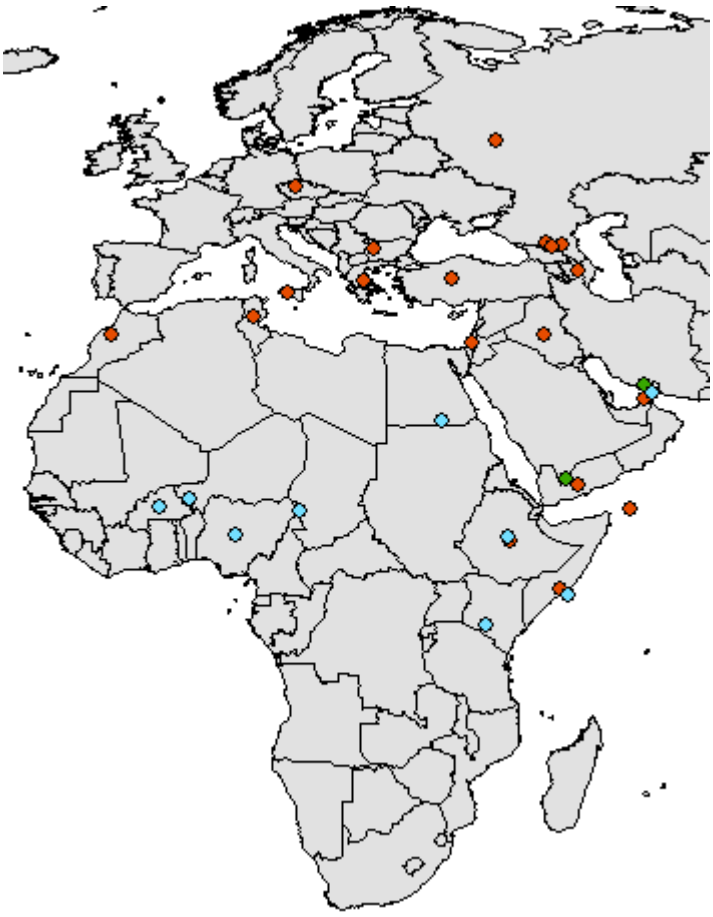

Figure 26 Geographic distribution of the populations used for complete mtDNA sequencing.
In the map the populations belonging to the N1/N2/X, J/T and L4/L6 lineages are represented in red, green and light blue dots, respectively.

The mtDNA diversity had been previously characterised in populations from eastern Africa,[99] Arabian Peninsula,[253; 257; 258] and African Sahel,[264] by sequencing the HVS-I and HVS-II, using a procedure described previously.[265] This information was used to assign samples to haplogroups, following the most recent phylogenetic evidence, reported on the PhyloTree website[266] and I checked the classification with

the Haplogrep software[267] and manually rechecked it again. Then I selected the informative haplotypes in order to perform the complete sequencing, as follows:

1) In order to test the role of the Arabian Peninsula as the suggested first stage of the “southern coastal route" for the out-of-Africa migration, I selected 85 samples for complete mtDNA sequencing, belonging to haplogroups N1, N2 and X from Arabian Peninsula, eastern African, North African, Near Eastern and European populations (indicated by red dots in Figure 26). The complete mtDNA sequences are deposited in GenBank database with accession numbers JQ245723–JQ245807 and also available in Table S1.

2) I investigated later demographic expansions in Arabia by selecting 19 Dubai and 25 Yemen samples belonging to haplogroups J and T, some of which were previously associated with the Neolithic expansion and which attain considerable frequencies in the Arabian Peninsula (represented in Figure 26 as green dots). The complete mtDNA sequences are available in Table S1.

3) The genetic exchanges across the Red Sea were investigated through the study of 26 L4 and L6 African lineages (represented in Figure 26 as light blue dots), which are related with the out-of-Africa L3 haplogroup. The complete mtDNA sequences are available in Table S1.

5.2. DNA extraction

Although the samples were previously characterised, in some cases I had to re-extract the DNA from the original samples. Depending on the amount of sample available, I used two different methods: Chelex® 100 and phenol-chloroform protocols.

5.2.1. Chelex® 100 protocol

Chelex 100 (Bio-Rad) is an ion exchange resin used to purify other compounds via ion exchange and its ability to bind transition metal ions. The Chelex appears to protect the DNA from the effects of the heating through sequestering divalent heavy metals that could damage the DNA. The following Chelex 100 protocol describes the procedure to obtain a total of 200 µl of DNA.

- Cut 1 cm^2 square of blood spot on the FTA™ card (Whatman™) or transfer 10-15 µl of blood sample to a 1.5 ml tube;

- Add 1 ml of sterilized water and briefly vortex during 30 minutes;
- Centrifuge at 20817g for 4 minutes and discard the supernatant containing the cellular fragments, leaving the substrate in the tube;
- Add 200 µl of Chelex 100 (5%) to the tube containing the substrate and Incubate at 56ºC during 30 minutes;
- Briefly vortex the tubes for a few seconds;
- Incubate the tubes at 100ºC during 8 minutes;
- Briefly vortex the tubes for a few seconds;
- Centrifuge during 4 minutes at 20817g;
- Store the resulting DNA at -20ºC for posterior amplification.

5.2.2. Phenol-chloroform extraction

A phenol-chloroform extraction is a liquid-liquid extraction. The phenol-chloroform combination denatures proteins, reduces the partitioning of poly(A)+ mRNA into the organic phase and reduces the formation of insoluble protein complexes at the interphase. The following protocol describes the procedure to obtain a total of 100 µl of DNA (100 ng/µl).

1. Add 1 ml of whole blood lysis buffer to 300 µl of blood sample and mix (if the sample consists in a mouthwash, hair or blood spot go directly to step 3);

Lysis buffer:
150 mM NaCl
25 mM EDTA

2. Centrifuge at 19357g for 3 minutes and remove supernatant;
3. Suspend the pellet with 44 µl of 10% SDS, 200 µl RBC lysis buffer and 5 µl of proteinase K (50 µg/ml). A master mix should be prepared first;

RBC lysis buffer:
0.32 M sucrose
1% Triton x100
5 mM $MgCl_2$
12 mM Tris-HCl pH 7.5

4. Digest at 54ºC for 2 hours and then at 37ºC overnight;
5. Add 250 µl (5M) NaCl and leave on ice for 40 minutes;
6. Centrifuge at 19357g for 30 minutes;

7. Add 500 µl of phenol: chloroform: Isoamyl Alcohol (25:24:1), mix and centrifuge at 19357g for 10 minutes.

8. Transfer the upper aqueous phase to a new tube and repeat the step 7.

9. Transfer the upper aqueous phase to a new tube and add 150 µl of (7.5M) ammonium acetate and 1 mL of 100% ethanol (ice-cold).

10. Leave at -20ºC overnight

11. Centrifuge at 19357g for 20 minutes and remove supernatant.

12. Wash the pellet in 500 µl of 70% ethanol (ice-cold).

13. Centrifuge at 19357g for 10 minutes and remove supernatant.

14. Repeat the steps 12 and 13.

15. Air-dry the pellet.

16. Suspend the pellet in 100 µl of water.

The quantity and quality of DNA were measured using the full-spectrum spectrophotometer NanoDrop ND-1000. The purity of DNA was assessed by the ratio of absorbance at 260 nm and 280 nm; a ratio of ~1.8 indicates pure DNA and lower values may indicate the presence of proteins, phenol or other contaminants. Another ratio of absorbance at 260 nm and 230 nm (indicative pure value of ~2.0) was also assessed, as this may indicate other contaminants which absorb at 230 nm. All the DNA extracted was diluted to a final concentration of 100 ng/µl.

5.3. Whole-genome amplification

Some of the original samples were already exhausted, so I had to amplify the low amount of DNA available by using the Illustra GenomiPhi V2 DNA Amplification Kit (supplied by GE Healthcare). This kit contains random hexamer primers that anneal at multiple sites of the DNA template, allowing the whole-genome amplification.

- Mix 1 µl of template DNA with 9 µl of sample buffer (GenomiPhi Sample Buffer provided with the kit) ;
- Heat at 95ºC for 5 minutes;
- Cool at 4º C to avoid the double-strand DNA re-formation;
- Combine 9 µl of reaction buffer with 1 µl of enzyme mix on ice (solutions provided in the kit) and add to the sample above;
- Incubate at 30ºC for 2 hours;
- Heat the sample to 65ºC for 10 minutes and cool at 4ºC;

- Dilute the product 10 times.

5.3.1. Polymerase Chain Reaction (PCR)

PCR (Polymerase Chain Reaction) is a method developed by Kary Mullis in the 1980s. PCR is based in the ability of DNA polymerase to synthesize a new strand of DNA, making it possible to obtain millions of copies of a given fragment of DNA from a small amount of starting DNA.

All samples were already typed for the HVS-I and HVS-II regions of mitochondrial DNA. For amplification of these regions two pairs of primers were used, the L15997 (5'-CAC CAT CAC CCA AAG TAG CT-3') and H16401 (5'-TGA GGA GGA TTT TGG CAC TG-3'); L48 (5'-CTC ACG GGA CTC CAT GCT GC-3') and H408 (5'-CTG TTA AAA GTG CAT ACC GCC A-3') for HVS-I and HVS-II respectively. The characterization of these regions allowed the classification into haplogroups and I selected the samples from the specific haplogroups to perform complete mtDNA sequencing.

In order to perform the complete sequencing of the mtDNA I applied two different protocols. The 85 samples belonging to haplogroups N1, N2 and X and the 45 J and T samples were amplified at the University of Leeds, while the 26 L4 and L6 lineages were amplified at IPATIMUP. The main difference in the sequencing protocols between labs is the number of primers used, due to the size of the capillary installed in the sequencer machines, implying different size limits for the fragments to be sequenced.

5.3.1.1. Protocol at IPATIMUP

The complete mtDNA was amplified by using 32 overlapping fragments described in Table 1. For each sample, the reaction mix contained a final volume of 25 µL:

- H_2O
- 200µM dNTPs (Bioline™)
- 10 x GoTaq® DNA Polymerase buffer (Promega™):
  10mM Tris-HCL (pH8,8)
  50mM KCL
  1,5mmol/L $MgCL^2$

- 0.2 µM primers
- 5u/µL of GoTaq® DNA polymerase (Promega™):
- 100 ng/ µL DNA Sample

The conditions of the amplification reaction were as follows:

Initial denaturation step:	95ºC for 2 minutes
35 cycles:	95ºC for 30 seconds
	60ºC for 30 seconds
	72ºC for 30 seconds
Final extension:	72ºC for 10 minutes

Table 1 Primers used in the protocol at IPATIMUP to amplify the complete mitochondrial genome

Fragment name	Fragment size (bp)	Primer sequence (5' – 3')
P1F-L16340	681	AGCCATTTACCGTACATAGCACA
P1R-H408		TGTTAAAAGTGCATACCGCCA
P2F-L382	603	CAAAGAACCCTAACACCAGCC
P2R-H945		GGGAGGGGGTGATCTAAAAC
P3F-L923	607	GTCACACGATTAACCCAAGTCA
P3R-H1487		GTATACTTGAGGAGGGTGACGG
P4F-L1466	629	GAGTGCTTAGTTGAACAGGGCC
P4R-H2053		TTAGAGGGTTCTGTGGGCAAA
P5F-L2025	609	GCCTGGTGATAGCTGGTTGTCC
P5R-H2591		GGAACAAGTGATTATGCTACCT
P6F-L2559	591	CACCGCCTGCCCAGTGACACAT
P6R-H3108		TCGTACAGGGAGGAATTTGAA
P7F-L3073	640	AAAGTCCTACGTGATCTGAGTTC
P7R-H3670		GGCGTAGTTTGAGTTTGATGC
P8F-L3644	623	GCCACCTCTAGCCTAGCCGT
P8R-H4227		ATGCTGGAGATTGTAATGGGT
P9F-L4210	625	CCACTCACCCTAGCATTACTTA
P9R-H4792		ACTCAGAAGTGAAAGGGGGCTA
P10F-L4750	599	CCAATACTACCAATCAATACTC
P10R-H5306		GGTGATGGTGGCTATGATGGTG
P11F-L5278	593	TGGGCCATTATCGAAGAATT
P11R-H5832		GACAGGGGTTAGGCCTCTTT
P12F-L5781	626	AGCCCCGGCAGGTTTGAAGC

P12R-H6367		TGGCCCCTAAGATAGAGGAGA
P13F-L6337	601	CCTGGAGCCTCCGTAGACCT
P13R-H6899		GCACTGCAGCAGATCATTTC
P14F- L6869	578	CCGGCGTCAAAGTATTTAGC
P14R-H7406		GGGTTCTTCGAATGTGTGGTAG
P15F-L7379	580	AGAAGAACCCTCCATAAACCTG
P15R-H7918		AGATTAGTCCGCCGTAGTCG
P16F-L7882	506	TCCCTCCCTTACCATCAAATCA
P16R-H8345		TTTCACTGTAAAGAGGTGTTGG
P17F-L8299	603	ACCCCCTCTAGAGCCCACTG
P17R-H8861		GAGCGAAAGCCTATAATCACTG
P18F-L8799	638	CTCGGACTCCTGCCTCACTCA
P18R-H9397		GTGGCCTTGGTATGTGCTTT
P19F-L9362	609	GGCCTACTAACCAACACACTA
P19R-H9928		AACCACATCTACAAAATGCCAGT
P20F-L9886	617	TCCGCCAACTAATATTTCACTT
P20R-H10462		AATGAGGGGCATTTGGTAAA
P21F-L10403	612	AAAGGATTAGACTGAACCGAA
P21R-H10975		CCATGATTGTGAGGGGTAGG
P22F-L10949	617	CTCCGACCCCCTAACAACCC
P22R-H11527		CAAGGAAGGGGTAGGCTATG
P23F-L11486	629	AAAACTAGGCGGCTATGGTA
P23R-H12076		GGAGAATGGGGGATAGGTGT
P24F-L12028	615	GGCTCACTCACCCACCACATT
P24R-H12603		ACGAACAATGCTACAGGGATG
P25F-L12572	591	ACAACCCAGCTCTCCCTAAG
P25R-H13124		ATTTTCTGCTAGGGGGTGGA
P26F-L13088	618	AGCCCTACTCCACTCAAGCAC
P26R-H13666		AGGGTGGGGTTATTTTCGTT
P27F-L13612	614	AAGCGCCTATAGCACTCGAA
P27R-H14186		TGGTTGAACATTGTTTGTTGG
P28F-L14125	602	TCTTTCTTCTTCCCACTCATCC
P28R-H14685		CATTGGTCGTGGTTGTAGTCC
P29F-L14650	604	CCCCATTACTAAACCCACACTC
P29R-H15211		TTGAACTAGGTCTGTCCCAATG
P30F-L15162	597	CTCCCGTGAGGCCAAATATC
P30R-H15720		GTCTGCGGCTAGGAGTCAAT
P31F-L15676	524	TCCCCATCCTCCATATATCC
P31R-H16157		TGATGTGGATTGGGTTTTTATGTA

P32F-L15996	446	CTCCACCATTAGCACCCAAAGC
P32R-H16401		TGATTTCACGGAGGATGGTG

5.3.1.2. Protocol at the University of Leeds

The complete mtDNA was amplified using 22 overlapping fragments described in Table 2. For each sample, the reaction mix contained a final volume of 35 µL:

25 µL:

- H_2O
- 200µM dNTPs (Bioline™)
- 10 x GoTaq® DNA Polymerase buffer (Promega™):
  10mM Tris-HCL (pH8,8)
  50mM KCL
  1,5mmol/L $MgCL^2$

- 0.2 µM primers
- 5u/µL of GoTaq® DNA polymerase (Promega™):
- 100 ng/ µL DNA Sample

The PCR reaction consists in the following cycle:

Initial denaturation step:	94ºC for 5 minutes
10 cycles:	94ºC for 45 seconds
	57ºC for 45 seconds (decrease of 0.5ºC per cycle until 52.5ºC)
	72ºC for one minute and 20 seconds
30 cycles:	94ºC for 45 seconds
	52ºC for 45 seconds
	72ºC for 1 minute and 20 seconds
Final extension:	72ºC for 10 minutes

Table 2 Primers used in the protocol at the University of Leeds to amplify the complete mitochondrial genome.

Fragment name	Fragment size (bp)	Primer sequence (5' – 3')
P1F -15873	1085	TACTCAAATGGGCCTGTCCT
P1R-388		TGGTTAGGCTGGTGTTAGGG
P2F-16413	884	TGAAATCAATATCCCGCACA
P2R-727		AGGGTGAACTCACTGGAACG
P3F-449	1018	TTATTTTCCCCTCCCACTCC
P3R-1466		GGCCCTGTTCAACTAAGCAC
P4F-1331	1012	AAGGTGTAGCCCATGAGGTG
P4R-2342		AGGCTTATGCGGAGGAGAAT
P5F-2007	1125	TGGTGATAGCTGGTTGTCCA
P5R-3169		GGAAGGCGCTTTGTGAAGTA
P6F-2835	1022	CCAACCTCCGAGCAGTACAT
P6R-3894		GGTTCGGTTGGTCTCTGCTA
P7F-3587	940	CCCTGGTCAACCTCAACCTA
P7R-4526		GATGAGTGTGCCTGCAAAGA
P8F-4346	1125	GAACCCATCCCTGAGAATCC
P8R-5470		GTGGTAAGGGCGATGAGTGT
P9F-5162	935	TCGCACCTGAAACAAGCTAA
P9R-6096		TTACAAATGCATGGGCTGTG
P10F-5910	916	GCCGACCGTTGACTATTCTC
P10R-6825		CGGAGGTGAAATATGCTCGT
P11F-6643	1176	TCCTACCAGGCTTCGGAATA
P11R-7818		AGGGCGATGAGGACTAGGAT
P12F-7494	1040	CATGGCCTCCATGACTTTTT
P12R-8533		TATTTGGAGGTGGGGATCAA
P13F-8389	945	ATGGCCCACCATAATTACCC
P13R-9333		GGAGCGTTATGGAGTGGAAG
P14F-9183	993	CCTCTACCTGCACGACAACA
P14R-10175		GCACTCGTAAGGGGTGGAT
P15F-9815	1044	CCACGGACTTCACGTCATTA
P15R-10858		AATTAGGCTGTGGGTGGTTG
P16F-10609	1159	TAACCCTCAACACCCACTCC
P16R-11767		GCGTTCGTAGTTTGAGTTTGC
P17F-11402	1143	TGACTCCCTAAAGCCCATGT
P17R-12544		TGGCTCAGTGTCAGTTCGAG
P18F-12227	1073	CTAACTCATGCCCCCATGTC
P18R-13299		TTGGTTGATGCCGATTGTAA
P19F-12913	1156	TCCAACTCATGAGACCCACA

P19R-14068		AGGTGATGATGGAGGTGGAG
P20F-13714	1143	GGAAGCCTATTCGCAGGATT
P20R-14856		AGGAGTGAGCCGAAGTTTCA
P21F-14478	1121	CAACCATCATTCCCCCTAAA
P21R-15598		GACGGATCGGAGAATTGTGT
P22F-15195	1245	TATCCGCCATCCCATACATT
P22R-16439		GCACTCTTGTGCGGGATATT

5.4. Gel electrophoresis

Gel electrophoresis is a method for separation of macromolecules based on their size and charge. The technique allows the separation of the nucleic acid molecules (negatively charged due to its sugar-phosphate), by applying an electric field to move the negatively charged molecules through a matrix.

In order to confirm the mtDNA amplification by PCR I used two different electrophoresis techniques based on two distinct types of gel - agarose gel at the University of Leeds and polyacrylamide at IPATMUP.

5.4.1. Agarose gel

Agarose gel electrophoresis is the most popular technique used for the separation of moderate and large-sized nucleic acids. The process is fast and it has a wide range of separation but a relatively low resolution power.

The matrix consisted of a 2% agarose gel (2 g of agarose, 100 mL of 0.5 x TBE buffer and a drop of 5 mM ethidium bromide for visualization of the fragments). The length of the DNA bands was confirmed using a ladder of 100 bp intervals (Promega). The samples were subjected to a constant voltage at 120 V and 500 mA for approximately 20 minutes and the fragments were visualized using a UV transilluminator.

5.4.2. Polyacrylamide gel

A polyacrylamide gel is a stable chemically cross-linked gel. Although implying a more time consuming preparation, it has a higher resolving power and therefore is good for separation of low molecular weight fragments.

The gel was prepared manually in a manufactured tape (3 mm thick) containing a hydrophilic film (Gel Bond).

The gel consists in the following solution: 4.4 ml of the stock solution T9C5 gel (9% acrylamide and 5% piperazine) in a buffer consisting of 0.375 M Tris / HCl (pH 8.8), 350 µl glycerol (100%), 250 µl ammonium persulphate (1%) and 10 µl of TEMED. The buffer (0.125 M Tris / Glycine, pH 8.8) was soaked in a paper strip "Whatman" (10 cm x 2 cm) and placed directly on the gel. Bromophenol blue was used as a tracking dye to monitor the progression of the run as it migrates at a known rate. The electrophoretic system is a horizontal Multiphor plate (Pharmacia) with temperature controlled at 4°C. The electrical conditions were controlled by a fixed voltage of 200 V.

The DNA fragments were visualized by silver nitrate staining, according to the method described by Budowle et al.[268] Initially, the DNA was fixed by immersing the gel in 10% ethanol for 10 minutes under stirring, and then 1% nitric acid for 5 minutes, also under stirring. After washing the gel with distilled water for 10 seconds (twice), it was placed in a solution of 0.2% silver nitrate for 20 minutes under stirring and in the absence of light. After washing the gel with distilled water, this was placed in 0.28 M sodium carbonate, 0.02% formaldehyde and stirring until visualization of the fragments. In order to stop the reaction the gel was placed in 10% acetic acid for 10 seconds. Finally the gel was washed with water and retained after drying at room temperature.

The samples with good quality amplification followed for automated sequencing.

5.5. Capillary Sanger sequencing

The complete sequencing was performed by using the Sanger method, based on the use of dideoxynucleotides (ddNTP's) in addition to the normal nucleotides (NTP's) found in DNA. The dideoxynucleotides contain a hydrogen group on the 3' carbon instead of a hydroxyl group (OH) being responsible for stopping the replication process when they are incorporated into the growing strand of DNA, resulting in varying lengths of short fragments of DNA. These short fragments will be separated by capillary electrophoresis and because the ddNTPs are

fluorescently labelled, each nucleotide at the end of each fragment can be identified. The sequencing was performed only in the direct-strand, except in some cases, as when the sequence bears a poly-C stretch, for which the sequencing had to be repeated with the reverse-strand direction.

The samples from the University of Leeds were sent to the GATC Biotech company, and later, to Eurofins MWG Operon for complete mtDNA sequencing using the ABI 3730xl DNA Analyzers (AB Applied Biosystems). The direct-strand primers used for the sequencing reaction are described in the Table 3.

Table 3 Primers used in the University of Leeds for sequencing reaction.

Fragment name	Primer sequence (5' – 3')
P1R-131	ACAGATACTGCGACATAGGG
P2F-16521	TAAAGCCTAAATAGCCCACA
P3F-614	AATGTTTAGACGGGCTCAC
P4F-1402	AAACTTAAGGGTCGAAGGTG
P5F-2176	AAAGCAGCCACCAATTAAG
P6F-2897	ATCCAATAACTTGACCAACG
P7F3638	TAGCCGTTTACTCAATCCTC
P8F-4410	CAGCTAAATAAGCTATCGGG
P9F-5191	CACCCTTAATTCCATCCAC
P10F-5999	TCTAAGCCTCCTTATTCGAG
P11F-6643	TCCTACCAGGCTTCGGAATA
P12F-7614	AAGACGCTACTTCCCCTATC
P13F-8423	CTATTCCTCATCACCCAACT
P14F-9213	CACCAATCACATGCCTATC
P15F-9922	CCTGATACTGGCATTTTGTAG
P16F-10689	GGCCTAGCCCTACTAGTCTC
P17F-11452	TGCCGCAGTACTCTTAAAAC
P18F-12246	CTAACAACATGGCTTTCTCA
P19F-12973	CTACTAGGCCTCCTCCTAGC
P20F-13723	TTCGCAGGATTTCTCATTAC
P21F-14546	ATAATAACACACCCGACCAC
P22F-15324	CAACACTCCACCTCCTATTC

At IPATIMUP the sequencing protocol was performed by me until the submission of the final product to the sequencing machine (completed by a specialised technician). The IPATIMUP protocol is describe as follows:

5.5.1. First purification

Prior to the sequencing reaction, the fragments were purified by using Microspin column purification S-300 HR (GE Healthcare) following the manufacturer's instructions:

- The tip of the column was cut out, and the column was introduced in a 1.5 ml tube and centrifuged at 740 g for 1 minute.
- The cap was removed and the column was inserted in a new 1.5 tube; the total PCR product was pipetted into the column.
- Finally the column with the PCR product was centrifuged at 750 g for 2 minutes.

5.5.2. Sequencing reaction

The direct-strand primers (Table 4) used in the amplification reaction were also used in the reaction mix (final volume of 5 µL):

- 1 µl of Big Dye Terminator Cycle Sequencing Ready Reaction (AB Applied Biosystems)
- 1 µl of the dilution buffer (AB Applied Biosystems)
- 0.5 µl of the direct-strand primer (2.5 mM)
- 2.5 µl of the DNA purified above.

The PCR reaction consists in the following cycle:

Initial denaturation step:	96ºC for 4 minutes
35 cycles:	95ºC for 15 seconds
	50ºC for 9 seconds
	60ºC for 2 minutes
Final extension:	60ºC for 10 minutes

Table 4 Primers used in IPATIMUP for the sequencing reaction.

Fragment name	**Primer sequence (5' – 3')**
P1F-L16340	AGCCATTTACCGTACATAGCACA
P2F-L382	CAAAGAACCCTAACACCAGCC
P3F-L923	GTCACACGATTAACCCAAGTCA
P4F-L1466	GAGTGCTTAGTTGAACAGGGCC

P5F-L2025	GCCTGGTGATAGCTGGTTGTCC
P6F-L2559	CACCGCCTGCCCAGTGACACAT
P7F-L3073	AAAGTCCTACGTGATCTGAGTTC
P8F-L3644	GCCACCTCTAGCCTAGCCGT
P9F-L4210	CCACTCACCCTAGCATTACTTA
P10F-L4750	CCAATACTACCAATCAATACTC
P11F-L5278	TGGGCCATTATCGAAGAATT
P12F-L5781	AGCCCCGGCAGGTTTGAAGC
P13F-L6337	CCTGGAGCCTCCGTAGACCT
P14F- L6869	CCGGCGTCAAAGTATTTAGC
P15F-L7379	AGAAGAACCCTCCATAAACCTG
P16F-L7882	TCCCTCCCTTACCATCAAATCA
P17F-L8299	ACCCCCTCTAGAGCCCACTG
P18F-L8799	CTCGGACTCCTGCCTCACTCA
P19F-L9362	GGCCTACTAACCAACACACTA
P20F-L9886	TCCGCCAACTAATATTTCACTT
P21F-L10403	AAAGGATTAGACTGAACCGAA
P22F-L10949	CTCCGACCCCCTAACAACCC
P23F-L11486	AAAACTAGGCGGCTATGGTA
P24F-L12028	GGCTCACTCACCCACCACATT
P25F-L12572	ACAACCCAGCTCTCCCTAAG
P26F-L13088	AGCCCTACTCCACTCAAGCAC
P27F-L13612	AAGCGCCTATAGCACTCGAA
P28F-L14125	TCTTTCTTCTTCCCACTCATCC
P29F-L14650	CCCCATTACTAAACCCACACTC
P30F-L15162	CTCCCGTGAGGCCAAATATC
P31F-L15676	TCCCCATCCTCCATATATCC
P32F-L15996	CTCCACCATTAGCACCCAAAGC

5.5.3. Final purification

At the end of the sequencing reaction, the samples were subjected to a final purification step to remove the unincorporated nucleotides and primers. This final purification was done by using columns of Sephadex™ (GE Healthcare), which functions as a gel filtration. The method consists in inserting the column into 1.5 mL tube charged with 750 µL of Sephadex and centrifuged at 1700 g for 4 minutes. The column was removed and placed in a new 1.5 ml tube. The 5 µL of the sample was then charged into the column and centrifuged at 1700 g for 4 minutes. A total

of 8 µL of ultrapure formamide (AB Applied Biosystems) was added to the purified sample and the remaining procedure was done by the IPATMUP technician.

5.6. Sequencing analysis

All the sequencing results were analysed using a 4.10.1 demo version of the Sequencher Software (Gene Codes Corporation, Ann Arbour, MI USA). Sequencher is a bioinformatics program for DNA sequence analysis which allows the alignment of the entire mtDNA genome against a reference sequence. I used the revised Cambridge Reference Sequence (rCRS)[49] as reference sequence. This tool allows multiple sequence alignment and has comprehensive DNA sequence editing tools that identify all the nucleotide variants using the same standard numbering system used for the rCRS (1-16569 bp).

All ambiguous positions were checked a second time by using the reverse primer. The transitions were indicated by the position alone, while the transversions were represented by the position followed by the mutated base. For the indels, a “d” or a “i” were added after the mutated position in case of deletion and insertion, respectively; when the indel occurred in a stretch of a repeated sequence, the last position is the one annotated.

5.6.1. Sequence data management

5.6.1.1. VNALL

Sequencher Software available is a demo version and for that reason the software did not allowed me to assemble and save all the fragments into a single fragment. Therefore I used the software VNALL kindly provided by Dr. Vincent Macaulay, to create a FASTA/PHYLIP format file from a single input file that could be readable by other bioinformatics tools.

5.6.1.2. Geneious

For a fast download of sequences in FASTA format from GenBank (http://www.ncbi.nlm.nih.gov/genbank/), I used the program Geneious

(http://www.geneious.com). The demo version of the Geneious software allows one to search in PubMed (by author, title, abstract content or accession number) and have the references saved directly into a local folder in our copy of Geneious.

5.6.1.3. BioEdit

All the sequences were aligned against the revised human mitochondrial reference sequence, the rCRS, by using the BioEdit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html). BioEdit is a program which includes the ClustalW algorithm that allows multiple sequences to be aligned and saved as FASTA files. The automated alignment was checked manually, in order to make uniform the inclusion of deletions or insertions in poly-nucleotides stretches.

5.6.1.4. mtDNASyn

In order to facilitate the identification of polymorphic positions relative to the rCRS in large datasets of complete mtDNA sequences I used mtDNASyn.[269] This program works by importing a FASTA file with the sequences aligned; the program identifies the positions that are variable relative to the reference; and the list of these variants can be exported as a .txt file which can be used in the software for haplogroup affiliation and as a .rdf binary file to be used in the Network software for phylogenetic inference.

5.6.2. Haplogroup affiliation

Haplogroup and subhaplogroup affiliation of the samples was done by using the HaploGrep software[267] and double-checked manually against PhyloTree.[270] The HaploGrep is based on the PhyloTree standard[270], which regularly updates the annotated mitochondrial tree estimated from worldwide data. The application generates the top ten best results of haplogroup affiliation and the phylogenetic position of the respective haplogroup in the general tree.

5.6.3. Samples for comparison

5.6.3.1. Published complete mtDNA samples

I also collected other complete mtDNA sequences already published and deposited in GenBank. I used 300 complete sequences belonging to the haplogroups N1, N2 and X (Table S2), 1676 samples J/T complete sequences (Table S2) and 31 complete L4 and L6 sequences (Table S2).

5.6.3.2. HVS-I/HVS-II database

With the help of my co-workers (Joana Pereira, Marta Costa, Pedro Soares and Luísa Pereira), I constructed an Excel database for control region (HVS-I and HVS-II) mtDNA sequences already published and deposited in GenBank or reported in papers. I organized all the HVS-I/HVS-II sequences from Arabia while Marta Costa was responsible for the European sequences, Joana Pereira organized all the Near East and North Africa sequences and Pedro Soares the African ones. Finally, Pedro Soares and Luísa Pereira checked and confirmed the haplogroup affiliations.

The resulting database included 42,485 mtDNA sequences covering Europe, the Near East, Arabia, sub-Saharan and North Africa with the control region variants (HVS-I and in some cases HVS-II), haplogroup, geographic region, reference, any RFLPs tested and ethnic group whenever this information was provided.

The information from the database was used in various analysis such as the construction of the interpolation maps and the reconstruction of preliminary reduced-median networks to be employed in the founder analysis (information for Supplementary Tables is provided in the corresponding Results sections).

5.7. Phylogenetic and statistical analyses

The main analyses I applied to the mtDNA data were: inferring the phylogenetic tree which represents the evolutionary relationships among sequences; dating of nodes within the tree; founder analysis, to trace and date the movements of maternal lineages from a source region to a new territory; interpolation maps

representing the distribution of the haplogroup frequency across a broad geographical region.

5.7.1. Phylogenetic analysis and age estimation

Phylogenetic trees are graphs that group species/individual genetic sequences in a hierarchical order of descent, allowing us to understand their evolutionary relationships. Trees are applied to data from non-recombining systems from human populations, in order to infer the history of recent human migrations.[86; 271]

A tree consists of nodes connected by branches, with internal nodes representing ancestral species or sequences. Trees can be defined as unrooted or rooted (Figure 27).[2] The former only specify the relationship between the taxa but not the evolutionary path, while the latter include information on the oldest ancestor from which all the other taxa diverge, and therefore gives information about the evolutionary pathways. This node is identified as the most recent common ancestor (MRCA) of the sequences or species in the tree.[172] It is possible to transform an unrooted tree into a rooted tree by two different methods. The first is to assume that the root falls midway along the longest branch on the tree (mid-point rooting) and the second is to add an outgroup to the tree. An outgroup is a taxon which diverged before the most recent common ancestor of the group under consideration. For instance, the chimpanzee is often used to root human trees; partial human mitochondrial trees focusing on a certain haplogroup can be easily rooted by including a sample from another human haplogroup. In a phylogenetic tree, the groups are monophyletic if all the taxa within the group descend from a single common ancestor forming a natural clade, and paraphyletic if it does not include all the descendants of a single common ancestor.[2]

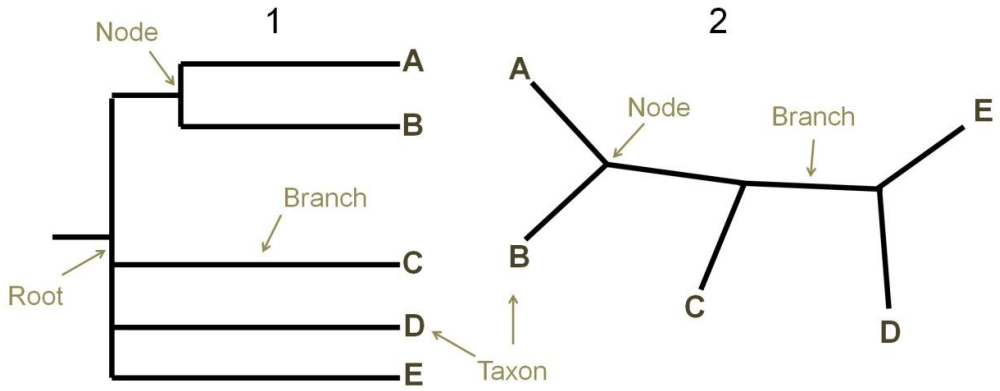

Figure 27 Phylogenetic trees.
An example of a rooted tree (1) and an unrooted tree (2).

There are many methods for reconstructing phylogenies from genetic data. The methods based on optimality criteria such as maximum parsimony (MP), maximum likelihood (ML) or Bayesian inference, consider all the possible trees and choose the tree that best fits the data. Of course, the implicit assumptions of each method will greatly influence the global structure of the evolutionary inference. Alternatively, methods based in a clustering algorithm use a distance matrix and an interactive clustering process that combines taxa together in a hierarchical fashion. The distance methods include algorithms such as neighbour-joining (NJ) and Unweighted Pair-Group Method with Arithmetic Mean (UPGMA).[2] The main concern when using distance methods is that information is lost by converting character data to distances, leading to just an overall relationship between the tree and data.[271]

In this work I employed the MP and Bayesian methods, and I also used ML to infer branch lengths.

5.7.1.1. Maximum parsimony, Network and ρ statistic

MP is a non-parametric phylogenetic method that attempts to minimise the amount of evolutionary change needed to explain the data.[2] Therefore MP chooses the tree (or trees) with the shortest tree length, minimizing the number of homoplasies. A drawback is that MP will tend to *under-estimate* the actual amount of change, unless the rate of change and the level of homoplasy are very small.[2]

In mtDNA it is not straightforward to infer the MP tree. Reticulation and ambiguous data render unclear which is the most parsimonious pathway, so that the evolutionary path can be best represented by networks. Networks represent alternative plausible evolutionary pathways, allowing us to explore the most likely alternatives and make explicit decisions when choosing between them.

The Network software can predict ancestral unobserved haplotypes, identify where homoplasy occurs, estimate the number of recurrences, and estimate ages of nodes when a mutation rate is provided[2; 272]. There are two alternative methods to infer the network, named: reduced-median and median-joining networks. The reduced-median network resolves the reticulations from the median network by eliminating the least likely links. When using large samples sizes, the method distinguishes when a reticulation in the network may justifiably be resolved on a compatibility basis in conjunction with a frequency-based argument and leaving plausible evolutionary solutions.[272] Median-joining networks are less accurate than

the reduced median approach, but can be useful for large datasets. The algorithm departs from the minimum-spanning trees, which are combined within a single (reticulate) network, to which a few consensus sequences (median vectors) of three mutually close sequences at a time are added.[2]

In order to find the MP trees that best represent the evolutionary history of the haplogroups analysed in this work, I used the reduced-median algorithm in the Network 4.6 software (www.fluxus-engineering.com).[273] The highly recurrent positions 16182, 16183 and 16519 were excluded from the analysis as well as the indel positions 309, 315, 524 and 573. By default, the weighting value of each position was scored as 10. However, in order to decrease the level of reticulation the weights of the ten fastest mutations[116] were diminished to 5. The remaining reticulation was resolved manually, by employing several criteria,[274] such as: choosing the slowest-evolving mutations as occurring only once and the fastest recurring several times; when both mutations are slow, high frequency haplotypes were assumed to be more ancestral.

After manually resolving the networks reticulation, the branching order for the trees is thus defined, and then I constructed the most parsimonious trees by hand in an Excel file.

Merging lineages by moving back through the tree from the present allows us to estimate the time to the MRCA (TMRCA). In the rooted MP trees of this thesis, I estimated the ages of nodes by the average distance rho (ρ), which was developed by Forster et. al.[117] The ρ statistic represents the average number of mutational changes sites between a set of sequences and a specified common ancestor.[2; 117]

$\rho = \sum_{i=1}^{m} (n_i / n)\rho_i$ (ρ_i refers to the last common ancestor of the pooled population and n_i the sample size)

The estimation of the variance is $\sigma^2 = \frac{\left(\sum_{i=1}^{k} n_i{}^2 l_i\right)}{n^2}$,($l_i$ refers to the branch lengths, n_i is the number of samples associated with each branch length, k represents each node and n the total number of samples).

The ρ statistic can be used with any molecular clock and is a useful and reliable tool.[117] For mtDNA it has been shown that the amount of non-synonymous mutations is higher at the tips of the tree and decreases with increasing age of the branches, which would lead to an overestimation of the age estimates for the

younger branches if no correction were applied. Soares et al.[116] published a correction curve that takes into account the role of purifying selection over time, by calculating the amount of non-synonymous mutations at various phylogenetic time depths and normalising it by the corresponding amount of synonymous mutations (which are neutral to selection) and, therefore, correcting the mutation rate at any time depth. I used a mutation rate estimate for the complete mtDNA sequence of one substitution in every 3,624 years, correcting for purifying selection by using the calculator provided by Soares et al.[116] These authors also estimated a mutation rate for synonymous mutations only at one substitution every 7,884 years, and one mutation every 16,677 years for HVS-I (in the range 16,051-16,400 bp), that I also applied here after correcting for purifying selection as referred. Standard errors were calculated based on 95% confidence intervals ($\rho - 1.96 \times \sigma^2$; $\rho + 1.96 \times \sigma^2$) as in Saillard et al.[275]

5.7.1.2. Maximum likelihood

Maximum likelihood (ML) is a statistical method that infers evolutionary history by maximizing the probability of obtaining the observed data.[86] It estimates the amount of evolutionary change by using evolutionary models, which are based on several parameters such as mutation rate, transition/transversion ratio and diverse base composition among loci.[2] ML allows strict or relaxed molecular clock models, that is, constant or variable rates across lineages in the inferred phylogeny, as well as substitution heterogeneity across sites, or, in other words, variability in rates between regions of the studied molecule.

PAML (Phylogenetic Analysis by Maximum Likelihood) is a package of programs (BASEML, BASEMLG, CODEML, EVOLVER, PAMP, YN00, MCMCTREE, and CHI2) that use ML analysis.[276] The program BASEML can compare and statistically evaluate phylogenetic trees inferred with a variety of different models.[276] It has been shown,[116] in large mtDNA databases, that the Kishino and Yano 1985 model (HKY85) displays a significantly improved performance when compared with other models. The HKY85 distinguishes between the rate of transitions and transversions (the probability of a transition to occur is higher than a transversion) and allows unequal base frequencies.

For my data, I provided the MP trees inferred using network to PAML 3.13[276] assuming the HKY85 mutation model with gamma-distributed rates (approximated by a discrete distribution with 32 categories). This allowed me to obtain ML

estimates of branch lengths, that then I converted to time using the clock calculator published in Soares et al.[116] as described for ρ.

5.7.1.3. Bayesian inference

The Bayesian method estimates parameters of an underlying distribution based on the observed distribution.[2] It applies a prior knowledge to determine the probability distribution of the reconstructed haplotypes, such as constant size population, exponential growth, and logistic (or expansion) growth.[277] The application of Bayesian methods grew up by introducing the Markov-chain Monte Carlo method (MCMC), which removed many of the computational problems. MCMC chooses an individual at random and estimates its haplotype assuming that all the other haplotypes are correctly reconstructed, and goes on repeating this process a large number of times.

The Bayesian skyline plot (BSP) is a model for estimating ancestral population dynamics given the known genotypes. It uses MCMC to estimate a distribution of effective population size through time given the genotypes and a specific nucleotide substitution model.[278]

BEAST is a powerful and flexible software package for Bayesian analysis of molecular sequences. As the ML, it also allows strict or relaxed molecular clocks[279] and substitution heterogeneity across sites. BEAST allows a flexible choice of priors on parameters, providing a general estimation and hypothesis testing of evolutionary models.

The BEAST program contain other two programs, BEAUti and Tracer, responsible for assisting in generating the input and analyzing the output, respectively. The BEAUti software (Bayesian Evolutionary Analysis Utility) is a BEAST graphical user interface that allows the user to import files for a number of simple model combinations and specify the settings for the MCMC sampler. Tracer is a graphical and statistical tool for analysing the output of BEAST and checking its performance and accuracy.[279; 280]

I generated BSPs[278] using BEAST 1.4.6[279], in order to detect population growth associated with the different haplogroups analyzed (N(xR), L4/L6, and J/T). For the haplogroup N(xR) lineages, I used a total of 363 complete sequences (I only left one representative of each of the Native American haplogroups X2a1 and X2a2, to avoid a signal of population expansion from the American continent) and used L2, L4 and L6 sequences as outgroups, with a relaxed molecular clock (lognormal in distribution across branches and uncorrelated between them) and the HKY model

of nucleotide substitutions with gamma-distributed rates. For this analysis, I used a mutation rate previously calibrated using internal calibration points in the L3 phylogeny, the ancestor of haplogroup N, of 2.6129 $x10^{-8}$.[99] BSPs were designed estimate the effective population size through time using random sequences from a given population, but have also proved effective with individual haplogroup data.[281] Thus, I generated BSPs by haplogroup (N1, N2, N2 without W in Finns, and X) and geographical regions (Europe, Near East/Caucasus/Arabian Peninsula and overall). With these two sub-sets of analyses I aimed to distinguish which haplogroups were mainly responsible for the increases in effective population size observed in the overall N(xR) analysis, and observe if a given region carried the signal for a given population increase in its specific N(xR) sequences, indicating a possible expansion within that region involving several subclades. The same parameters were used to generate BSPs for the L4 and L6 lineages (L4 alone and L4+L6) and the JT clade (J alone in Arabia, J in the Near East, T in Arabia and T in the Near East.

For all studies (haplogroup N(xR), JT and L4/L6), I ran 50,000,000-100,000,000 iterations, with samples drawn every 10,000 MCMC steps, after a discarded burn-in of 10,000,000 steps. I checked for convergence to the stationary distribution and sufficient sampling by inspection of posterior samples. I visualized the BSPs with Tracer v1.3. I used a generation time of 25 years and forced the larger subhaplogroups to be monophyletic in the analysis: MCMC updates which violated this assumption were rejected. In order to perform a systematic comparison and description of the increment periods in the effective population size of the BSP, I calculated a rate of population size change through time.

5.7.2. Founder analysis

In order to detect the time of migrations from one region to another and their impact upon the modern mtDNA pool, Richards et al.[259], developed a new method known as the founder analysis. Founder analysis is a method for the analysis of non-recombining DNA sequence data aiming to identify and date the proportion of lineages that moved into a new territory at a specific time, using the accumulated variation of a detected founder type within that region. This method requires the establishment of a source and a sink population and the presence of similar sequences in both locations, indicating the occurrence of a probable migration event. The founder analysis also takes into account the effects of both back-migrations and recurrent mutation and defines different levels of stringency for

identifying founders – *f0, f1*, *f2* and *fs*. The *f0* criterion takes into account all the possible founders, thus being more exposed to recurrent mutations and back-flow and allowing a higher number of false estimates;[259] thus I will not use *f0* criterion in the founder analysis. Both *f1* and *f2* reduce the impact of recurrent mutation and back-migrations by excluding sequences that result from parallel mutations, so that sequence matches are not at the tips of the source phylogeny. Founders must have at least one (*f1*) or two (*f2*) derived branches in the source population. In order to allow for variation in the frequency of founder cluster in the sink, the criterion *fs* was created and defines that the extent of variation required in the source population is stipulated by the frequency of the founder cluster in the sink population, as the probability of back-migration to the source depends on the frequency of the cluster in the sink population. The absolute frequency of the *fs* criterion was corrected by using logarithms to the closest integer.[259] However, the frequency based correction applied to the *fs* criterion was rejected due to the rather arbitrary formula used to identify the founders.[282]

I employed founder analysis for the N(xR), L(xMN) and Eurasian haplogroups. For the N(xR) sequences I mainly performed founder analyses from Southwest Asia to Europe as well as exploratory analyses from the Near East into North Africa. For the African L(xMN) haplogroups, I investigated the migration from Africa into Arabia plus the Near East. For the Eurasian haplogroups, I investigated first migration from the Near East, Iran and Pakistan into the Arabian Peninsula, and then migration from Arabia plus Near East into East Africa and, independently, into North Africa.

I assumed a strict division between source and sink populations and two criteria (*f1* and *f2*) for identifying founder sequences to partly account for homoplasy and back migrations. By using the Excel database of control region (HVS-I and HVS-II) sequences, the first step was to identify all the sequences present in the database for the haplogroups N(xR), African and other Eurasian. Subsequently, reduced-median networks in the range 16051–16400 bp were reconstructed for HVS-I, by converting the sequences to a Network-recognizable file (*tor* file) by using the program fm2net (developed by Christopher Snell, an MSc student at Leeds in Professor Martin Richards' group). The highly recurrent positions 16182 and 16183 were excluded from the analysis. By default, the weighting value of each position was scored as 10. However, in order to decrease the level of reticulation the weight of the ten fastest mutations[116] were diminished to a weight of 5. I performed the N(xR) network analyses and Pedro Soares obtained the networks for the African and other Eurasian haplogroups, which will be used in other founder analyses by our group to explore other migrations.

I then identified all the founders and descendants in the final networks, for the regions we are interested in for this work, by using an in-house computer tool.[283] The age of migration of each founder was estimated using the ρ statistic[117] and an HVS-I mutation rate of one mutation every 16,677 years.[116] In order to assess the error in the Bayesian partitioning across the different migration times realistically, the effective number of samples in each founder was calculated. This was obtained by multiplying the number of samples in each founder by a ratio of the variance assuming a star-like network and the variance calculated as in Saillard et al.[275] The distribution of founder ages was scanned for each region defining equally spaced 200-year intervals for each migration. For each case, I also investigated the proportion of introduction of lineages during putative migrations occurring in certain periods of time and represented the probabilistic proportion of introduction for each lineage at each of the putative migration periods in a graph resembling the images from the Structure analysis.

5.7.3. Interpolation maps

Combining geographic information with genetic data is a powerful approach for making inferences from genetic diversity and to understand the spatial impact of evolutionary processes.[2] Interpolation maps are performed by laying a fine grid on top of a map and estimating the gene frequency at each point. There are several methods of interpolation such as inverse distance weighted (IDW), kriging and natural neighbour. The IDW calculates the average of the neighbourhood data of each processing cell and assumes that each input point has a local influence that decreases with distance. The geographic location used is the centre of the distribution area from where individual samples of each population were collected.

ArcGIS is an integrated collection of GIS software for spatial analysis, data management and mapping. To determine and visualize the geographical distribution of the haplogroups studied in this work, I drew interpolation maps based on its frequency by using the ‘‘Spatial Analyst Extension’’ of ArcView version 3.2 (www.esri.com/software/arcview) and applying the IDW option with a power of two for the interpolation of the surface. All the frequency maps were generated using the data from the Excel database with control region (HVS-I and HVS-II) mtDNA sequences mentioned above. The frequencies used are listed in Tables S3, S7 and S10 for N(xR), JT and L4/L6, respectively.

I also constructed interpolation maps for the diversity values of ρ (which represents the average number of mutational changes sites between a set of

sequences and a specified common ancestor) and π (the mean number of nucleotide pairwise differences). These measures were calculated with the Network software, for the haplogroups I,W,X, J, T and L4, as reported in Table S4.

As frequencies and values of the diversity measures are very different for the various haplogroups, it was not possible to use the same scale in all the maps. This should be kept in mind when comparing the maps.

6. Genome-wide autosomal markers

6.1. Published genome-wide data

The genome-wide data for 709 samples from five geographic groups (Africa, Arabian Peninsula, Near East, Europe and South Asia) were collected from previously published data sets (Table 5). I did not include west North African populations because these have been already detailed surveyed for population structure and admixture with sub-Saharan Africans.[147]

Table 5 Samples used for genome-wide autosomal analysis.

Region	Populations	Abbreviation	N	Reference
Africa	Maasai, Kenya	Mas	19	HapMap
	Egypt	Egy	12	1
	Ethiopia	Eth	19	1
	Yoruba, Nigeria	Yor	21	2
	Bantu, Kenya	BN	19	2
Arabian Peninsula	Yemen	Yem	10	1
	Saudi Arabia	Sdi	20	1
Near East	Bedouin, Israel	Bdn	45	2
	Lebanon	Leb	7	1
	Syria	Syr	16	1
	Jordan	Jor	20	1
	Samaritan, Israel	Sm	3	1
	Druze, Israel	Drz	42	2
	Palestinian, Israel	Pal	46	2
	Iran	Irn	20	1
	Turkey	Tur	19	1
Jews	Yemen Jews	YJ	15	1
Europe	Belarus	Bel	9	1
	Lithuania	Lit	10	1
	France	Fr	28	2
	Orcadian, UK	Orc	15	2
	Russia	Ru	25	2
South Asia	Brahui, Pakistan	Brh	25	2
	Balochi, Pakistan	Blo	24	2
	Makrani, Pakistan	Mak	25	2
	Sindhi, Pakistan	Sin	24	2
	Pathan, Pakistan	Ptn	22	2
	Burusho, Pakistan	Bur	25	2
Total			709	

References: 1. Behar DM, et al. (2010) The genome-wide structure of the Jewish people. Nature 466, 238-242.

2. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100-1104.

The samples from Behar et al.[136] were genotyped using Illumina 610K and 660K bead arrays, while the samples from Li et al.[134] were screened with Illumina 650K bead arrays. The genotypes from the Maasai, an ethnic group located in Kenya, were obtained from the HapMap phase III release (http://hapmap.ncbi.nlm.nih.gov/). These data were published and quality control was performed by the authors, but I used PLINK 1.05[284] to re-check that individuals and SNPs had a genotyping success of 97%. PLINK is a freely available toolset for manipulation and quality control of large datasets of genotypes such as GWAS data. Using a compact binary file format to represent SNP data, as well as tools to transform the binary format to standard text-based formats, the software is able to perform data management, summary statistics, assessment of population stratification, association analysis and identify-by-descent estimations. PLINK uses the C/C++ programming language.

I used a Python in-house script to merge genotypes from the various chips and ended up with a total of 478,931 common autosomal single nucleotide polymorphisms (SNPs). The full dataset was pruned for linkage disequilibrium (LD), removing SNPs in strong LD ($r^2 > 0.1$) with nearby markers in a window of 50 SNPs (advanced by 10 SNPs each time); a total of 75,957 SNPs remained for further analyses.

6.2. Statistical analyses

A method that does not impose a bifurcating pattern on evolution is the principal component analysis (PCA), which is useful for revealing population relationships.[2] While principal component analysis seeks to discover the structure within the data through a non-parametric approach, there are other models, such as model-based ancestry estimation, that estimate ancestry coefficients as the parameters of a statistical model.[284] Approaches based on models-based ancestry estimation such as STRUCTURE, FRAPPE or ADMIXTURE are useful to estimate ancestries derived from multi-locus genotype data.[39]

6.2.1. Principal component analysis (PCA)

Principal component analysis (PCA) is a non-parametric method that estimates the proportion of the variance of a set of variables through some linear combination of these variables with minimum loss of information.[2 86] EIGENSTRAT is a popular implementation for transforming a number of possible correlated variables into a smaller number of uncorrelated variables (individual axes) known as principal components (PCs).[2; 284] A few simple graphs of PCA can be used to present a large fraction of the information contained in all genes tested. PCA is especially useful when there is considerable genetic exchange between close geographic neighbours.[86]

The principal component analysis (PCA) was performed for the allele diversity observed in the 75,957 SNPs genotyped in the 709 samples using smartpca, a program available in the EIGENSOFT package.[285] The statistical significance of each PC was evaluated through the Tracy-Widom statistics, computed by the EIGENSOFT tool twstats.

6.2.2. ADMIXTURE analysis

ADMIXTURE is a computationally fast maximum likelihood-based estimation approach suitable for inferring genome-wide ancestry at a large number of arbitrary SNP markers.[284 39] ADMIXTURE is able to accommodate the high number of markers that are incorporated currently in the chips, being more efficient than STRUCTURE.[284] As referred to before, the methods assumes that each individual's genotype data is drawn from multiple clusters, each cluster being defined by genotypic frequencies.

The 709 samples were run in the ADMIXTURE software.[284] I tested several numbers (from 3 to 6) of clusters or ancestral populations, *K*. The value of *K* was defined by applying a cross-validation, a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set, in this case, to check which *K* represents the most accurate population structure (the value of *K* with the lowest cross-validation error).

6.2.3. Estimating the date of admixture

Methods of local ancestry assessment, based on LD decay, infer ancestry at each locus in the genome and allow us to estimate times of admixture between two putative ancestor populations that are genetically 'close' to the ancestral mixing populations. ROLLOFF[142] measures the exponential decay of admixture-induced linkage disequilibrium (LD) as a function of genetic distance in a studied population by providing phased data of two putative parental populations; the date of admixture is calculated from weighted LD decay curves. According to Patterson et al.[286], ROLLOFF does not require a perfect local ancestry inference, which is difficult to ascertain for the shorter ancestry blocks resulting from older admixture events, and does not need accurate surrogates for the ancestral populations. So, ROLLOFF seems to be robust and to produce unbiased estimates for older events dating back up to 500 generations, as shown by simulations.[286]

In order to estimate the dates of admixture events in the populations analysed which display statistical evidence of admixture (from the results of PCA and ADMIXTURE), I therefore used the ROLLOFF method[142] implemented in the ADMIXTOOLS software package.[286] I ran the ROLLOFF method for the Arabian, Near Eastern and East African populations, using the unpruned set of 478,931 SNPs (as the method measures LD decay, all common SNPs must be considered), and by assuming Yoruba (African) and Italy plus Spain (European) as ancestral populations (data extracted from the 1000-Genomes database; http://browser.1000genomes.org/index.html). An East African ancestral population would be more suitable as putative ancestor for evaluating admixture in Arabia than the western African Yoruba population; this choice was due to the need for the ancestral population data be phased (haplotypes inferred), and 1000-Genomes provides currently the most reliable phased data available. There are two African populations genotyped in 1000-Genomes, the Yoruba from Nigeria and the Luhya from Kenya, but both share a West African Bantu ancestry, as demonstrated by my colleague Petr Triska at IPATIMUP (personal communication), who is studying in detail the Sahel region. For the European ancestral component we decided to use the Italian and Spanish samples and not the UK or Finnish because these two last samples showed some genetic drift in a preliminary analysis.

The correlation between SNPs was plotted as a function of the genetic distance for all chromosomes, and dates (in number of generations) were estimated by fitting an exponential distribution to the decay of the correlation coefficients. A good fit of the curve will lead to narrow confidence intervals of dates. The date (in number of

generations) and standard error for the admixture event is the average of dates for all chromosomes. The fitting of the curve was done in R, a language for statistical computing (including linear and nonlinear modelling, classical statistical tests, time-series analysis, classification and clustering) and graphics (well-designed publication-quality plots), which is very useful for exploring large data sets.

V. RESULTS

7. The first maternal descendants of the out-of-Africa migration

I reconstructed the complete mtDNA phylogeny for the three basal N subclades from 85 new sequences fully characterised by me (Table S1) and 300 previously published ones (Table S2). Data for haplogroup frequencies and diversity measures used in the interpolation maps are reported on Tables S3 and S4, respectively, and for the founder analyses with *f1* and *f2* criteria are displayed in Tables S5 and S6, respectively.

Figure 28 presents an outline topology of the total tree, indicating the primary branches with ages scaled against the ML estimates; detailed sections of the tree, with all age estimates (mean and 95% confidence interval; ML and rho for complete and synonymous diversities) are depicted in the following figures. The Age estimated for the haplogroup N in western Eurasia (excluding the highly frequent macro-haplogroup R) was ~60 ka, compatible with an age for the beginning of the Out-of-Africa migration in western Eurasia considering that at least by 50 ka modern humans were already present in Island Southeast Asia.[287] Nonetheless, this age estimate is quite close to the upper bound for the emergence of the ancestral haplogroup L3 in East Africa, estimated previously by us as having occurred at around 70 ka.[99] This seems to indicate that the date of origin of the N descendents was very close to the exit from East Africa.

7.1. Haplogroup N1

Analysing now in detail each of these branches, we can observe that haplogroup N1, is the oldest N(xR) clade at ~50–63 ka (Figure 29) and according to the structure of the tree it probably originated in Southwest Asia. This haplogroup has an early split in two branches: N1a′c′d′e′I comprising a series of subclades (N1a, N1c, N1d, N1e and I); and N1b.

The N1a′c′d′e′I branch is found across Southwest Asia (including the Near East and Arabia) and Europe and dates to 46–57 ka. As the most basal lineages within its subclades are restricted to Southwest Asia, this was its most probable place of origin. Within N1a′c′d′e′I, there is a series of consecutive splits. First, haplogroup N1c is originated, being most frequent in Southwest Asia, especially in Arabia and dating to ~47 ka (Figure 30D). Then, N1d detected so far only in India, splits from the cluster N1a'e'I which has a TMRCA of 42ka. Haplogroups N1a (dating to ~20

ka; Figure 31) and N1c are more frequent in Arabia and eastern Africa (Figure 30B and D). There are several deeply rooted Ethiopian, Somali, and Yemeni lineages within N1a, sharing the fast-evolving mutation at position 152 indicating that this subclade may have split before the N1a clade. But given the diversity of the subclade (dating to ~15 ka, compared to the age of N1a of ~20 ka) it is more likely that its split occurred within N1a (referred to as N1a2). Thus, N1a samples display deep diversity within eastern Africa and the southern part of the Arabian Peninsula, probably reflecting ancient gene flow (most probable during the Late Glacial period) across the Red Sea. N1a1 dating to the Late Glacial period (16–20 ka) is found at very low frequency across Europe. The founder analysis performed on HVS-I data for the haplogroup N1a indicated a Late Glacial entrance into Europe (with both *f1* and *f2* criteria), but the high frequency of some clades (N1a1a1 and N1a1a2) in central-European Neolithic burials may indicate that these clades might have re-expanded locally in the Neolithic period.

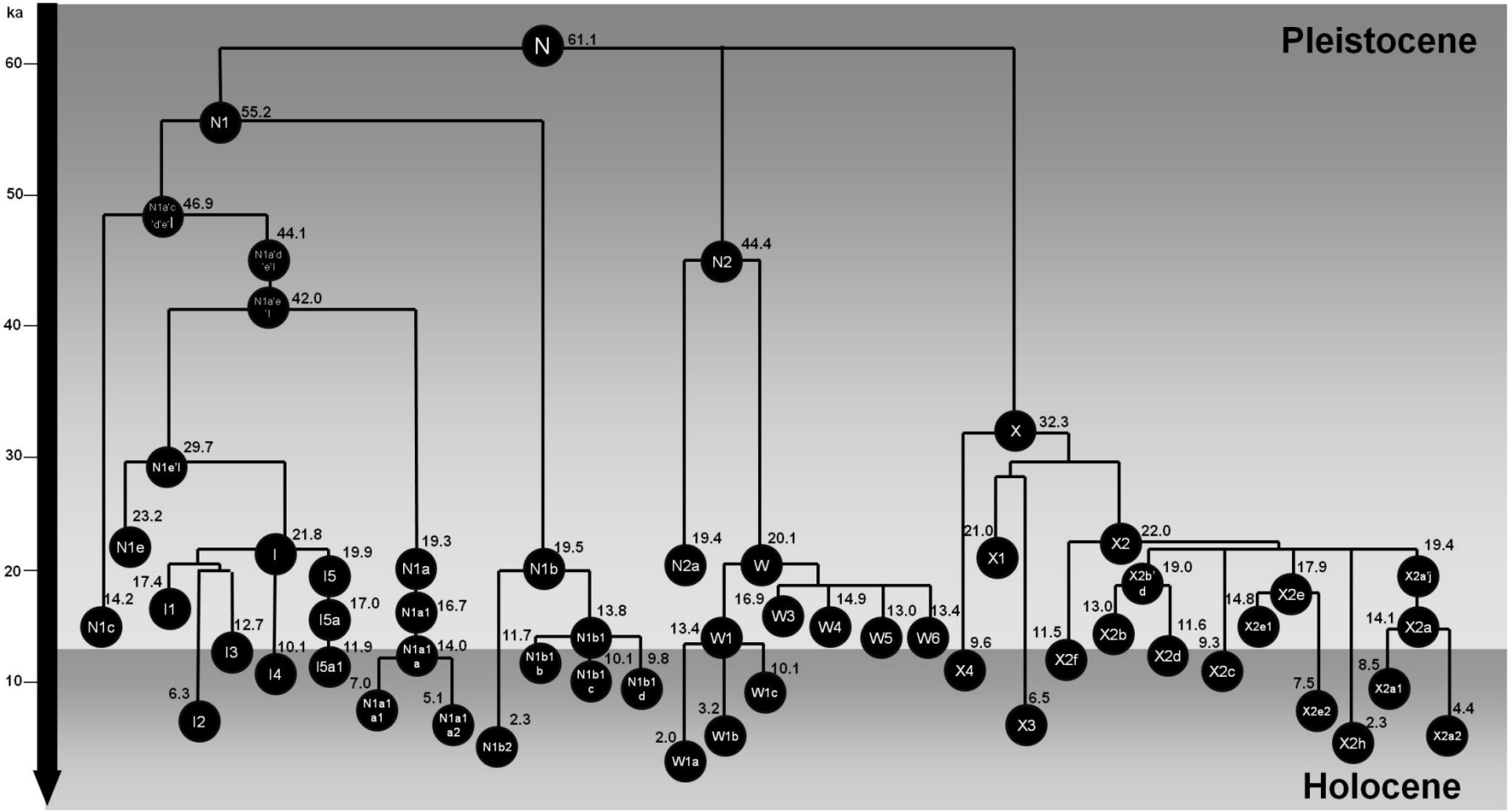

Figure 28 Schematic tree of haplogroup N (adapted from Fernandes et al.).[1]
Ages (in ka) indicated are maximum likelihood estimates obtained for the complete mtDNA genome.

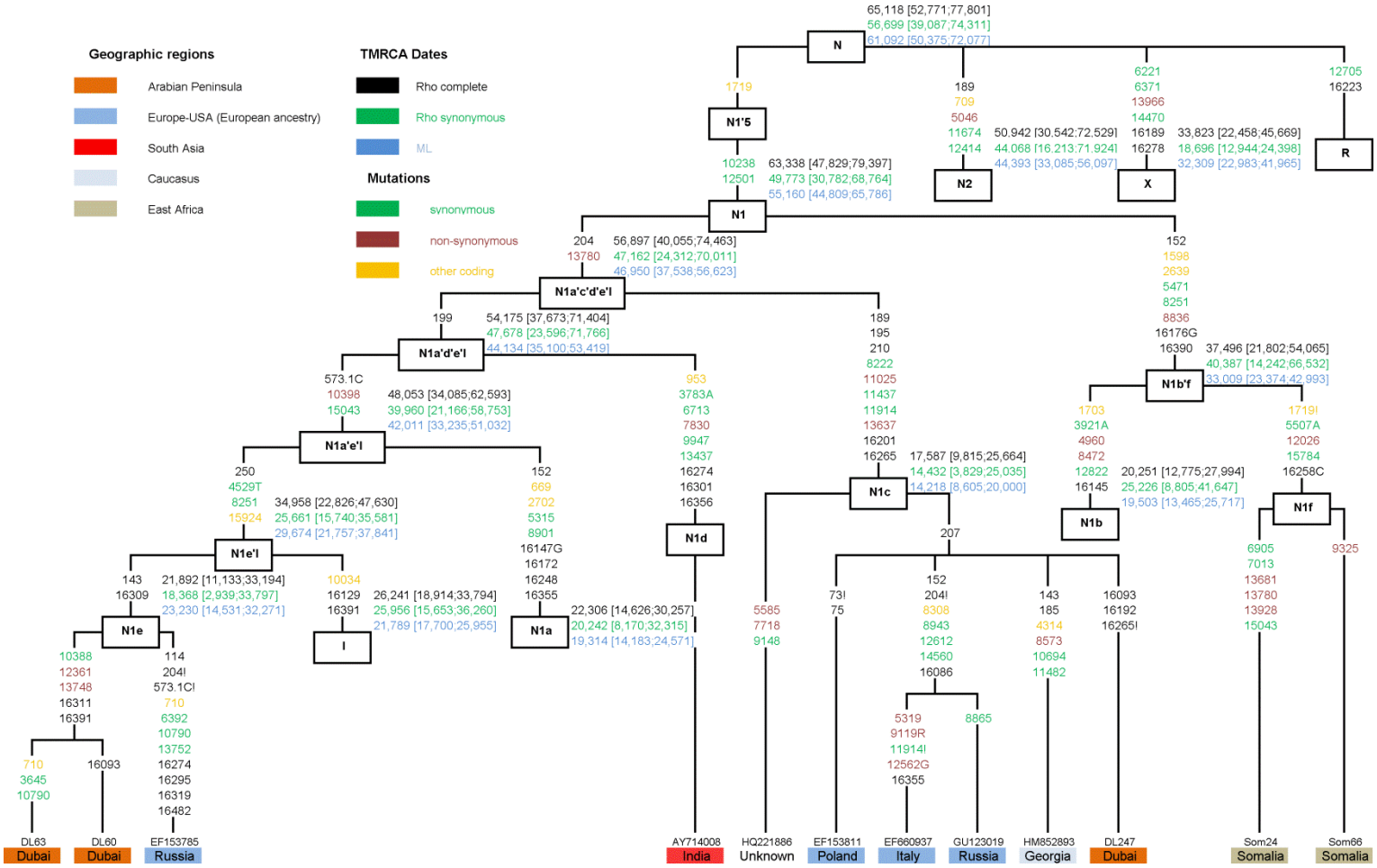

Figure 29 Phylogenetic tree of N(xR).
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

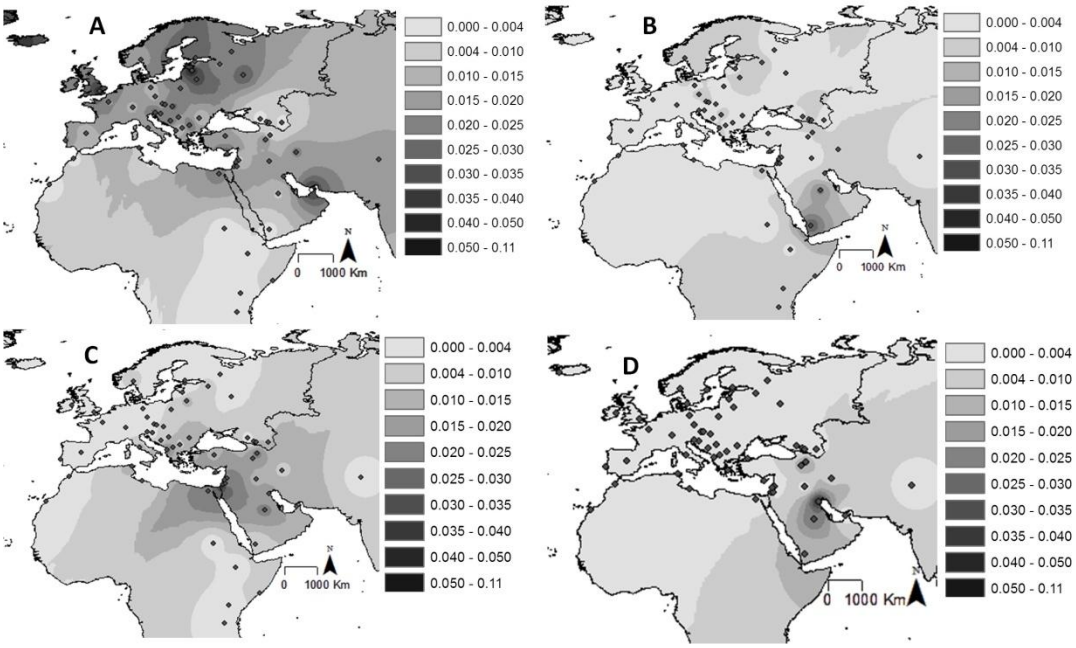

Figure 30 Frequency maps based on HVS-I data for haplogroups I (A), N1a (B), N1b (C) and N1c (D) (adapted from Fernandes et al.).[1]

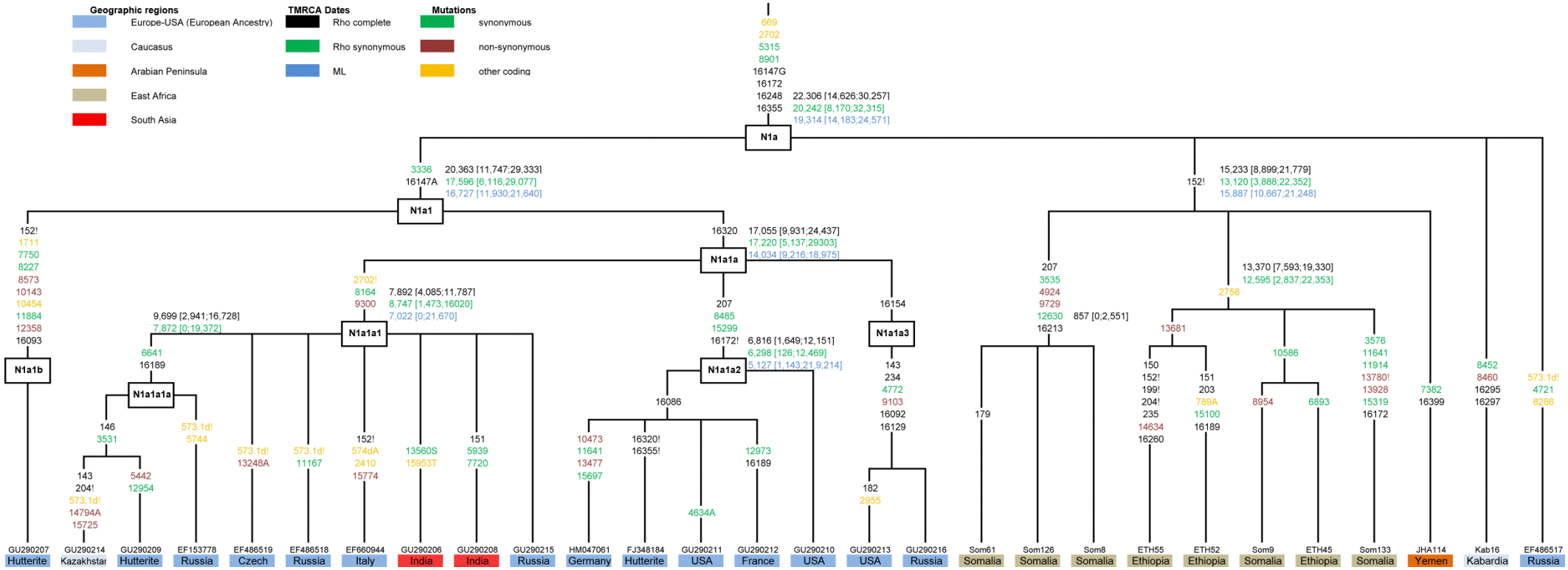

Figure 31 Phylogenetic tree of N1a.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

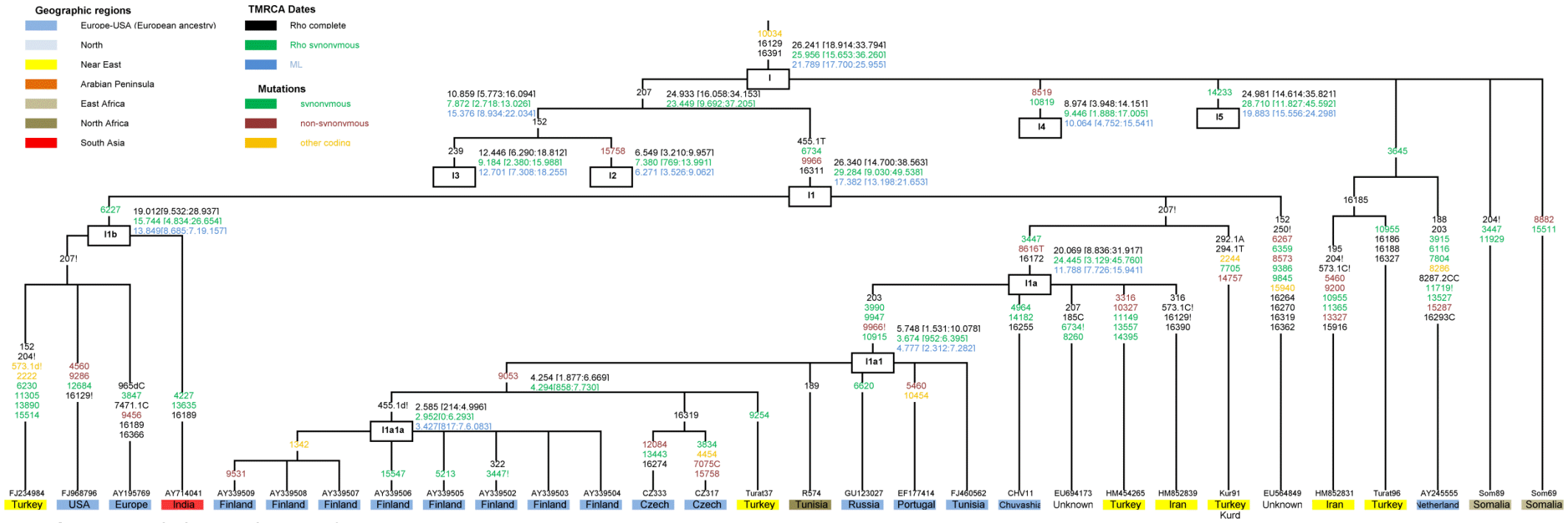

Figure 32 Phylogenetic tree of I1.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

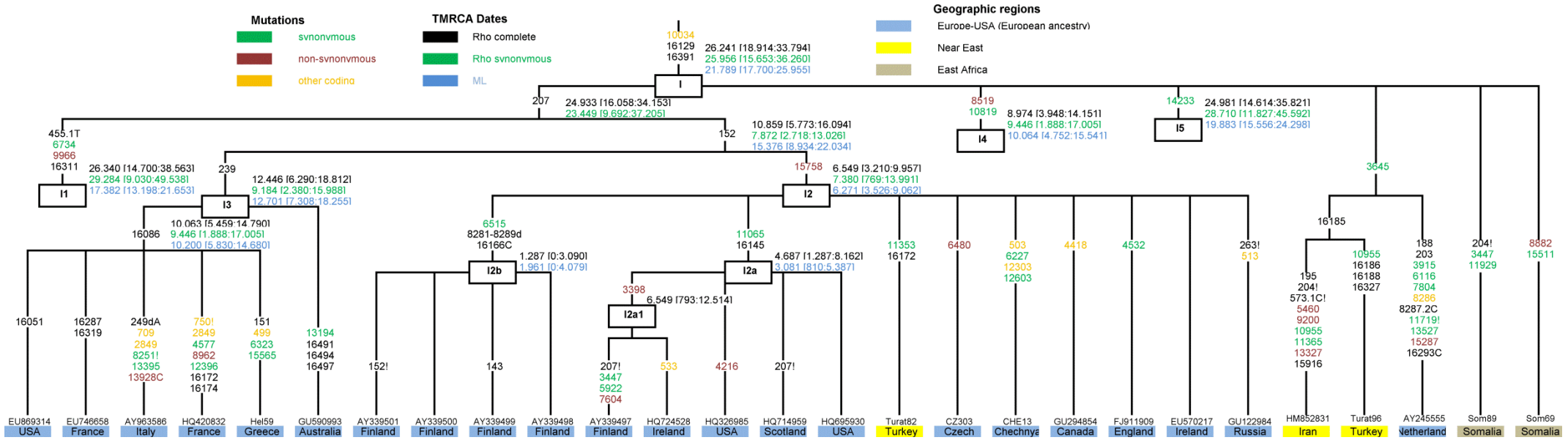

Figure 33 Phylogenetic tree of I2 and I3.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

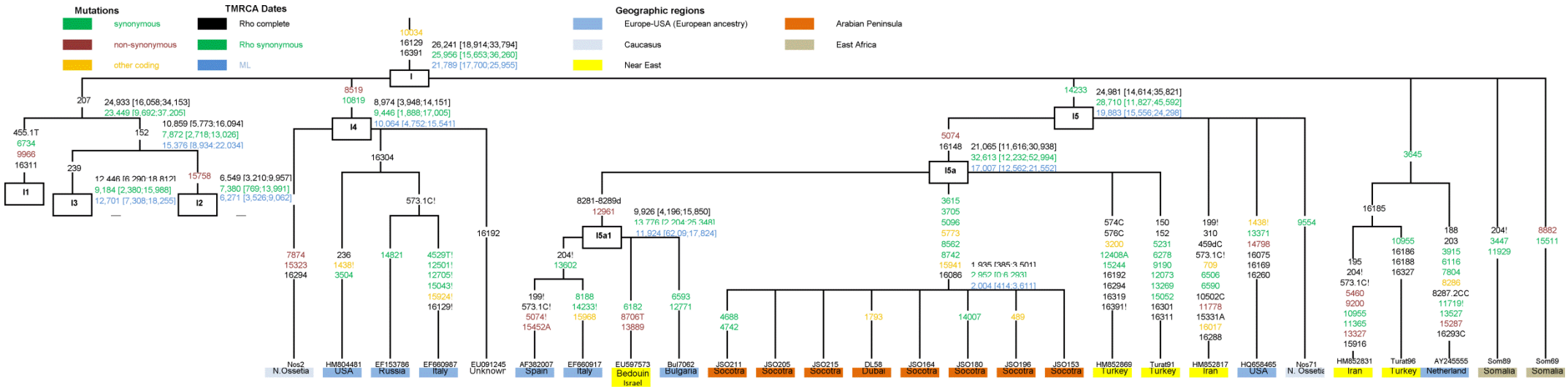

Figure 34 Phylogenetic tree of I4 and I5.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

N1e splits from haplogroup I at ~30 ka (Figure 29) and harbours three sequences in the tree located in the Arabian Peninsula and Russia. Finally, haplogroup I, dating ~25 ka, is overall most found in Europe (Figure 30A) and is by far the most frequent clade within N1. This haplogroup also has a frequency peak in the Gulf region and its highest diversity values are observed in the Gulf, Anatolia, and southeast Europe (Figure 35) suggesting a most likely origin in Near East and/or Arabia.

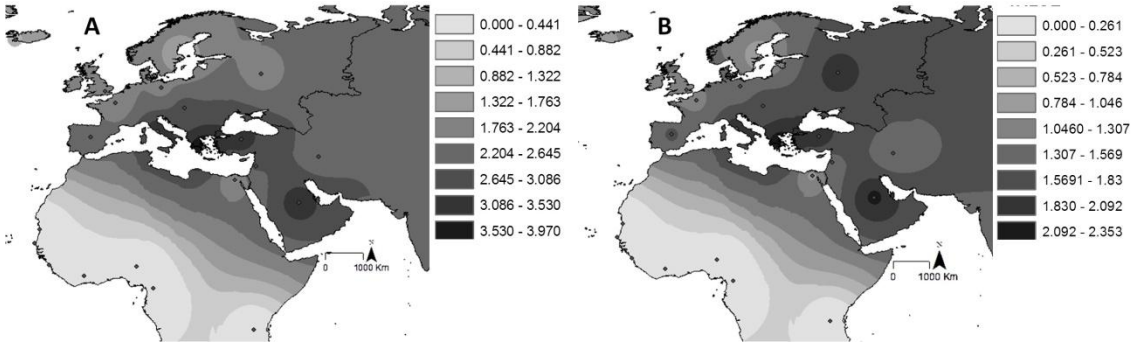

Figure 35 Distribution maps for haplogroup I for the diversity measures π (A) and ρ (B) based on HVS-I data (adapted from Fernandes et al.).[1]

A subhaplogroup of I, named I5a (Figure 34) shows founder effects signs on Socotra ~2 ka, an island found in the Gulf of Aden and which was settled during the Holocene.[254] I4 (Figure 34) and I2′I3 (Figure 33), dating to 10–15 ka ago, are both predominantly European. Looking to the HVS-I founder analysis, haplogroup I indicates a primarily peak during the Late Glacial period, but the I1a subclade peaks in the Neolithic period at ~6 ka ago when using both *f1* and *f2* criteria. This pattern is confirmed by the complete sequence tree and the I1a1, with an ML age estimation of ~5 ka ago, indicates a probable Near Eastern expansion.

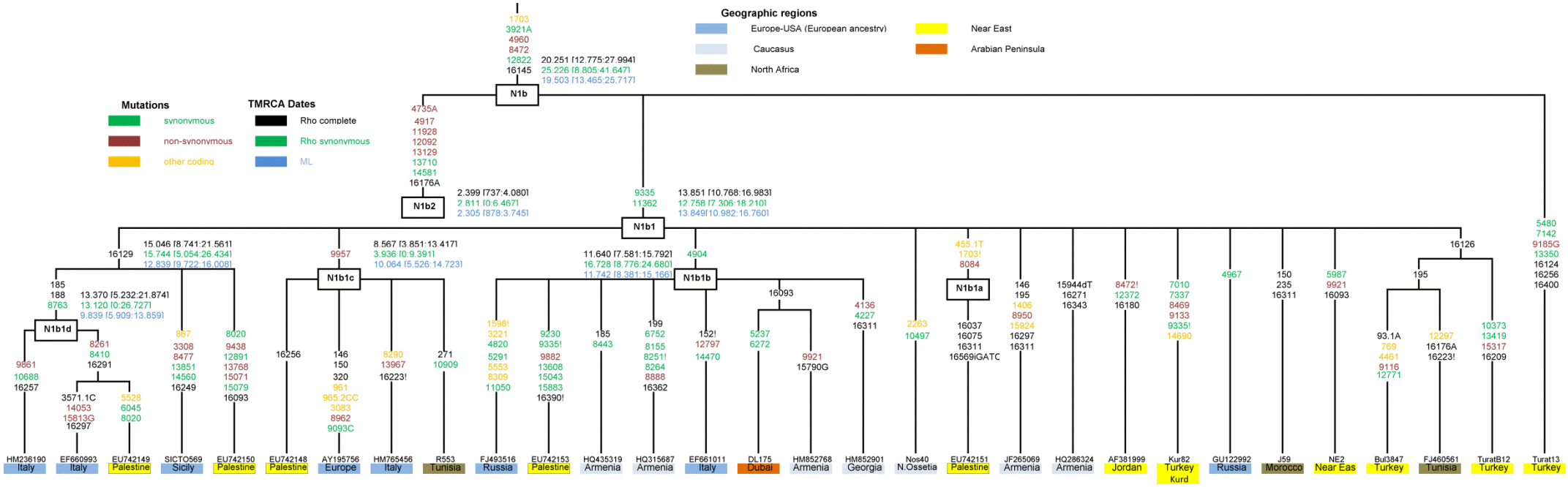

Figure 36 Phylogenetic tree N1b1**.**
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

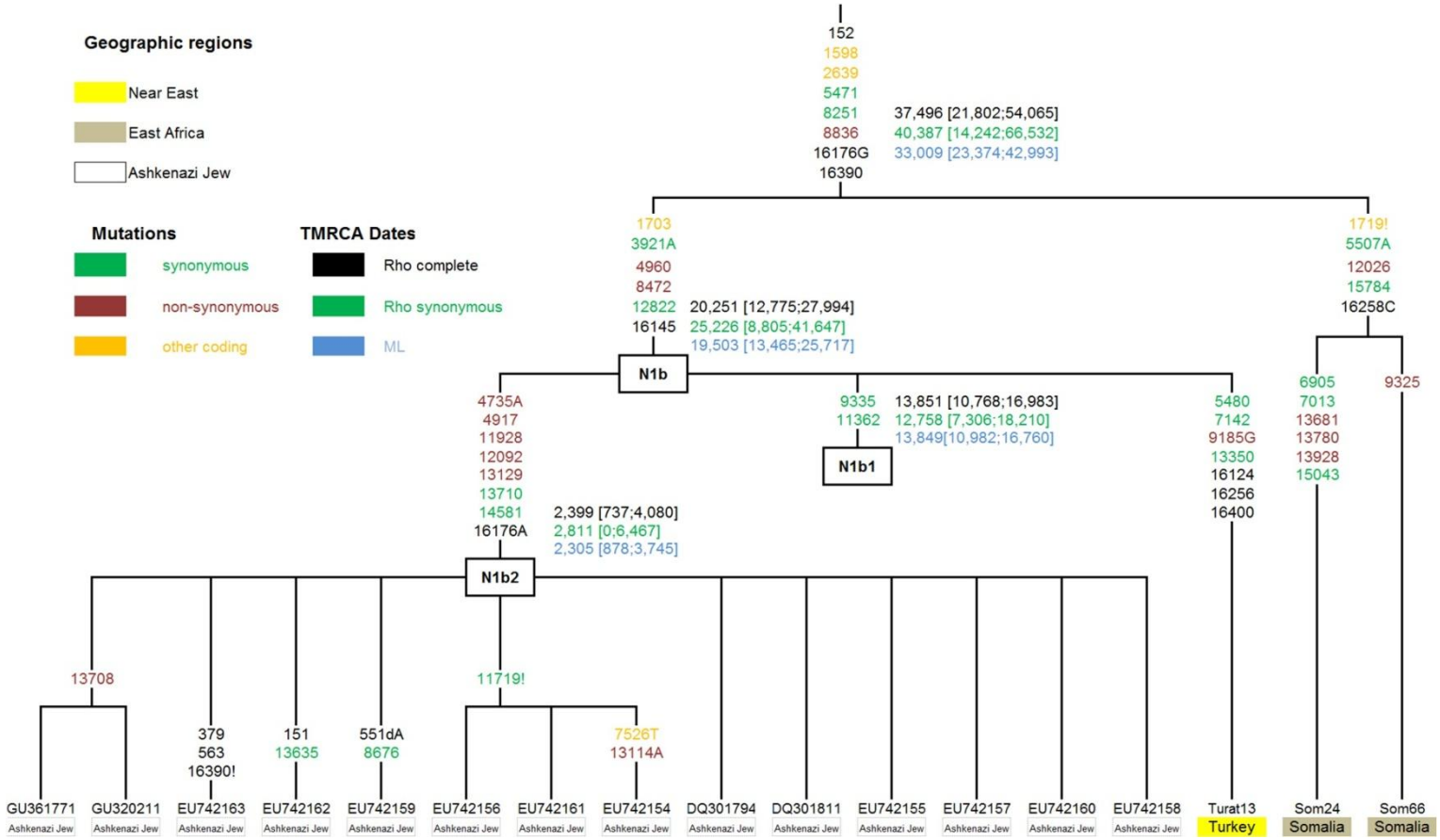

Figure 37 Phylogenetic tree N1b2.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

The deeper clade N1b′f, dating to 33–40 ka, comprises the haplogroup N1b (Figure 36; dates to 19–25 ka), primarily found in Southwest Asia (Figure 30C) and a clade with two Somali samples, sharing eight out of 15 polymorphisms with N1b, which I labeled as N1f. Given the N1b distribution (and N(xR) more generally), N1f probably resulted from ancient gene flow between Arabia and the Horn of Africa.

N1b itself has three basal branches: N1b1, N1b2, and N1b3 (represented by a single sample from Anatolia). N1b1 (Figure 36) is found in the Near East, Europe (rarely; mainly in central and eastern Mediterranean Europe), Arabia, and North Africa and dates to ~13–14 ka. Finally, N1b2 (Figure 37) is found mainly in Ashkenazi Jews, and its estimated age of ~2 ka indicates a recent founder effect among Ashkenazi ancestors. HVS-I founder ages indicate a primarily Neolithic expansion into Europe at ~8 ka, using an *f1* criterion. When using an *f2* criterion a peak at ~15 Ka indicate a Late Glacial expansion. This last scenario is perhaps more plausible given the HVS-I network (Figure 38) and suggests that N1b2 evolved within Europe. The predominantly European subclades, which date to ~10–13 ka ago, indicate that the expansion was indeed most likely pre-Neolithic.

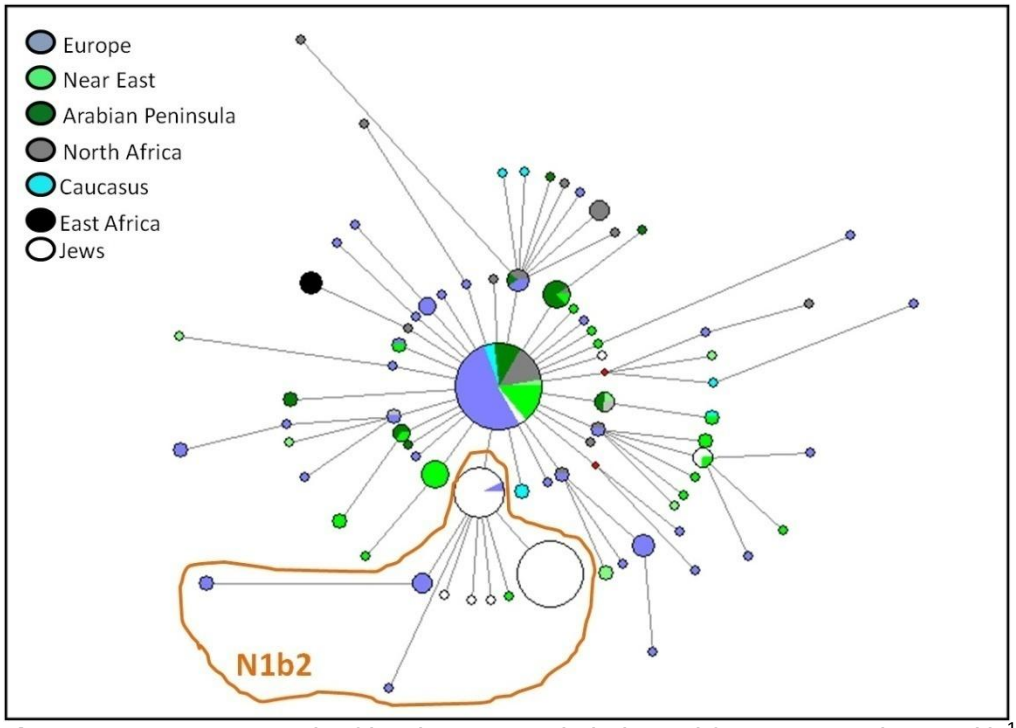

Figure 38 HVS-I network of haplogroup N1b (adapted from Fernandes et al.).[1] The areas of the circles are proportional to the frequency in the sample and the colours indicate the geographic origin according to the legend. The smallest circles are singletons. Branches are proportional to the number of polymorphisms.

Quite recently, our group, by focusing on the founder lineages of the Ashkenazi Jewish community,[288] provided further evidence on this issue. We completely

sequenced two Italian N1b2 lineages, revealing them to be ancestral to the Ashkenazi clade, raising the coalescence age for the newly defined, deeper N1b2 to 5.0 ka, and supporting the presence of this haplogroup in Europe since at least the Neolithic. It is thus more probable that N1b2 was introduced within Europe into the Ashkenazi community via females. This observation for N1b2, which makes ~9% of the Ashkenazi maternal pool, is corroborated by other European lineages that together sum up to ~80% of the Ashkenazi mtDNA pool.

The BSP of N1 (Figure 39) points to sequential Neolithic and Late Glacial expansions (observed in haplogroups I, N1a and N1b) with increments between 1.5 and 6.1 ka ago and between 8.8 and 13.4 ka ago (Table 6).

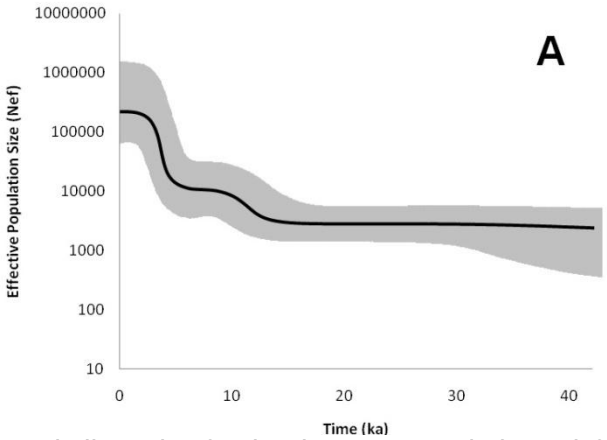

Figure 39 Bayesian skyline plot for haplogroup N1 (adapted from Fernandes et al.).[1]
The hypothetical effective population size variation through time is inferred.

Table 6 Peaks of rate of population size change through time.
The peaks were obtained from the BSPs for the general and geographically defined data, and periods of time where the rate of population size increase was of at least one individual per 100 individuals in a period of 100 years. Increment rate corresponds to the number of times the effective population size increased during this period

Data	Peak (ka)	Range (ka)	Increment
Overall	5.6; 12.63	1.78-14.67	42.49
Near East/Caucasus/Arabian Peninsula	9.34	7.70- 14.21	16.85
Europe	2.94	0.78-5.29	5.32
	12.8	11.18- 14.72	2.71
N1	3.6	1.51-6.06	20.05
	11.29	11.77-13.37	3.8
N2	6.29	3.15-10.11	19.99
N2 (without W Finns)	7.81	4.59-11.22	3.7
X	6.36	4.49-8.30	5.71
	13.72	11.73-16.16	3.46

7.2. Haplogroup N2

N2 dates around 44–50 ka (Figure 41) and comprises two separate branches: the rare N2a and the much more frequent haplogroup W. The few complete sequences of N2a are from eastern Europe and the Caucasus, but the HVS-I database also indicates a minor presence in Iran, Arabia, and Ethiopia (Figure 44A). Haplogroup W dated to ~20 ka is widespread (Figure 44B) and reaches >10% frequency in some eastern-European populations, with a strong founder effect in Finland 2–3 ka ago, and also in the Black Sea region, but it is less common in the Near East and Arabia. The HVS-I network (Figure 40) suggests a probable European origin, given that Near Eastern lineages appear largely nested within primarily European clades.

Haplogroup W primarily splits into two main clades: the predominantly European branch named W1 (Figure 43), dating to ~13 ka; and a clade including W3'W6, all dating to the Late Glacial period. The major clades within W are not recognizable at the HVS-I level, so the phylogeographic inferences are just based on the complete sequence tree.

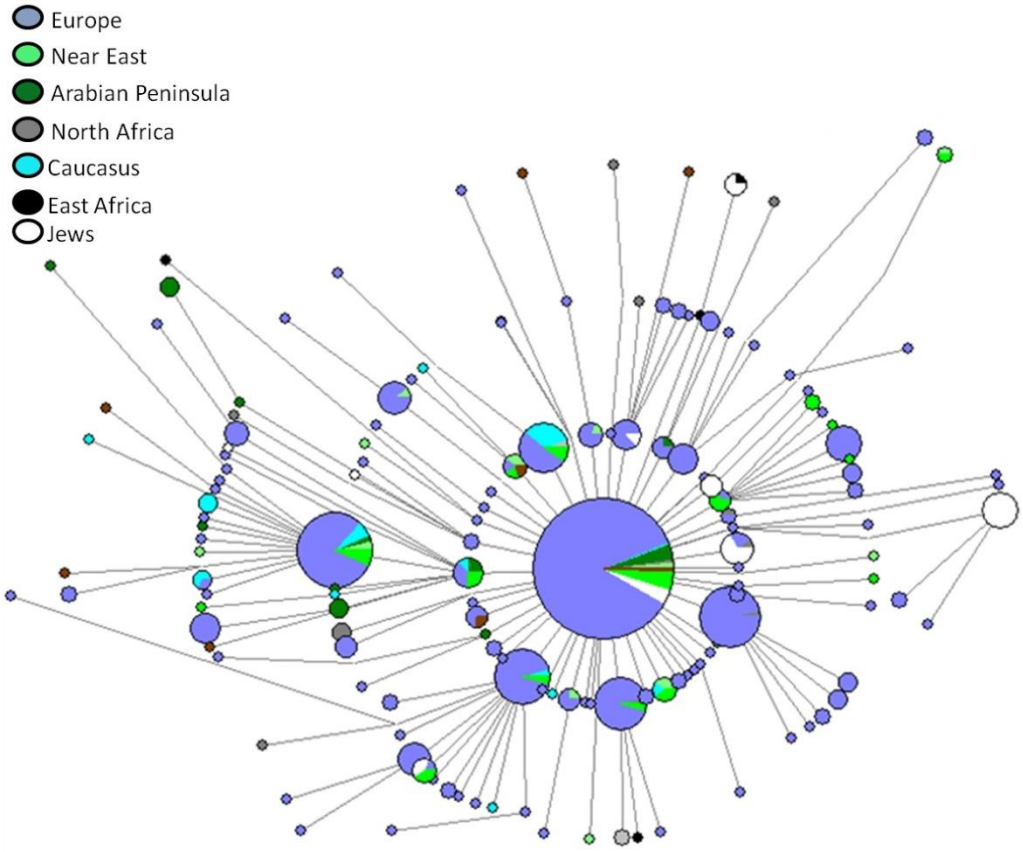

Figure 40 HVS-I network of haplogroup W.
The areas of the circles are proportional to the frequency in the sample and the colours indicate the geographic origin according to the legend. The smallest circles are singletons. Branches are proportional to the number of polymorphisms.

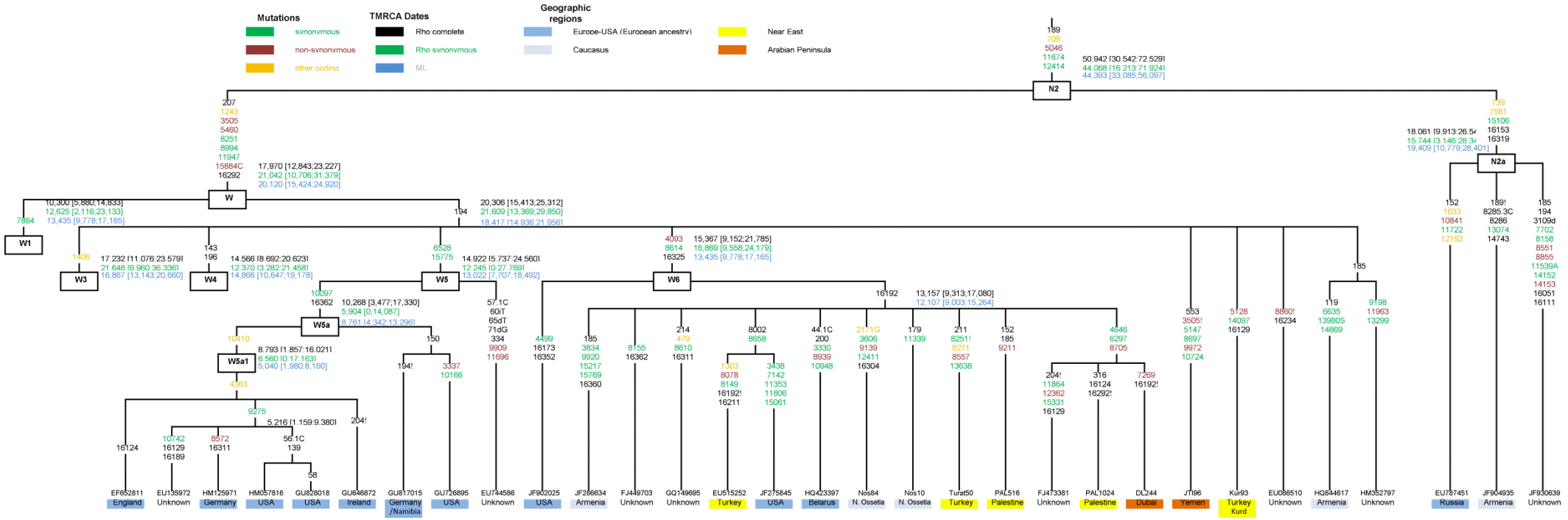

Figure 41 Phylogenetic tree of W5 and W6.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

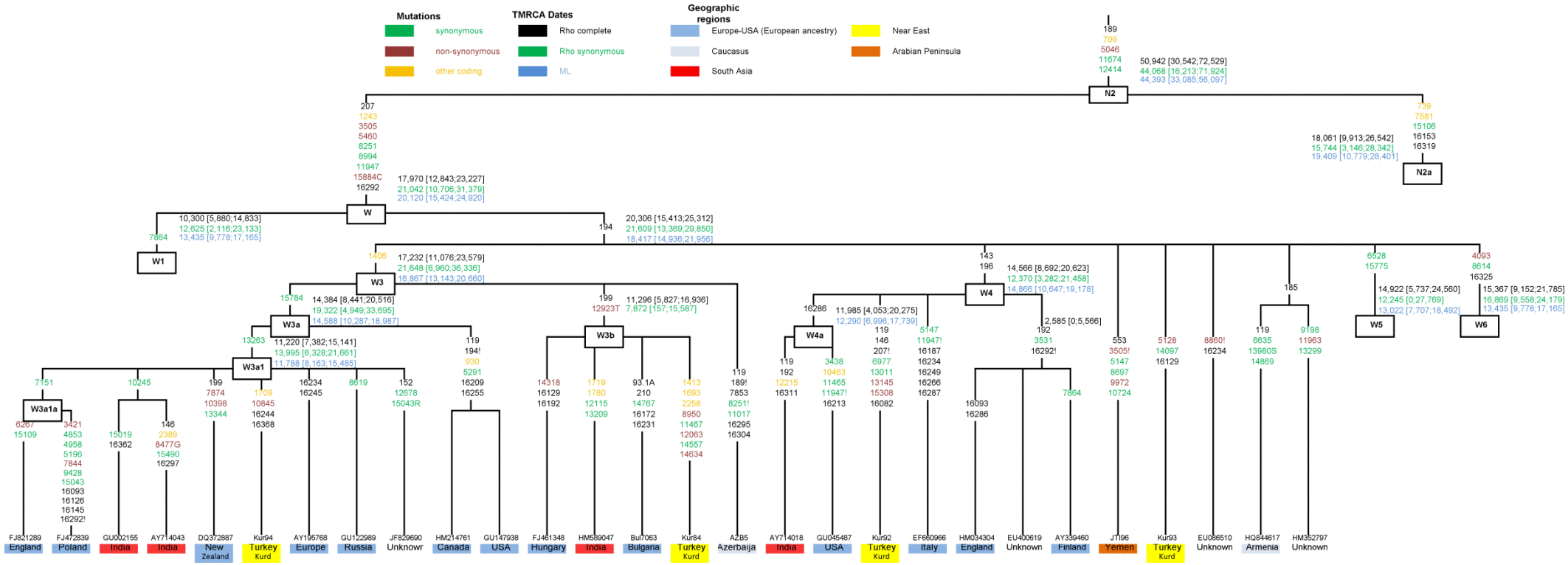

Figure 42 Phylogenetic tree of W3 and W4.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

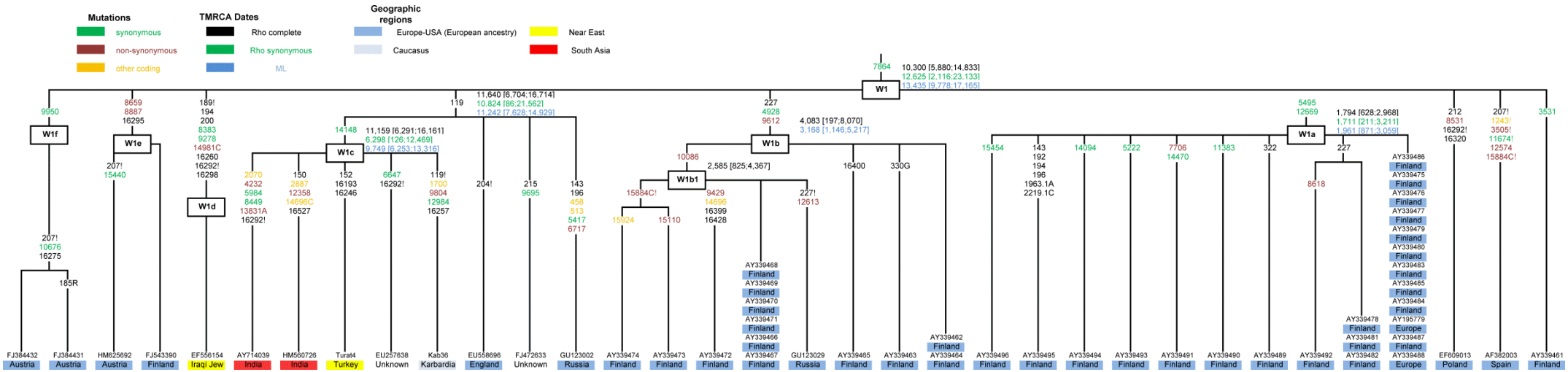

Figure 43 Phylogenetic tree of W1.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

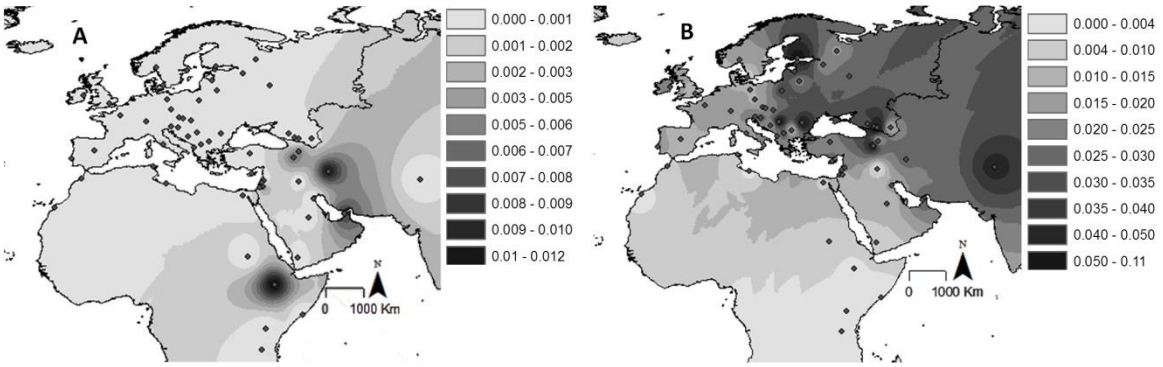

Figure 44 Frequency maps based on HVS-I Data for haplogroups N2a (A) and W (B) (adapted from Fernandes et al.).[1]

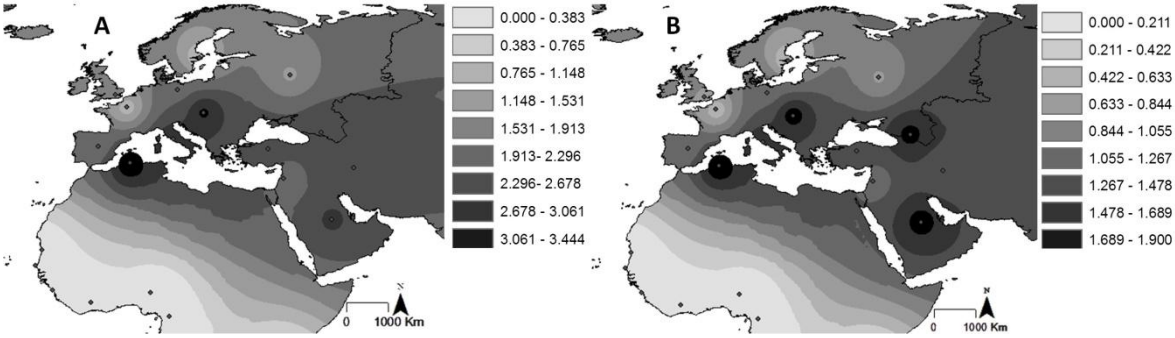

Figure 45 Distribution maps for haplogroup W for the diversity measures π (A) and ρ (B) based on HVS-I data (adapted from Fernandes et al.).[1]

The W3′6 subclade (Figure 41 and 42), includes some basal lineages from the Caucasus, Anatolia, and Yemen and dates to ~20 ka, suggesting in this case a centre of gravity in the Near East. The fact that W3, W4, and W6 have basal lineages in the South Caucasus and Anatolia points to these regions as the source of the European diversity. HVS-I founder ages of haplogroup W in Europe indicate peaks at 11.5 ka and at 14.8 ka when considering a *f1* and *f2* criterion, respectively. The possible founders (in W4, W6, W3a, and W3b) and the European W5 mostly date to ~13 ka ago, suggesting a Late Glacial expansion.

The BSP for N2 (Figure 46A) indicates an increment from 3–10 ka (Table 6). When considering the data without the highly drifted Finnish, (Figure 46B), the BSP shows a peak at ~8 ka (Table 6), suggesting that haplogroup W began its expansion during the Late Glacial period, and expanded locally within Europe during the Neolithic period.

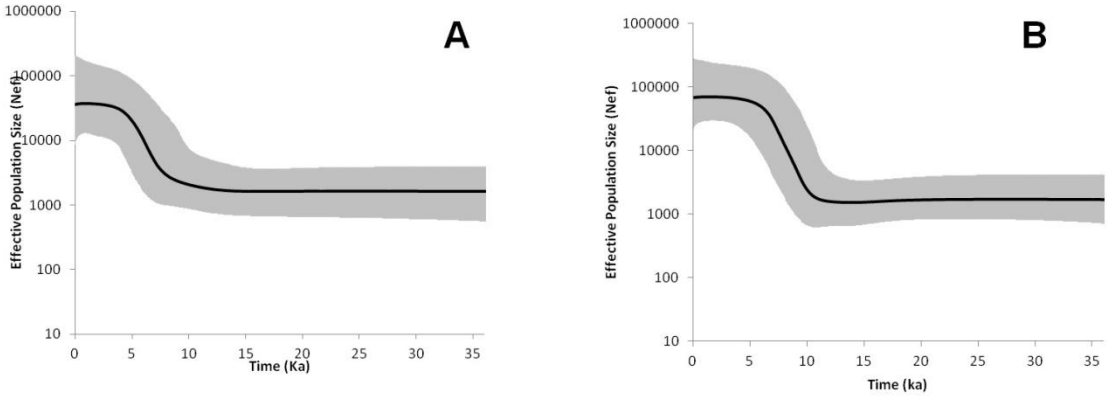

Figure 46 Bayesian skyline plot indicating hypothetical effective population size through time based on data from haplogroups N2 (A), and N2 without W Finns (B) (adapted from Fernandes et al.).[1]

7.3. Haplogroup X

Haplogroup X (Figure 47) dates to ~30 ka and is most frequent in the Near East (Figure 53). The diversity indices (Figure 54) of haplogroup X also clearly indicate the Near East as its point of origin. Haplogroup X primarily splits into two main clades: the very rare X4, with just two samples from Anatolia and Armenia; and a clade including X1'3 dating to ~29 ka. Haplogroup X1, dating to ~21ka, is largely restricted to the Near East and northeast Africa (from the HVS-I database). The haplogroup X3, is much younger (~6ka), and is not recognizable at the HVS-I level, but the complete sequence tree suggests a similar Mediterranean distribution.

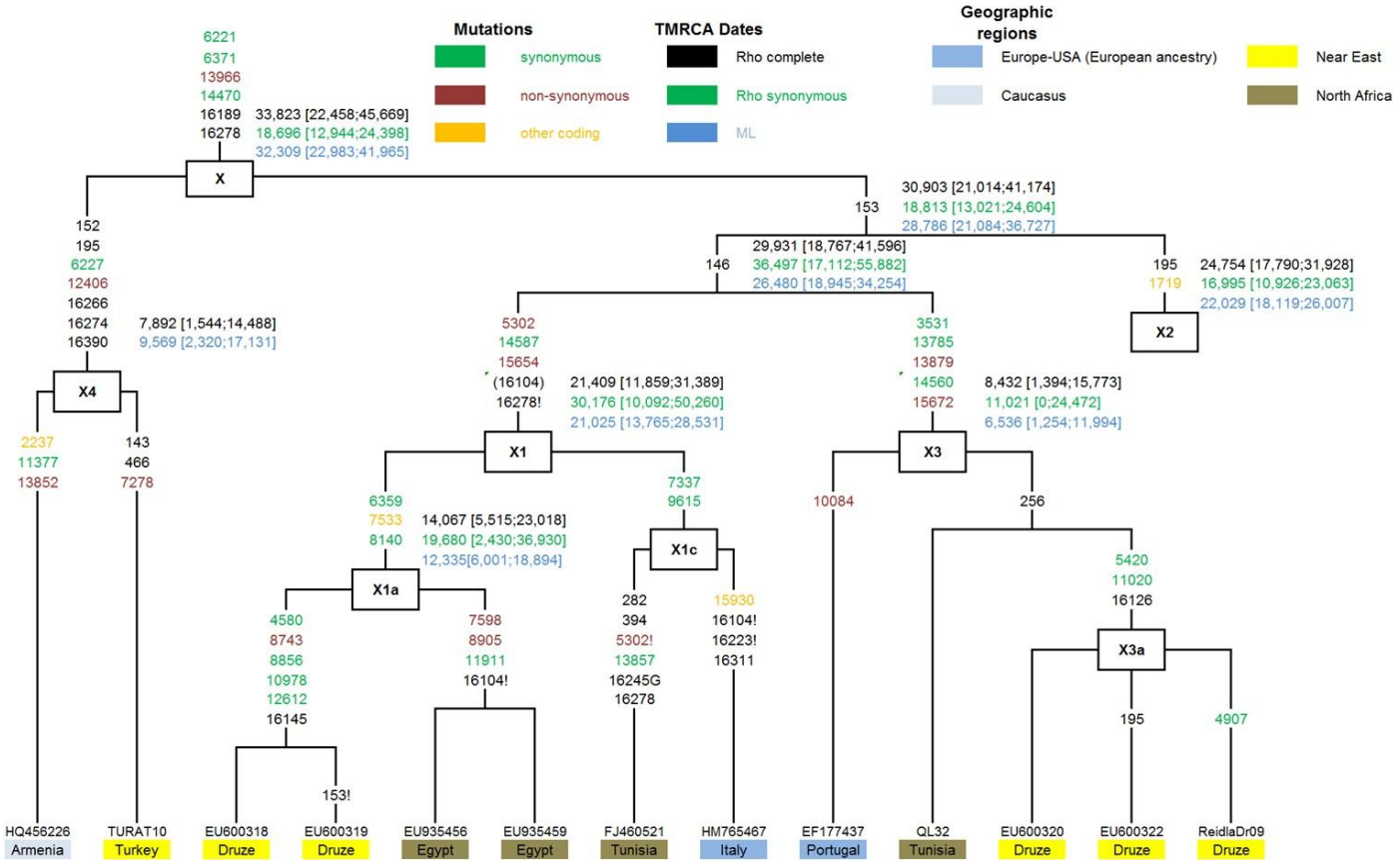

Figure 47 Phylogenetic tree of X1, X3 and X4.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

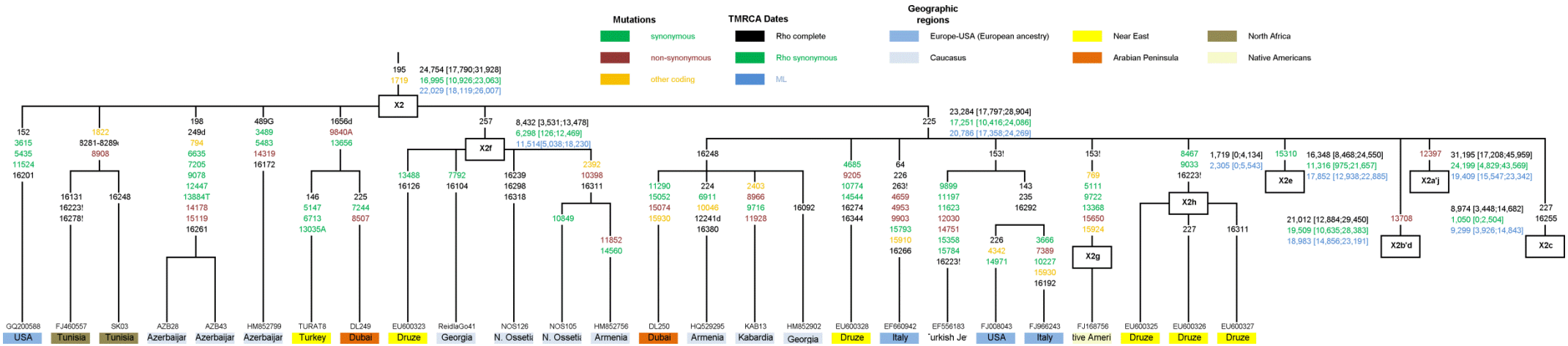

Figure 48 Phylogenetic tree of X2.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

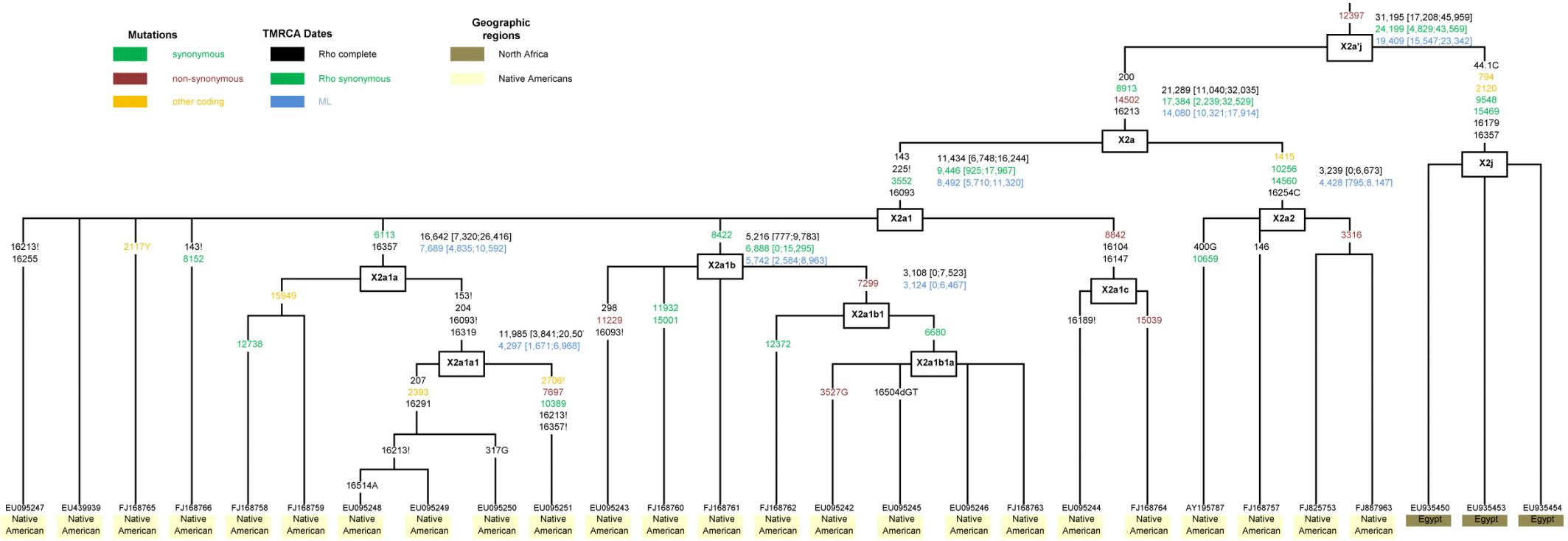

Figure 49 Phylogenetic tree of X2a'j.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

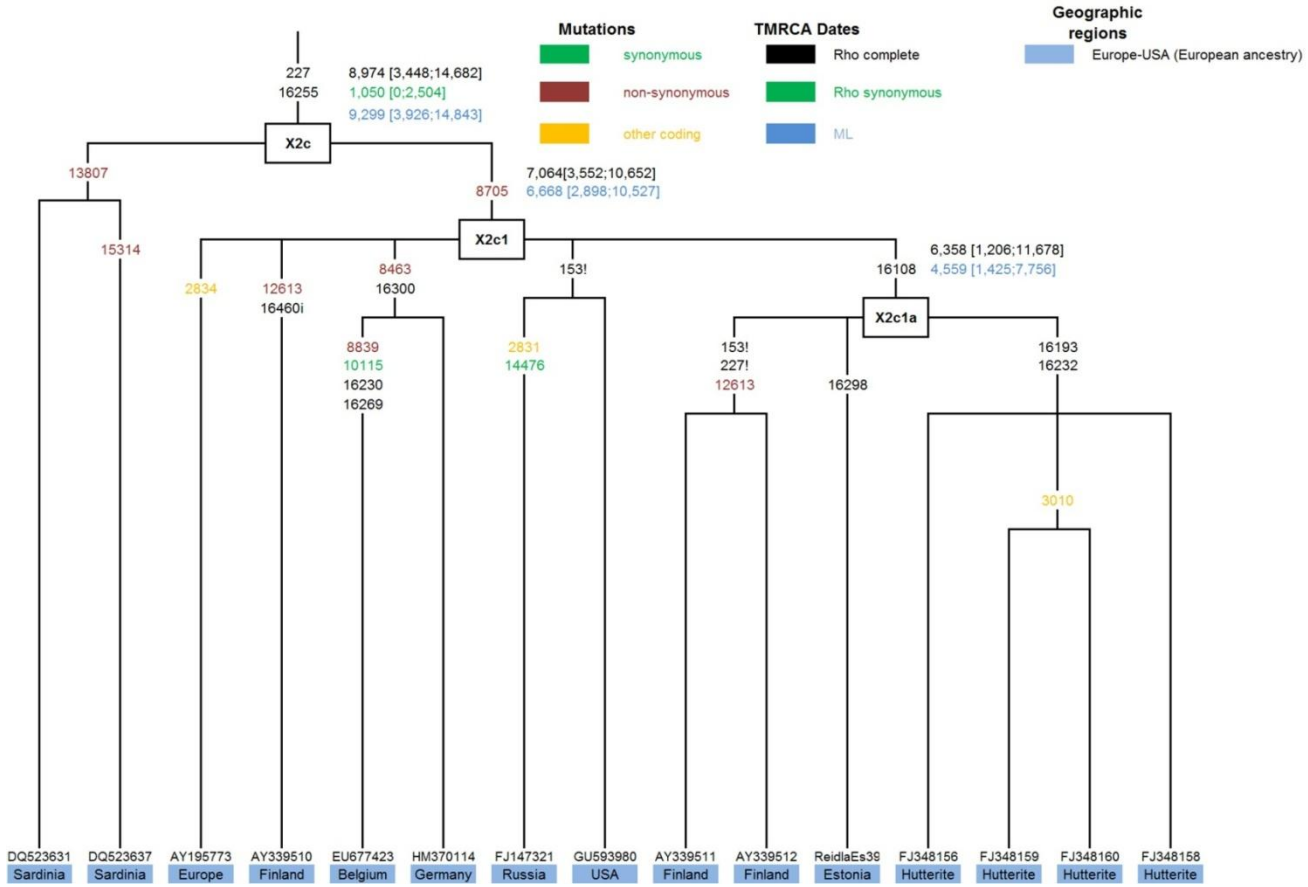

Figure 50 Phylogenetic tree of X2c.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

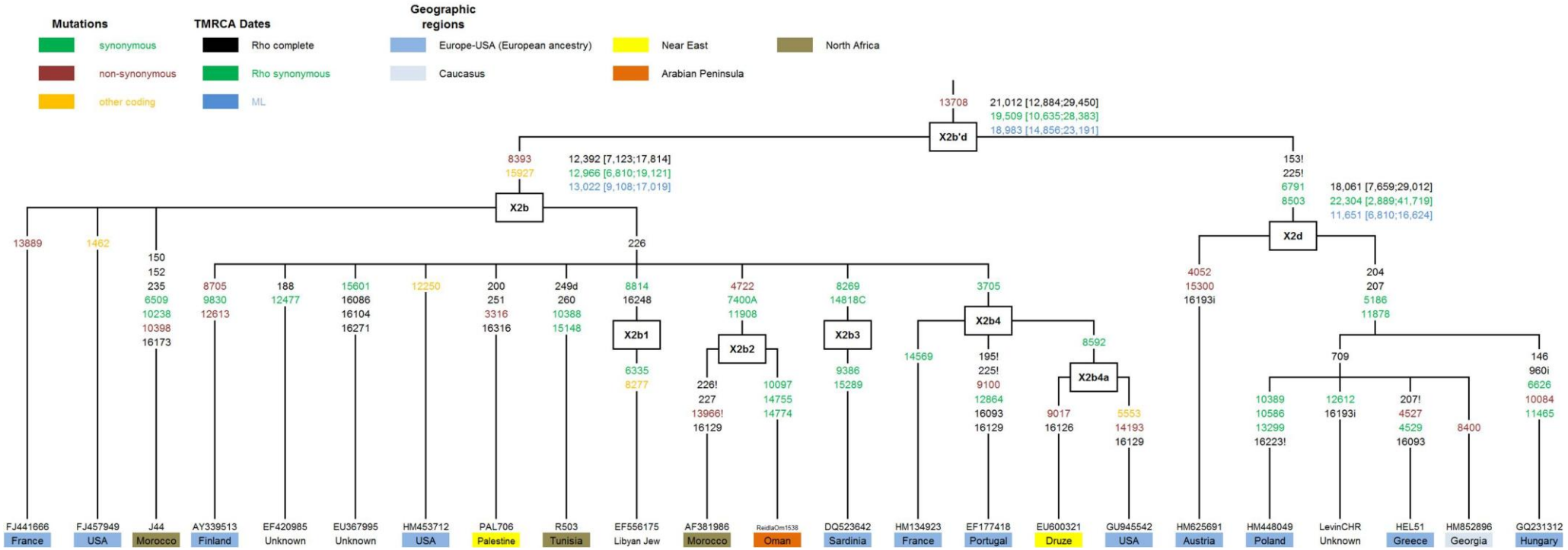

Figure 51 Phylogenetic tree of X2b'd.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

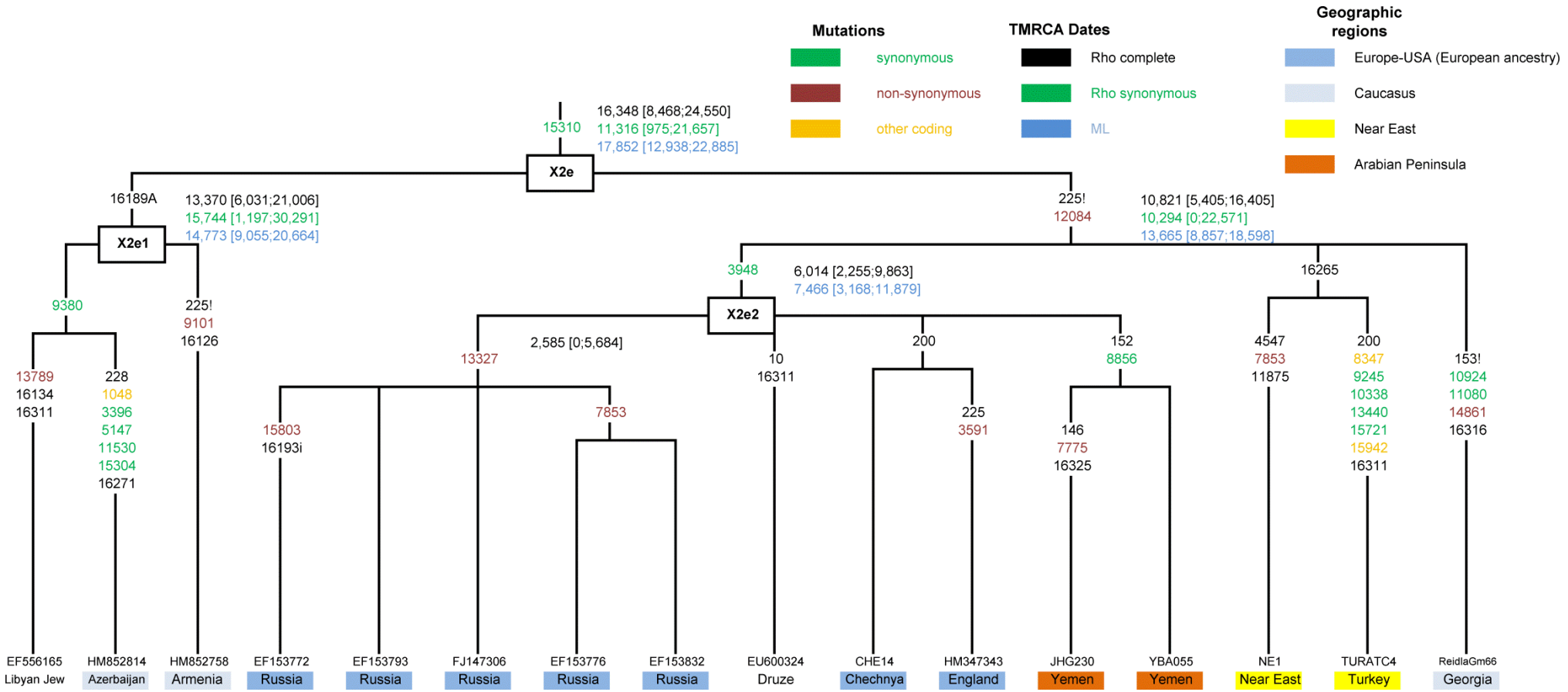

Figure 52 Phylogenetic tree of X2e.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

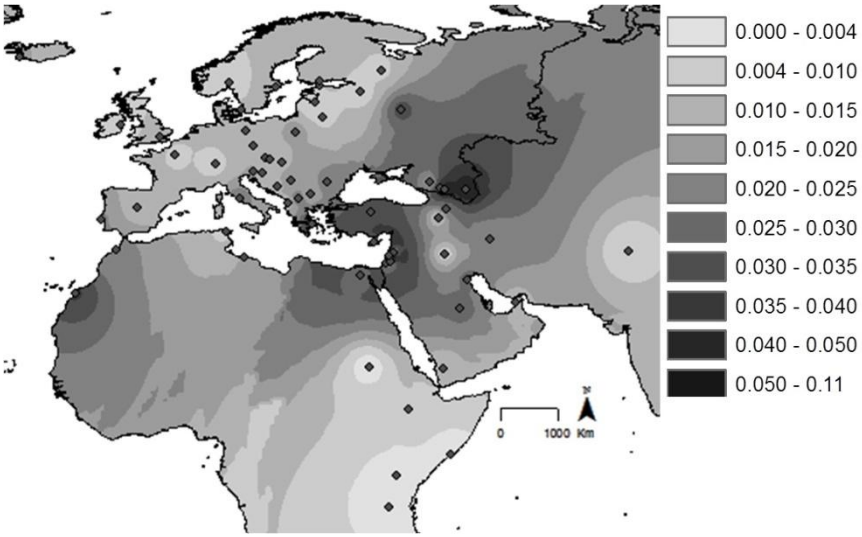

Figure 53 Frequency maps based on HVS-I data for haplogroup X (adapted from Fernandes et al.).[1]

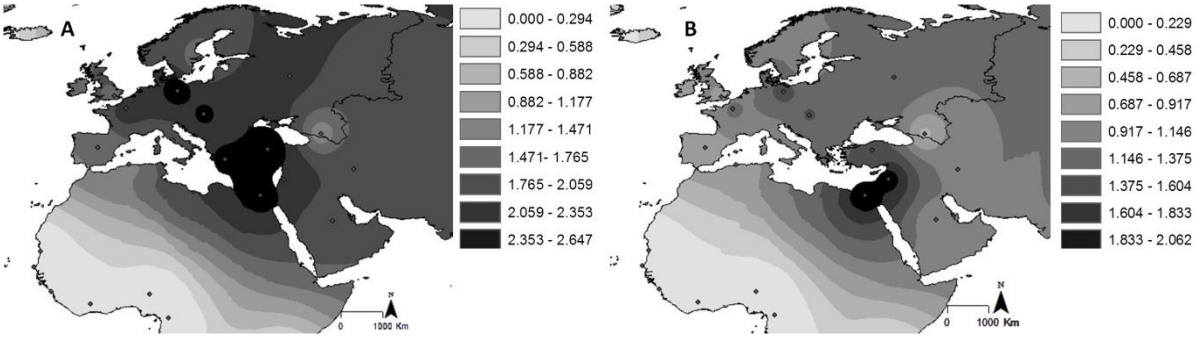

Figure 54 Distribution maps for haplogroup X for the diversity measures π (A) and ρ (B) based on HVS-I data (adapted from Fernandes et al.)**.**[1]

Haplogroup X2, dating to ~22ka, is by far the most frequent and widespread subclade. Its basal branches (including X2f) are largely restricted to the Near East, the Caucasus, and North Africa, except for the main branch defined by the polymorphism 225 which includes Near Eastern, North-African, and European-specific subclades, as well as a subclade (X2a) restricted to the Native Americans.

The root of X2+225 (Figure 48), dating to ~21 ka, is probably the major founder sequence for dispersals involving haplogroup X. A curious feature of the tree is the possible connection of X2a to the North-African clade X2j through a mutation at position 12397 (Figure 49). However, this mutation might be a recurrence; X2j appears to be extremely recent. The rare X2g (Figure 48), also found only in one Native American sample, indicates that the spread from the Near East toward the Americas could have begun as early as the emergence of the X2+225 clade, given that this could have been the only founder sequence. The X2b′d clade (Figure 51), dating to ~20 ka, and the haplogroup X2c (Figure 50), dating to ~9 ka, have most

likely an European origin, although the lower age may be indicating only an expansion within Europe. The founder sequence between the Near East and Europe is again the root of X2+225, suggesting the possibility that the expansion of this clade into Europe and toward America originated from the same initial process in the Near East. However, X2e2 (Figure 52), which dates to ~7 ka ago, might have spread to Europe during the Neolithic period.

HVS-I founder ages of haplogroup X in Europe indicate peaks at 15 ka, and at 16.5 ka when considering a *f1* and *f2* criterion, respectively; suggesting Late Glacial expansions of the haplogroup X (mostly X2+225). The BSP for X (Figure 55) indicates an increment also around the Late Glacial (Table 6) and an additional increment in the Neolithic period, as suggested by the analysis of clades X2c and X2e2.

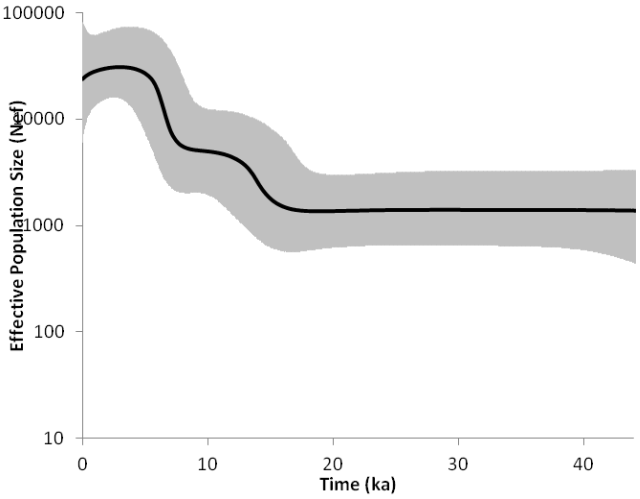

Figure 55 Bayesian skyline plot indicating hypothetical effective population size through time based on data from haplogroup X (adapted from Fernandes et al.).[1]

7.4. Overall pattern of migration and population expansion

The founder analysis for the overall data, by assuming a Near Eastern source and migrations into Europe, and using an extensive HVS-I database, led to the following results (Figure 56): (1) a poorly-defined peak ~12 ka with the *f1* criterion; and (2) two peaks, a small one dating to 6 ka and a larger one dating to 15 ka, using the *f2* criterion. These results suggest that the three clades are mostly related with Late Glacial or post-glacial expansions into Europe. We also performed a similar founder analysis of the three clades from the Near East into North Africa and obtained a single peak at 10.4 ka and 13.8 ka, using a *f1* and *f2* criterion respectively, also indicating a primarily post-glacial or Late Glacial expansion into this region.

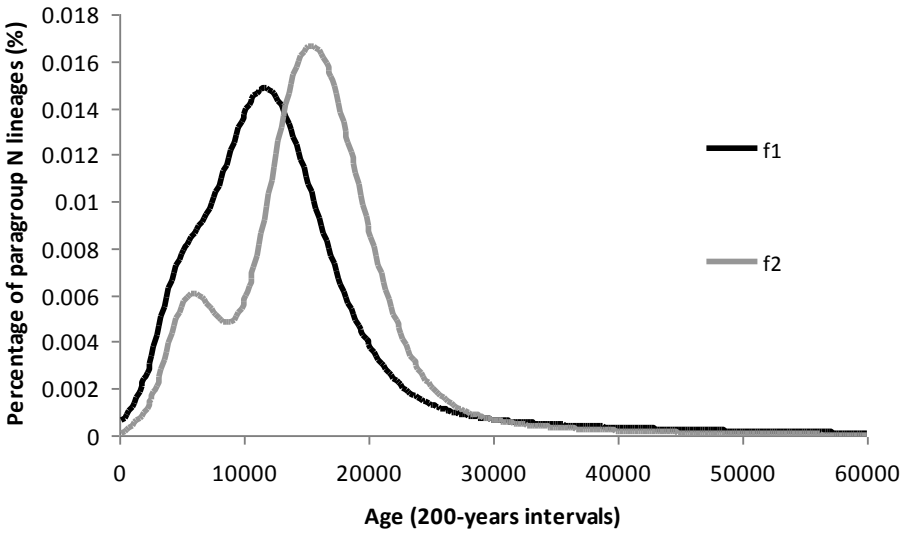

Figure 56 Probabilistic distribution of founder clusters across migration (adapted from Fernandes et al.).[1]
Times scanned at 200 yr intervals from 0–60 ka, using both an *f1* criterion (black line) and an *f2* criterion (grey line), when considering a migration from the Near East to Europe

The BSP obtained from the overall data (Figure 57A) shows continuous slightly stepped increase from ~15 ka almost to the present (Table 6). The BSP using only the European data (Figure 57B) separates two steps (~13 ka and ~6ka), while the BSP for the pool of the Near East and Arabian Peninsula (Figure 57C) indicates only a period of increase, initially gradual, from ~15 ka.

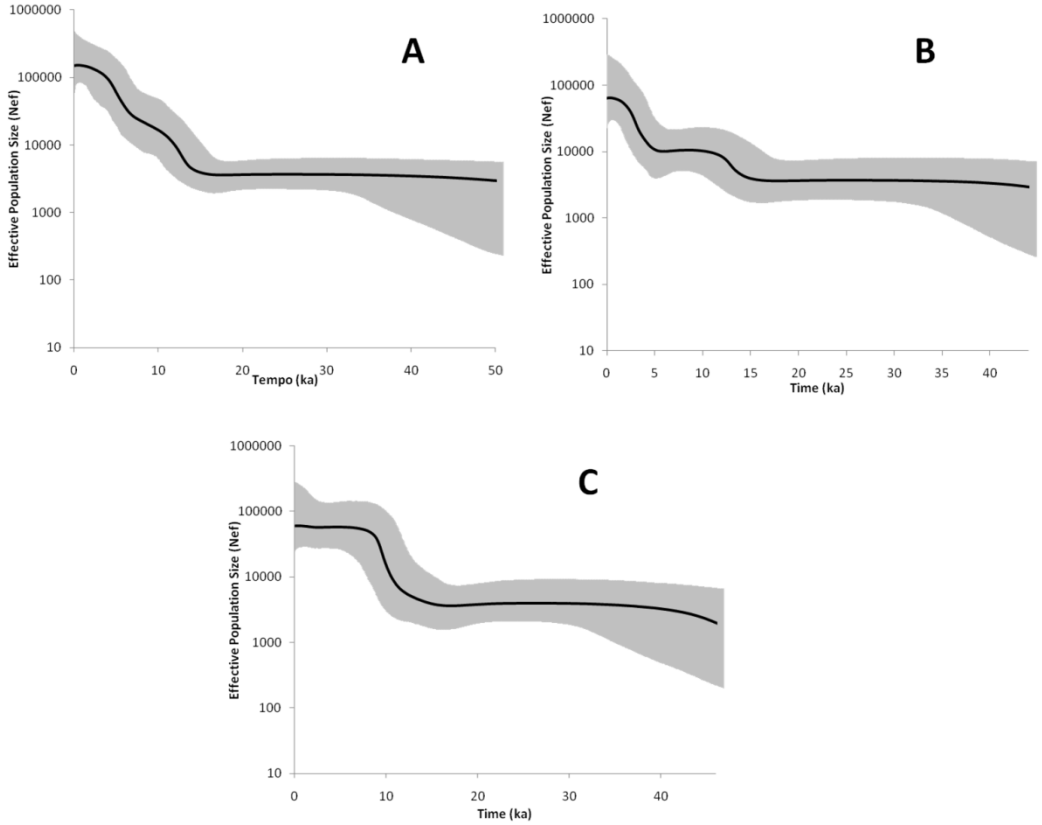

Figure 57 Bayesian skyline plot indicating hypothetical effective population size through time (adapted from Fernandes et al.).[1]
Bayesian skyline plot are based on data from the entire dataset (A), only Europe (B), and Near East, Arabian Peninsula and Caucasus (C).

8. Near Eastern and Pakistani maternal genetic influence into the Arabian Peninsula

Our group, have been previously engaged in fully characterising the diversity of the Arabian haplogroup R0a and related HV1, which together account for 16% of the Arabian pool.[102; 253; 257; 258; 260] The main aim of that characterisation was to add information to an important question in Arabian archaeological research, on population continuity *versus* replacement, since the time of the successful out-of-Africa migration across the south of the Peninsula. In this thesis, I tried to shed further light on this issue by performing complete mitochondrial sequencing of the second most frequent group of sequences in Arabia, the JT clade, and by performing founder analyses for the entire mtDNA pool in the region based on HVS-I diversity.

8.1. Insights from the complete sequencing of haplogroups JT

Last year, my colleague Joana Pereira[289] defended her PhD thesis at the University of Leeds, in which she performed a detailed JT analysis at a broad scale, mainly focused on the regions of the Near East, Europe and North Africa. She performed complete sequencing of 237 JT genomes and combined them with 1439 published sequences for the statistical analyses. From that large dataset, a total of 21 sequences are from Arabia, including 10 from Yemen, 10 from Dubai and one Saudi. In this thesis, I augmented the Arabian dataset of complete JT mtDNA lineages, by performing the characterisation of 44 more samples, 15 J and 10 T from Yemen, and 14 J and 5 T from Dubai (Table S1). The sequences used in the phylogenetic reconstruction are summed up in Table S7, and data for haplogroup frequencies and diversity measures used in the interpolation maps are reported in Tables S8 and S4, respectively.

In order to contextualise the sub-haplogroups in haplogroup JT which are present in Arabia, Figure 58 and 59 present an outline topology of the total tree, indicating the primary branches with ages scaled against the ML estimates according to Pereira;[289] detailed sections of the branches with Arabian samples, for which I recalculated the ages (mean and 95% confidence interval; ML and rho for complete and synonymous diversities) are depicted in the figures below.

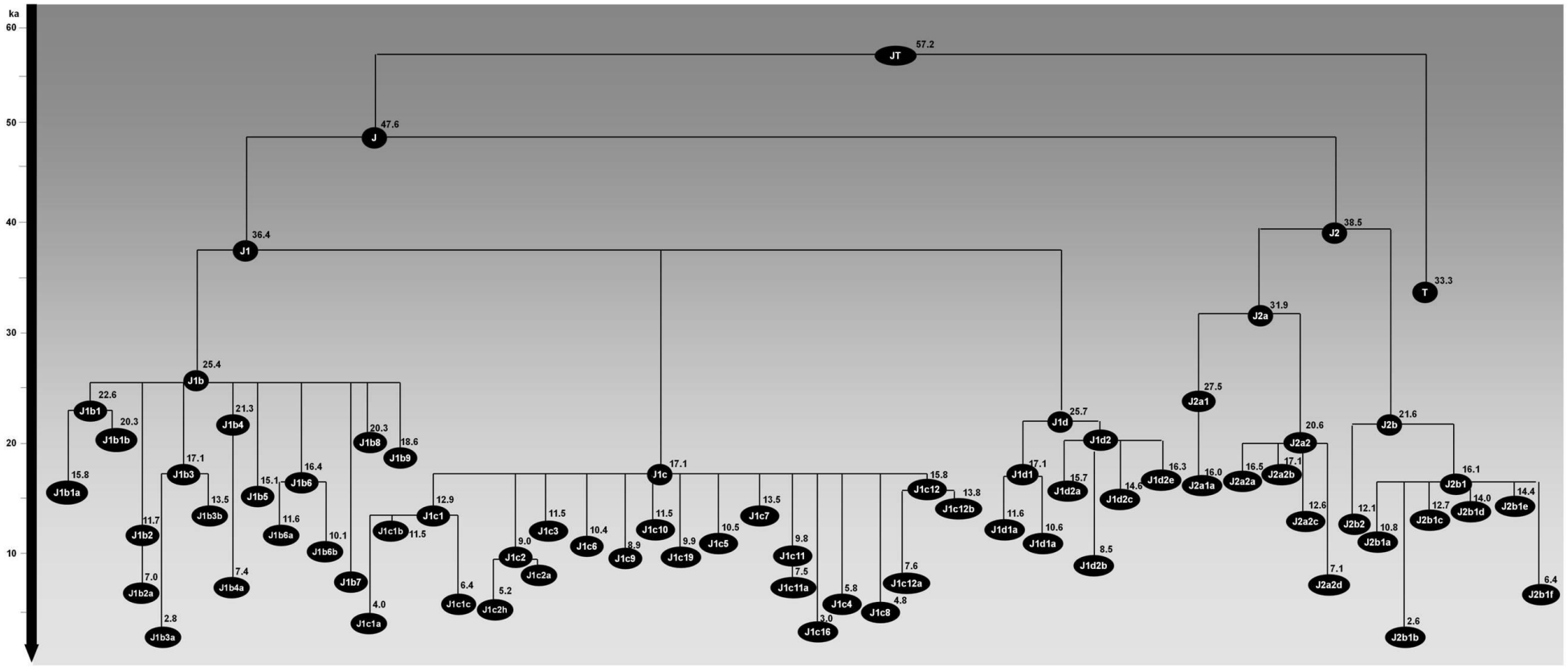

Figure 58 Schematic tree of haplogroup J.
Ages (in ka) indicated are maximum likelihood estimates obtained for the complete mtDNA genome.

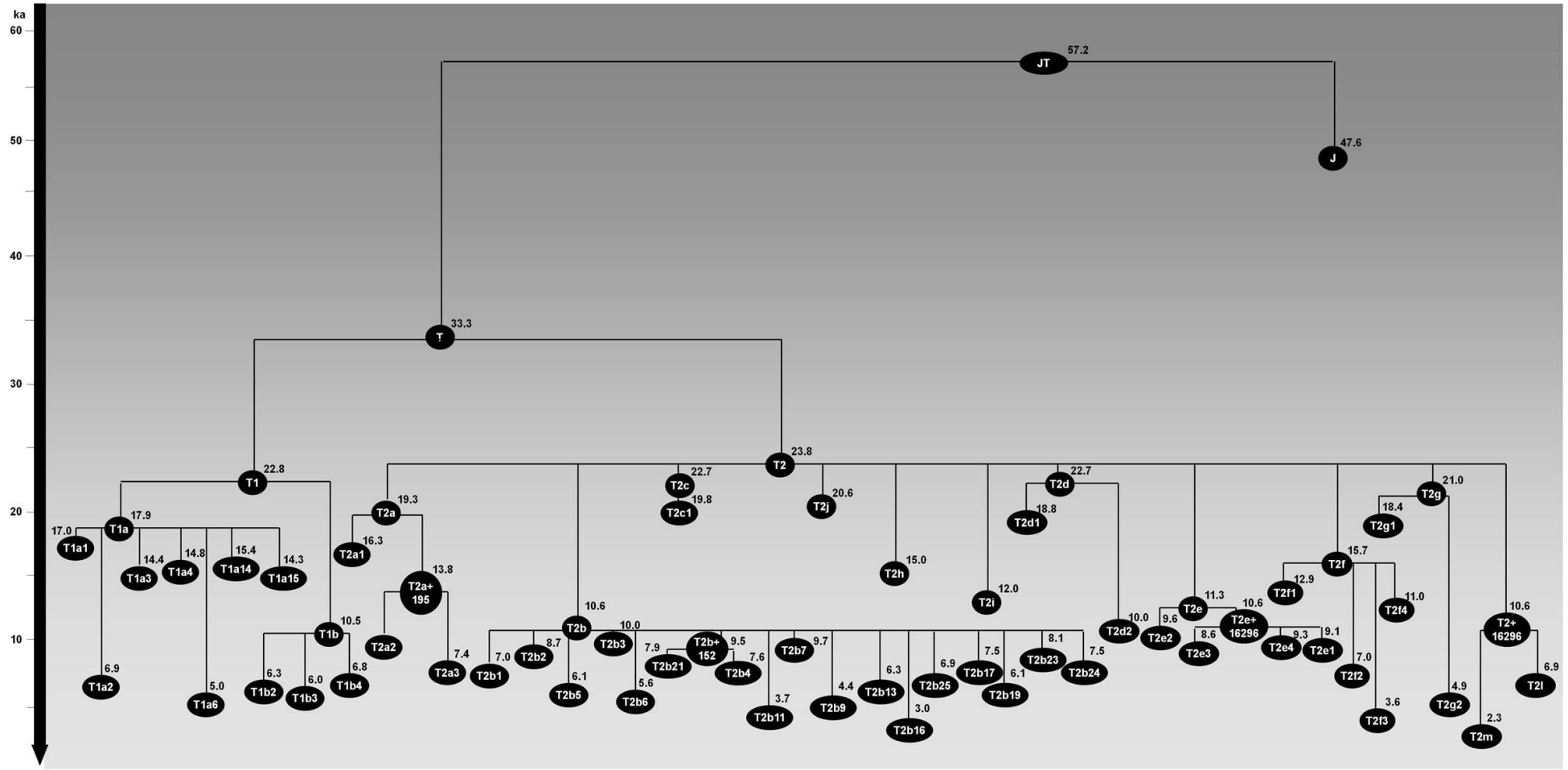

Figure 59 Schematic tree of haplogroup T.
Ages (in ka) indicated are maximum likelihood estimates obtained for the complete mtDNA genome.

As can be seen in the distribution maps represented in Figure 60, J is predominant in the Arabian Peninsula, especially in the central/eastern part, where it peaks at 20%, the highest frequency at a worldwide scale. While haplogroup T is broadly low frequent (5-7%) across the Peninsula, this is nevertheless quite close to its frequency in the Near East (7-12%).

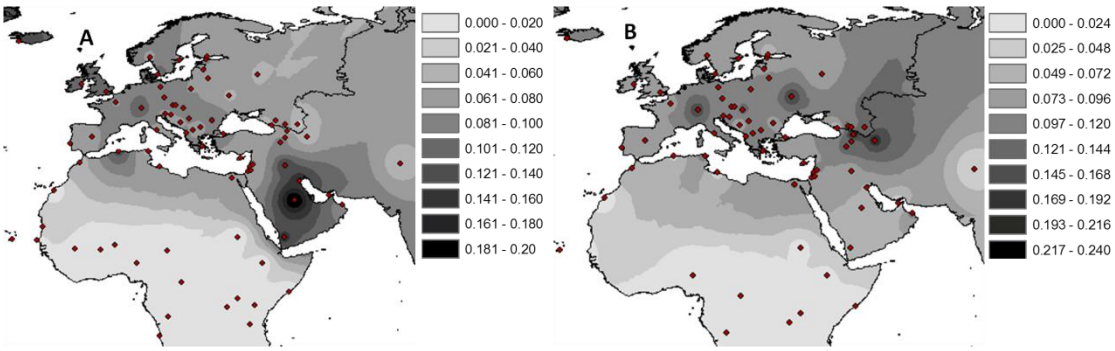

Figure 60 Frequency maps based on HVS-I data for haplogroups J (A) and T (B).

The maps of the diversity measures for J (Figure 61) and T (Figure 62) show that these lineages display higher diversities in the Near East and Arabian Peninsula region, followed by the Caucasus region and Indian subcontinent. The close frequency and diversity values between the Near East and Arabian Peninsula JT pools seem to indicate that both regions were probably in close contact around the time of emergence and dispersion of some JT sub-haplogroups, as further supported by the phylogeographic analyses described in detail bellow.

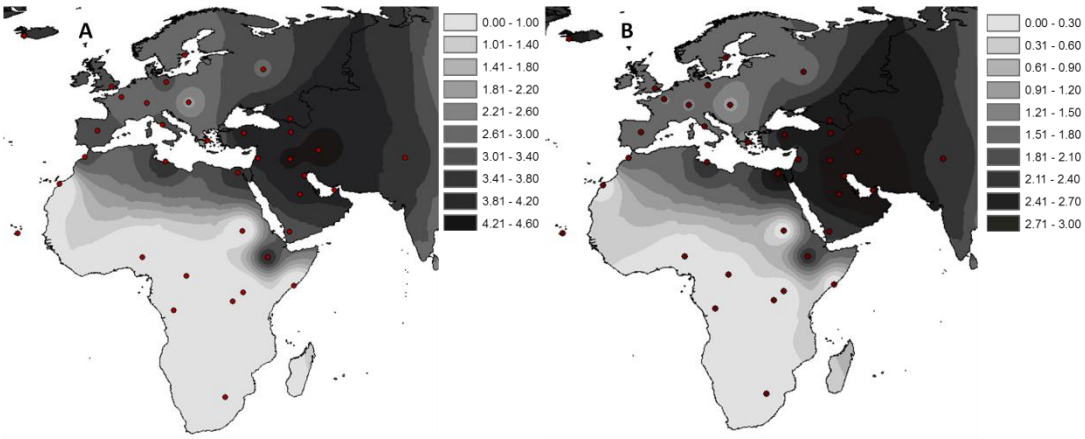

Figure 61 Distribution maps for haplogroup J for the diversity measures π (A) and ρ (B) based on HVS-I data.

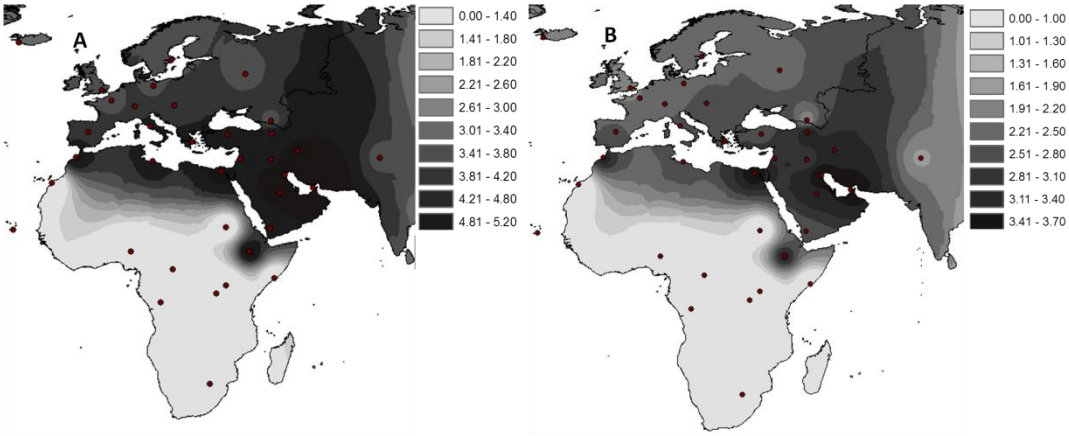

Figure 62 Distribution maps for haplogroup T for the diversity measures π (A) and ρ (B) based on HVS-I data.

Another line of evidence comes from the BSP analyses for JT in Arabia and Near East (Figure 63), for which population expansions were very close in time. The population expansion in haplogroup J occurs first in the Near East, around 15.0 ka until 11.0 ka, and soon after in Arabia, mainly between 12.0 and 8.0 ka. For T, the low number of sequences observed in Arabia (n= 21; about half the ones used for the other BSP inferences), all very diverse between them (as I will discuss below), renders the population expansion slightly older in Arabia (18.0-14.0 ka) than in the Near East (15.0-8.0 ka with slight break in the increase around 12.0 ka). Most likely, the sampling effect is leading to over-estimate the age of T in Arabia, as I did not find any evidence for it in the founder analysis that I will discuss later.

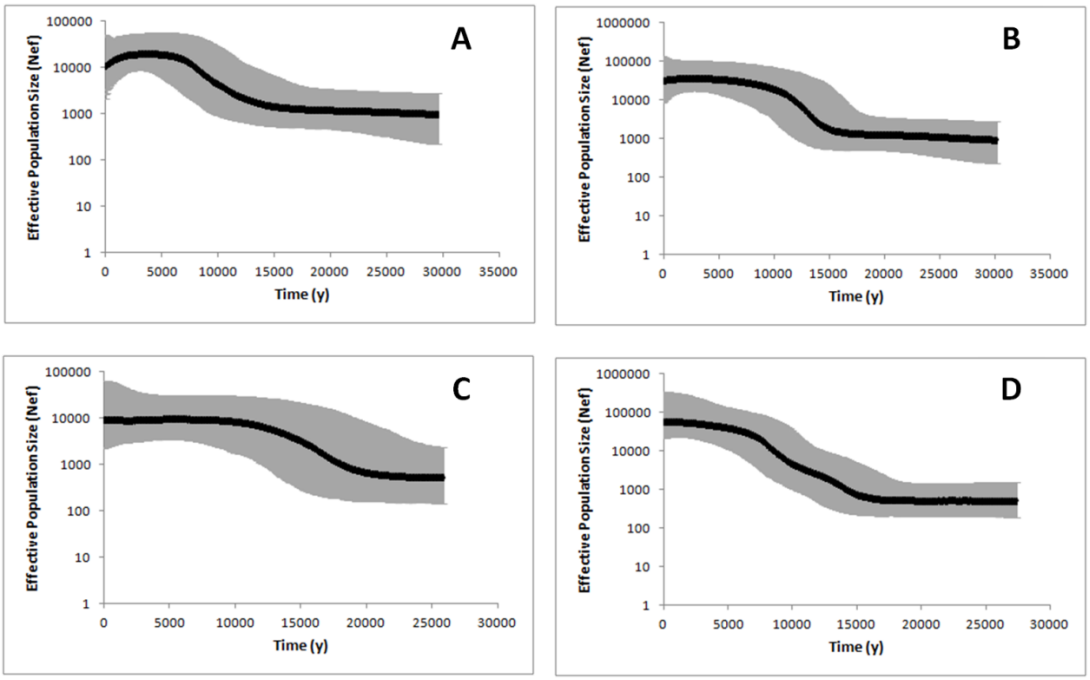

Figure 63 Bayesian skyline plot indicating hypothetical effective population size through time based on data from haplogroup J of Arabia (A) and Near East (B) and from haplogroup T of Arabia (C) and Near East (D).

The Arabian J samples cluster in specific sub-groups, both the frequent J1 and the rarer J2. According to Pereira[289], J1 and J2 seem to have begun to diverge in the Near East ~36.4 [24.4-49.0] ka and ~38.5 [26.5-51.0] ka, respectively; and some of the lineages migrated into Arabia and participated in local expansions.

J1 branches in three sub-haplogroups: J1b, J1c, and J1d. J1b, harbours many Arabian samples, displaying some interesting patterns. Pereira[289] argued that J1b is more frequent in the Near East and Caucasus, supporting a Near Eastern origin, at ~25.4 [19.3-31.6] ka and most of its subhaplogroups, such as J1b1 and J1b2, probably diverged within the Near East, after the LGM. In this work, by enlarging the sampling of complete J Arabian sequences, I discovered that J1b is an important sub-haplogroup in Arabia, reaching a considerable frequency there (Figure 64) and displaying some star-like branches (Figure 65).

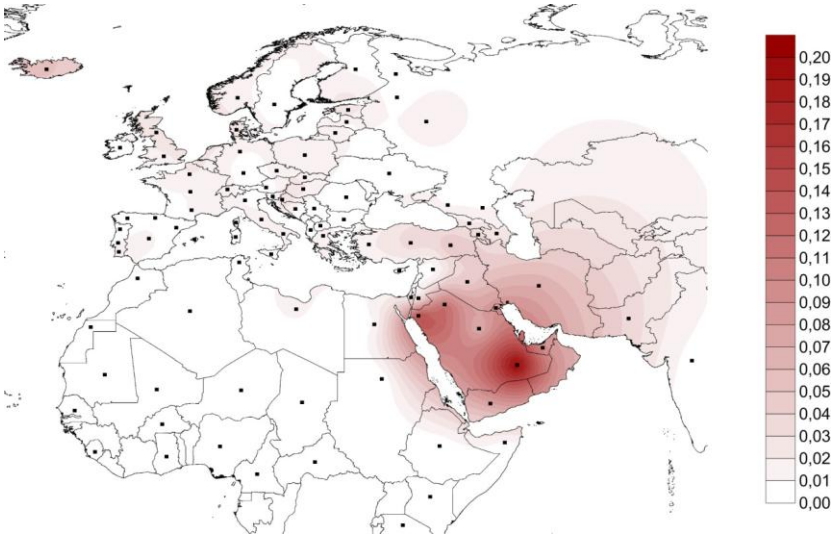

Figure 64 Frequency maps based on HVS-I data for haplogroups J1b (from Pereira).[289]

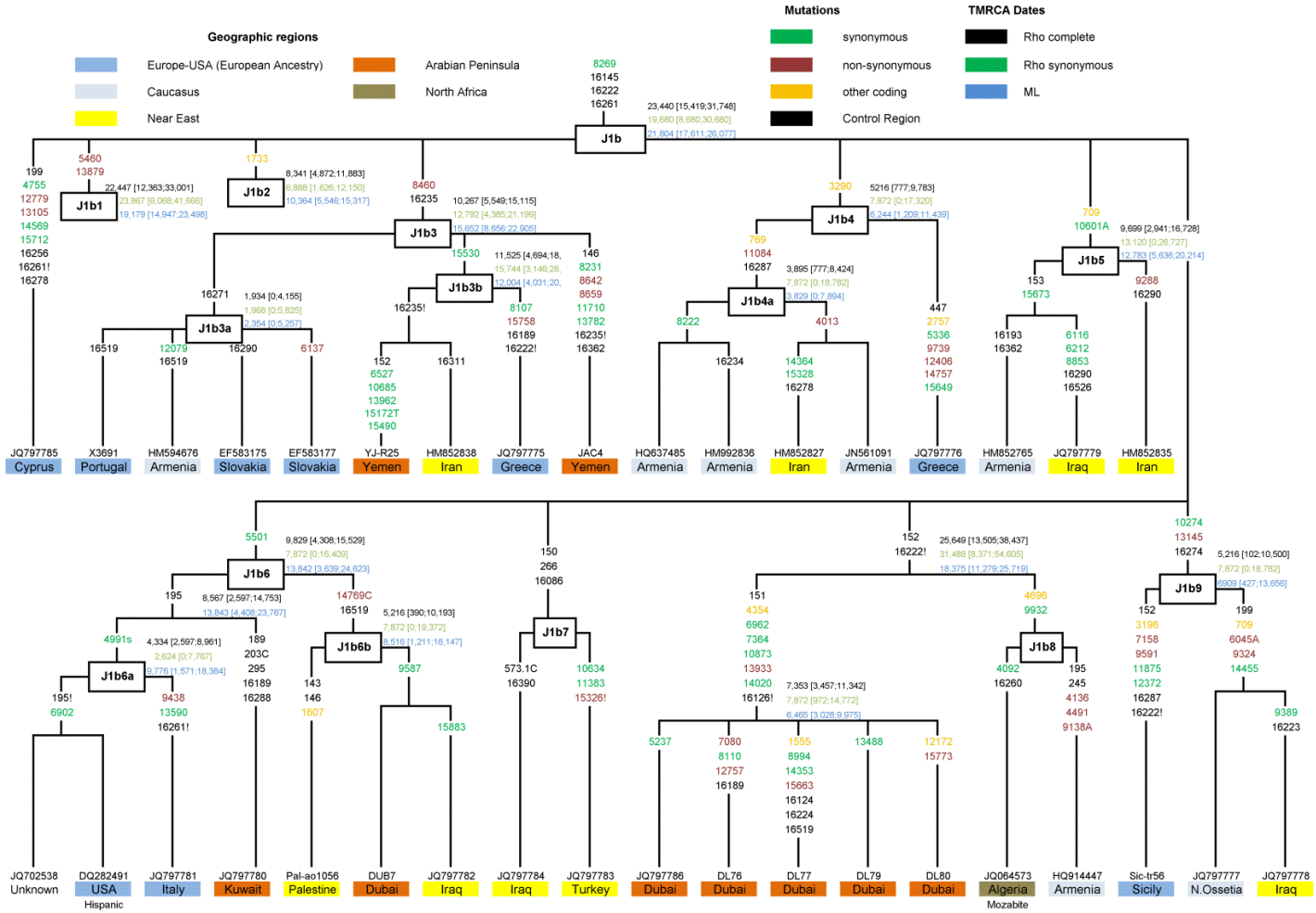

Figure 65 Phylogenetic tree of J1b.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

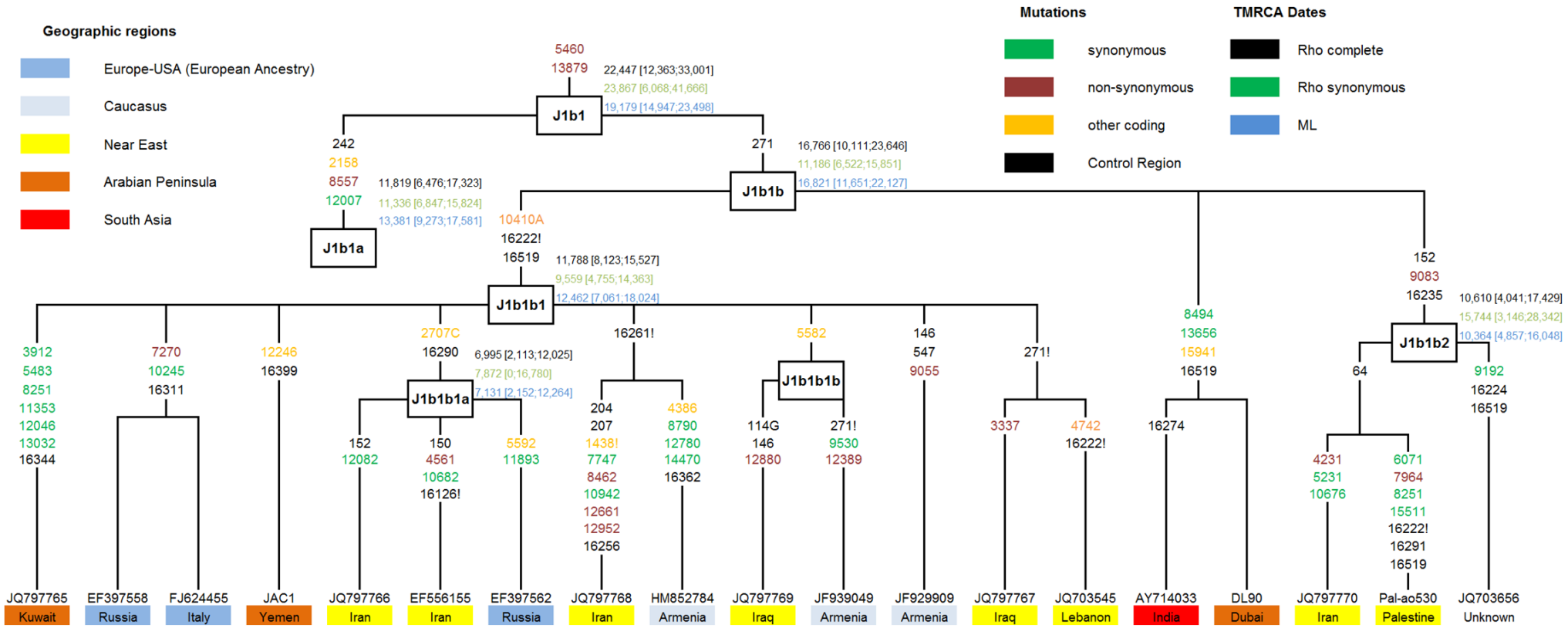

Figure 66 Phylogenetic tree of J1b1.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

The more globally frequent J1b1 (Figure 66) bears only three Arabian samples, one in the oldest (dating to ~22.6 [16.6-28.8] ka[289]) and more frequent J1b1a, for which Pereira[289] inferred for a Near Eastern origin although it is very frequent across Europe; and two Arabian samples in J1b1b, this haplogroup is clearly of Near Eastern origin and dates to ~20.3 [13.6-27.2] ka ago[289]. These sporadic Arabian J1b1 samples are likely to be of recent introduction, as shown by the close proximity between a Dubai (DL90) and Indian (AY714033)[290] sequences.

Curiously, the generally minor J1b sub-haplogroups have a much more interesting distribution inside Arabia. This is the case for sub-haplogroup J1b2 (Figure 67), defined by the transition at position 1733 (not identifiable when only characterising the HVS-I diversity), which by attending to the complete sequences seems to be located mainly in the Arabian Peninsula (and one basal Dubai sample supports its old presence in the region) and is also observed in the Near East, North Africa and southern Europe (most of these European sequences form a group restricted so far to this region and displaying no additional variation rendering it very recent). I dated J1b2 to 10.4 [5.5-15.3] ka. The J1b2 sub-groups with a clear dominance of Arabian samples (J1b2a and the one defined by polymorphisms at positions 152-16311) are dated at ~6.5 ka.

J1b3 (Figure 65) although rare is very widespread, from the Arabian Peninsula to the Caucasus and southeastern and central Europe; the age estimation is 15.6 [8.7-22.9] ka. J1b6 (Figure 65), splits into J1b6a which is mainly Arabian, and J1b6b that is observed in Europeans sharing the fast-evolving polymorphism at position 195 with a Kuwait sample; so most probably this sub-haplogroup had an Arabian origin at 13.8 [3.6-24.6] ka. An unnamed J1b branch, dated at 18.4 [11.3-25.7] ka and defined by the reverse transition at position 16222 and the transition at 152, both fast-evolving, unites a branch characterised by eight polymorphisms which has been so far only detected in Dubai (Figure 65), dating at 6.5 [3.0-10.0] ka; and the sub-haplogroup J1b8 (just characterised by two polymorphisms) observed in one Armenian and one Algerian sample.

Some of the clusters which are frequent in Arabia include individuals from North Africa, providing a tentative to link with the Arab Islamisation of North Africa from the 7th century AD.

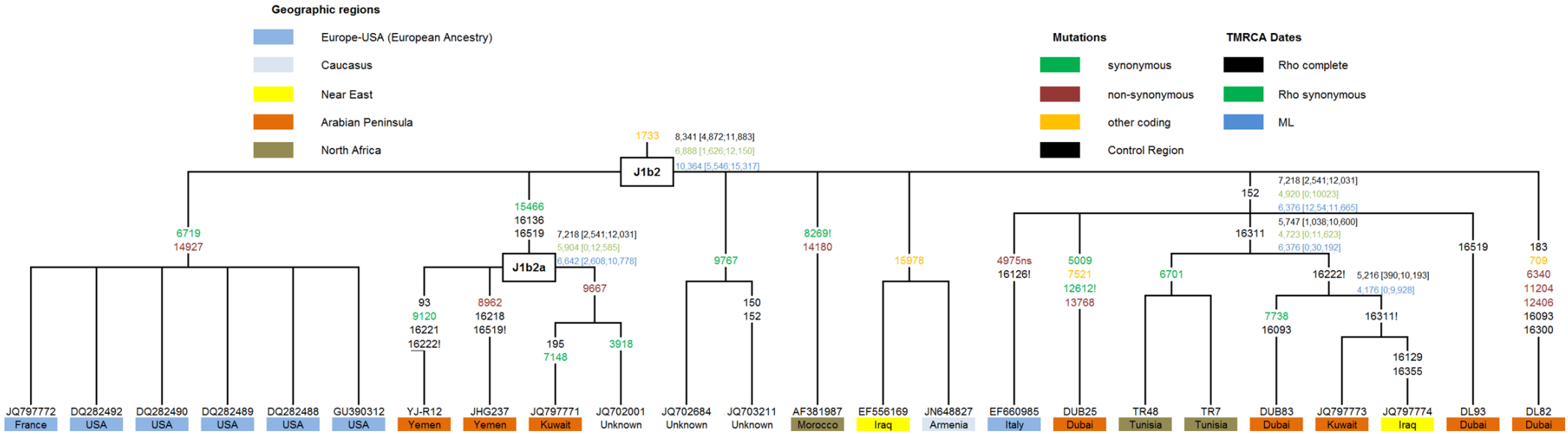

Figure 67 Phylogenetic tree of J1b2.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

No Arabian samples cluster in the J1c branch, which is mainly European and is also typical of the Ashkenazi Jews (72% of J Ashkenazi lineages are assigned to this sub-haplogroup).[288] Pereira[289] inferred for its origin in Europe, dating to the Late Glacial period (~17.1 [13.5-20.7] ka), and having a few lineages (J1c12) which are found especially in the Near East and southern Caucasus, probably due to recent back-migrations from Europe (also supported by Pala et al.).[101] These lineages seem not to have reached the Arabian Peninsula.

Haplogroup J1d is quite frequent in the Arabian Peninsula, and also present in southeast Europe, the Mediterranean basin, the Caucasus, the Near East, North Africa and eastern Africa. Given its wide geographic distribution, it could have been originated in the Near East (Figure 68), but as I characterised a Dubai sequence which is basal in the haplogroup, and in the J1d1a sub-group there is one Dubai and an Armenian basal sequences, an origin in Arabian Peninsula cannot be ruled out. I dated J1d to ~16.6 [12.9-20.4] ka, J1d1 to 13.7 [8.9-18.7] ka and J1d1a to 9.9 [6.2-13.7] ka.

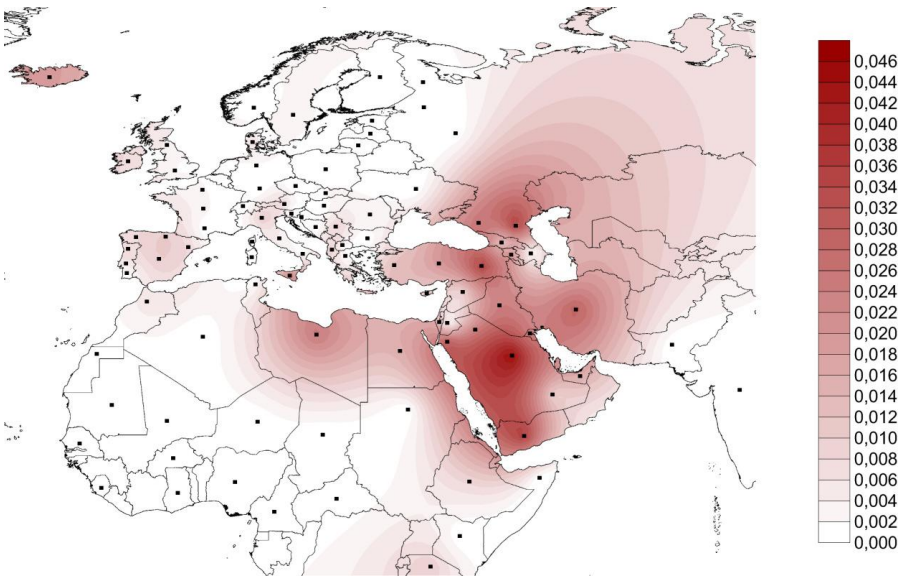

Figure 68 Frequency maps based on HVS-I data for haplogroups J defined by the transition at 16193, which mainly corresponds to J1d, but can also include J2d (from Pereira).[289]

The complete-mtDNA genome data seem to suggest that the Arabian Peninsula was the centre for the migration of J1d1a1 lineages to East Africa, reaching as far south as Tanzania (Figure 69), and dating to ~7.1 [4.4-9.9] ka (Figure 70). I will come back to this later in the founder analysis evaluation. The possible haplogroup J1d2 (defined by the fast evolving polymorphism at position 16519 by Pala et al.)[101] is more frequent to the north (the probable centre of radiation was the Fertile

Crescent) and east, in the Caucasus and Asia (Figure 71), and only J1d2c2, dated at 3 [0.2-5.8] ka, is mostly Arabian sharing an ancestor at 11.3 [6.4-16.4] ka with the recent (0.9 [0-2.2] ka) sub-group J1d2c1 restricted to the Caucasus.

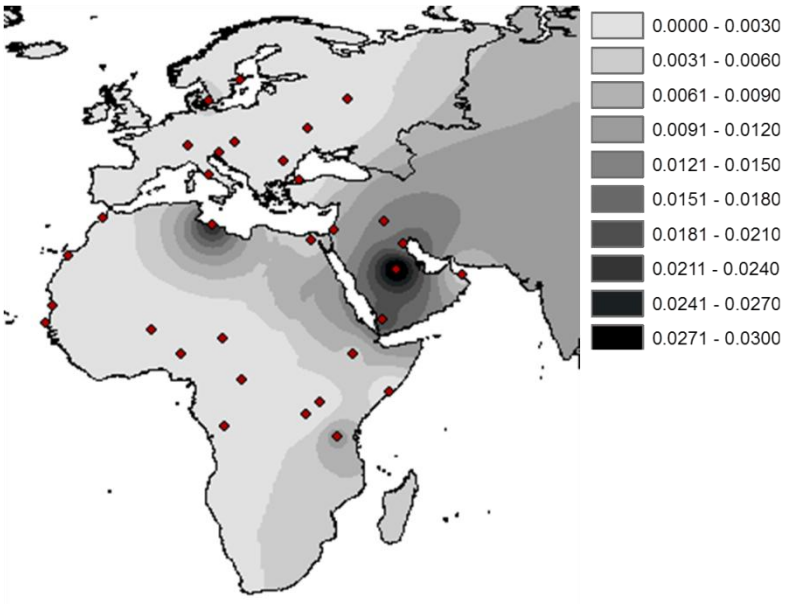

Figure 69 Frequency maps based on HVS-I data for the sub-haplogroup J1d1a.

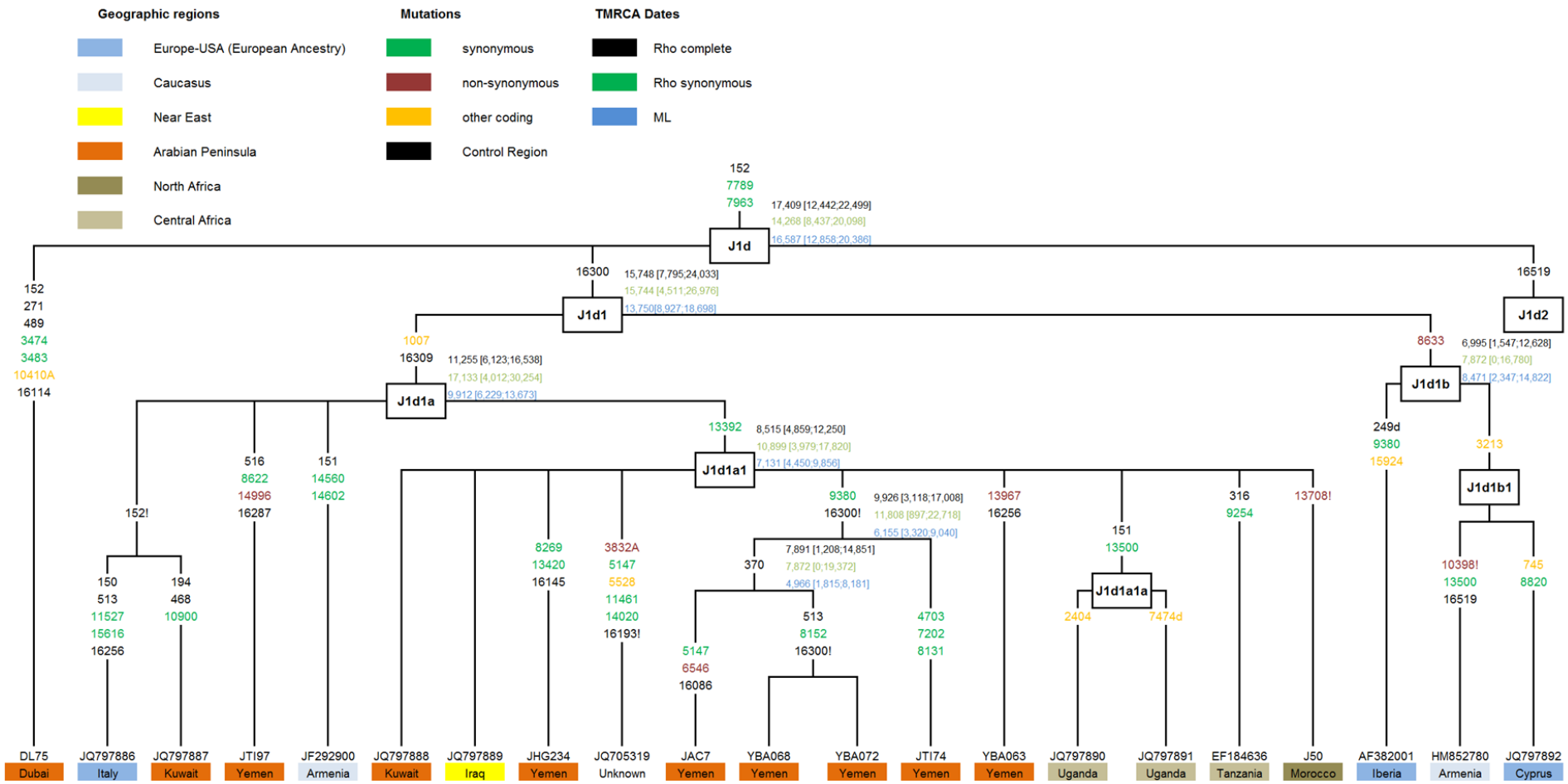

Figure 70 Phylogenetic tree of J1d1.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; deletions are indicated “d”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

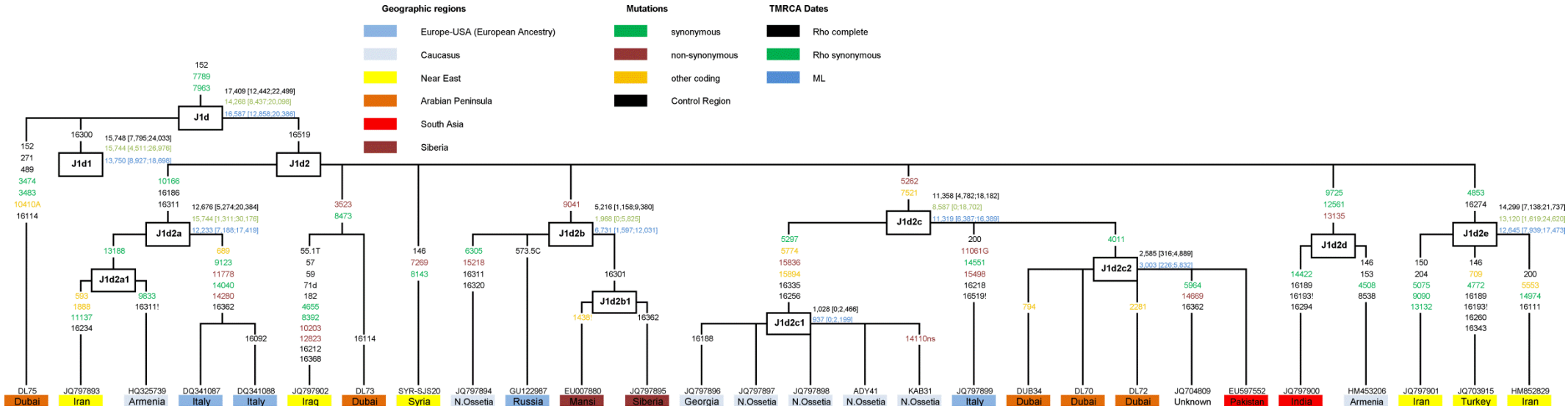

Figure 71 Phylogenetic tree of J1d2.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

J2 dates to ~38.5 [26.5-51.0] ka[289] and is widespread in Europe, especially in central and western regions, and parts of the Near East and North Africa, as well as the Arabian Peninsula (Figure 72). It presents two branches: J2a and J2b dating to ~31.9 [21.8-42.4] ka and ~21.6 [11.9-31.6] ka, respectively.[101; 289]

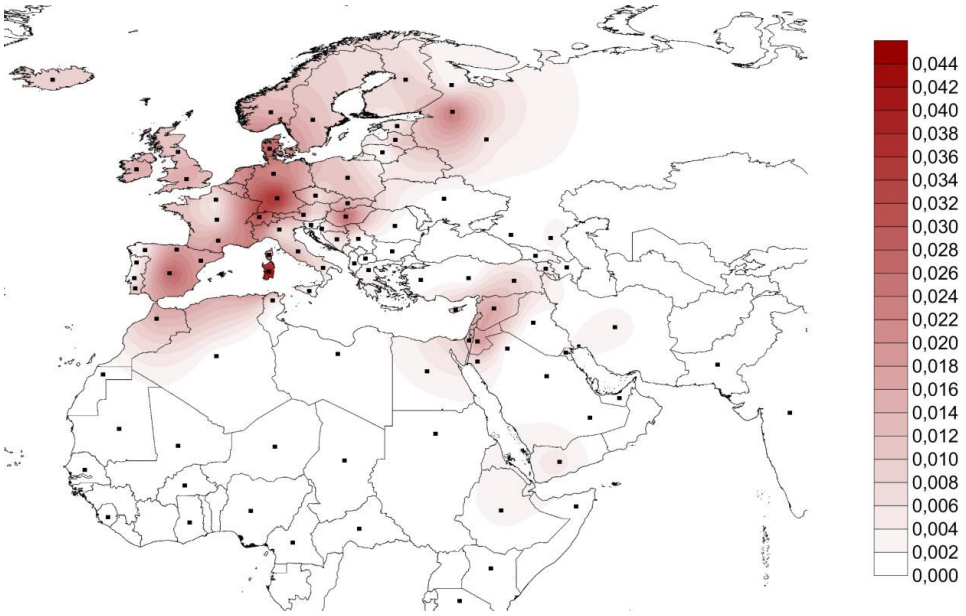

Figure 72 Frequency maps based on HVS-I data for haplogroup J2 (from Pereira).[289]

J2a splits into two subclades, J2a1 and J2a2, which have very different geographic distributions. J2a1 is clearly European with a coalescence time around ~27.5 [16.8-38.8] ka[289], and no Arabian complete sequences have been ascribed into this group.

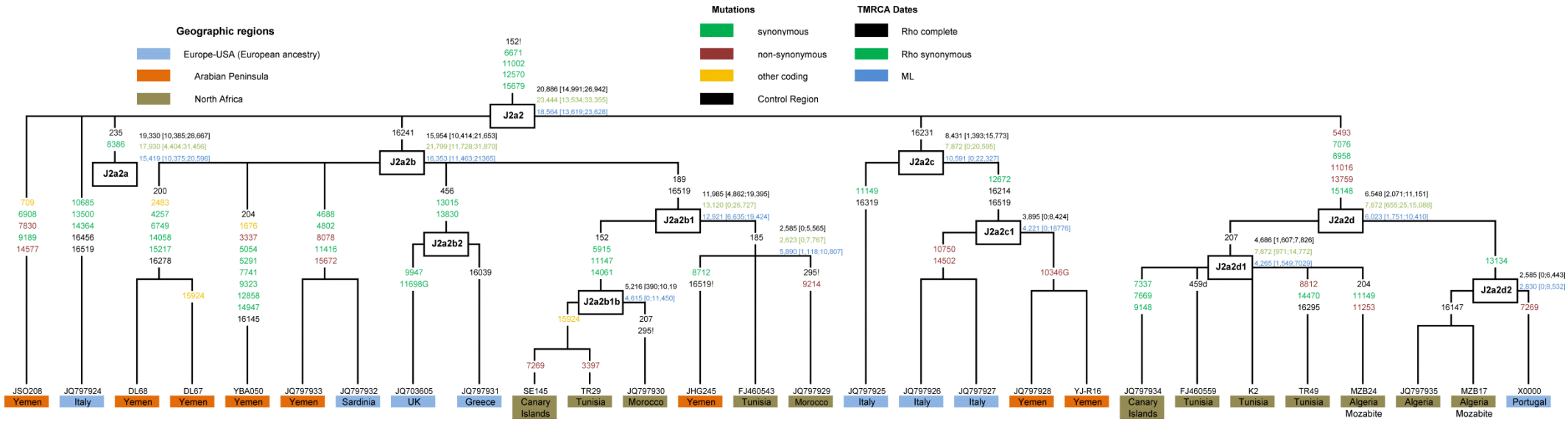

Figure 73 Phylogenetic tree of J2a2.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

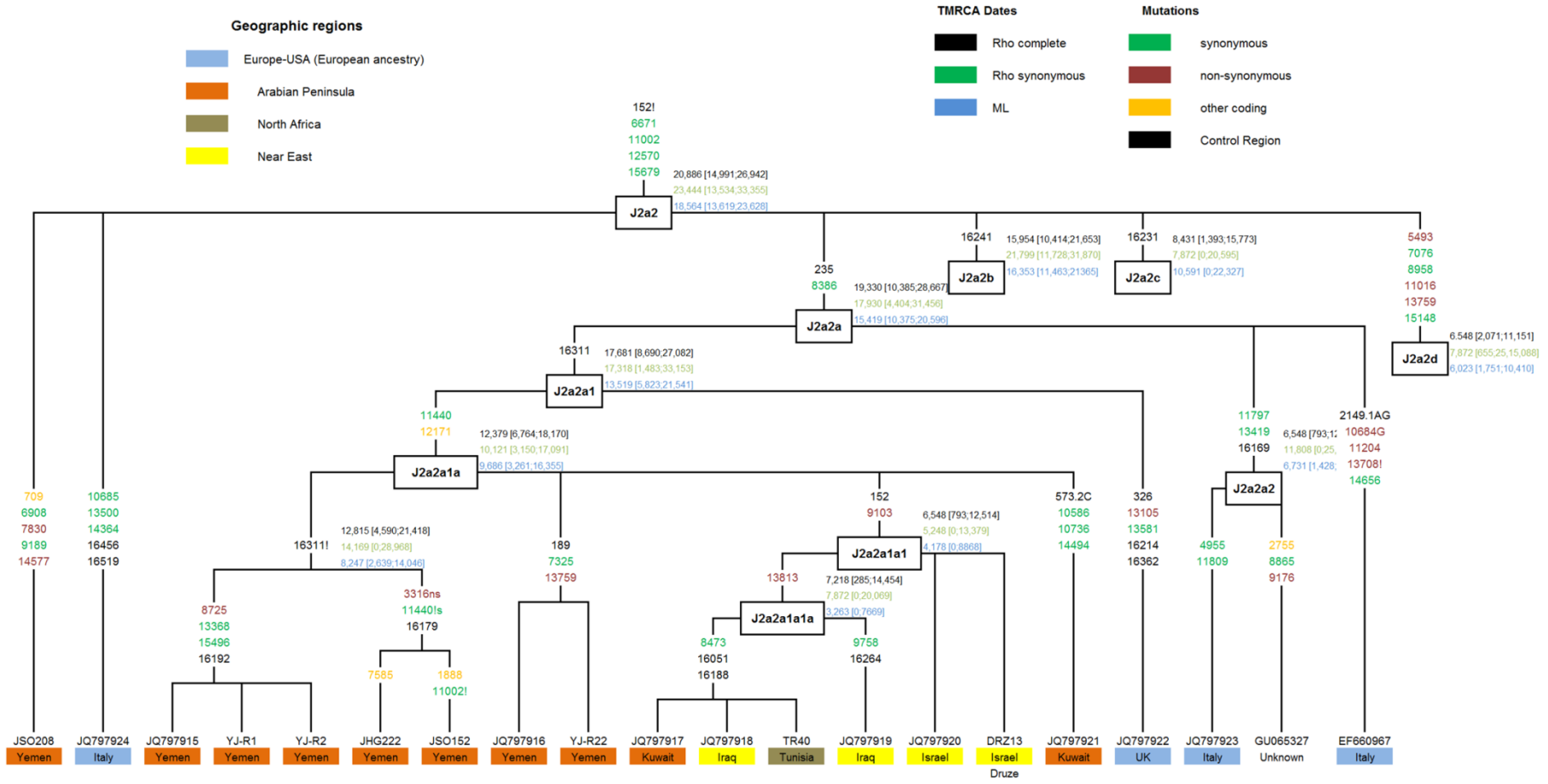

Figure 74 Phylogenetic tree of J2a2a.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; insertions are indicated by a dot followed by the number of repetition and the nucleotide position; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

J2a2 (Figure 73 and 74), coalesces about 18.6 [13.6-23.6] ka, and the geographic distribution, based on the complete sequence tree (not identifiable in the HVS-I diversity), includes the Near East, Arabian Peninsula, North Africa and the Mediterranean Basin, suggesting a re-expansion from the Near Eastern refugium shortly after the LGM along the Mediterranean and southerly into Arabia. There are two basal J2a2 lineages, located one in Yemen and one in Italy, and four sub-clusters: J2a2a, J2a2b, J2a2c and J2a2d. The sub-haplogroup J2a2a (Figure 74) dates to ~15.4 [10.4-20.6] ka and presents a basal European sequence, a rare European cluster (J2a2a2), and a more frequent J2a2a1 cluster defined by the fast-evolving substitution at position 16311, which has a basal European sequence and the cluster J2a2a1a mainly observed in the Arabian Peninsula, Near East and North Africa, dating to 9.7 [3.3-16.4] ka. The sub-haplogroup J2a2b (dating to 16.4 [11.5-21.4] ka), is frequent in the west of Arabia and in North Africa (Figure 75), and the phylogeny (Figure 73) shows that the Arabian samples are basal and clustered with North African ones in J2a2b1 (coalescent time of 12.9 [6.6-19.4] ka); the complete European J2a2b samples are mainly Mediterranean. J2a2c (dating to 10.6 [0-22.2] ka) is shared between Arabia and Europe, while J2a2d (dating to 6.0 [1.8-10.4] ka) is especially frequent in North Africa and shared with Europe.

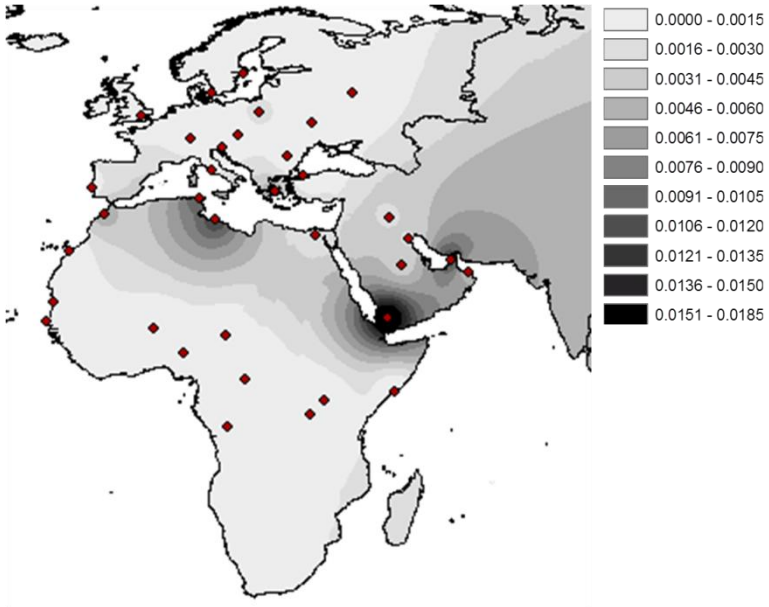

Figure 75 Frequency maps based on HVS-I data for the sub-haplogroup J2a2b.

J2b, dating to 31.9 [21.8-42.4] ka[289], is mainly represented by samples from Europe (especially the Mediterranean region and southeast Europe) and also from the Near East. So far, one Dubai complete sequence was ascribed to this clade,

and its distribution cannot be further explored through the HVS-I diversity, as it shares the polymorphism at position 16193 with the more frequent J1d.

A quite diverse pattern is displayed by the 20 Arabian samples incorporated into the complete haplogroup T mtDNA tree, as they do not cluster within specific sub-haplogroups, testifying that this clade is probable involved in more recent events. The complete T tree from Pereira[289] with the new Arabian samples is displayed in the CD attached. Two basal Arabian sequences are observed in T1a, another in T1b and T1a1a1, and there are several single instances of minor T1 sub-haplogroups (T1a3, T1a4, T1a6). Also for T2, one or two sequences are basal and observed in minor sub-haplogroups (T2a1a+8251, T2b+16239, T2c1b2, T2c1c, T2e3, T2i, T2g1a1a), which are widely distributed, ranging from Near East, North Africa, Caucasus and Europe, and usually having ages between 9 and 17 ka.[289]

8.2. HVS-I founder analyses

By applying founder analysis, I investigated the input of lineages from the Near East and Pakistan into the Arabian Peninsula (Figure 76). A total of 153 and 91 founder lineages were identified when using the *f1* and *f2* criteria, respectively, observed in 935 individuals (Tables S8 and S9). Attending to the diversity observed in the region, the periods for entrance of these founders peak at 1.0 and 11.0 ka for the *f1* criterion, while an additional peak at 16.0 ka is observable for the *f2* criterion.

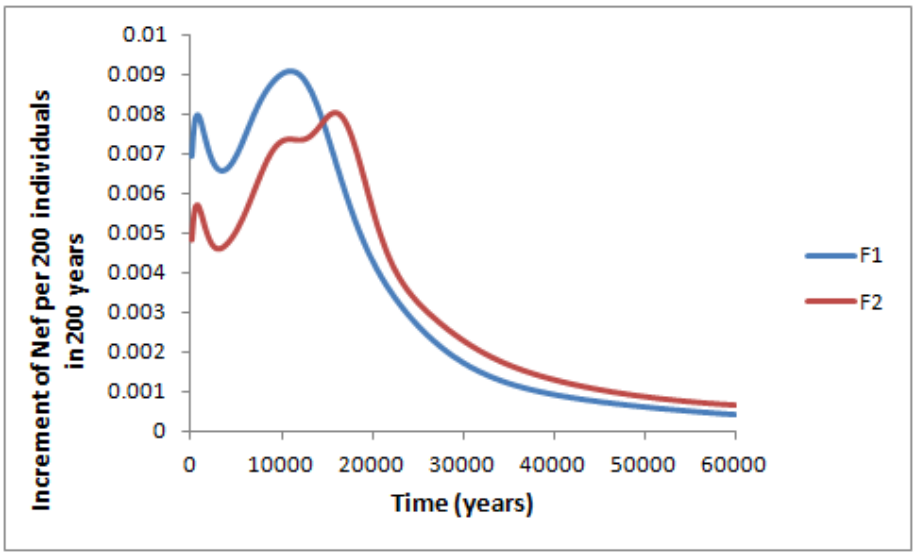

Figure 76 Probabilistic distribution of founder clusters across migration. Times scanned at 200 yr intervals from 0–60 ka, using both an *f1* criterion (blue line) and an *f2* criterion (red line), when considering a migration from the Near East and Pakistan to Arabian Peninsula

Given these peaks, I then imposed four main migration periods, at 1.0, 8.0, 11.0 and 16.0 ka, for the input of these lineages into Arabia (Figure 77), which would represent recent events, the Neolithic introduction, the Younger Dryas and the Late glacial period, respectively. The mean values were of 31% of the lineages having been introduced in the Late glacial period, 28% at ~11 ka, 25% at ~8 ka, and 15% at the recent period ~1 ka, for the *f1* criterion. When using the *f2* criterion, the dates of introduction are generally older, which tends to increase the proportion of sequences in the older migration period.

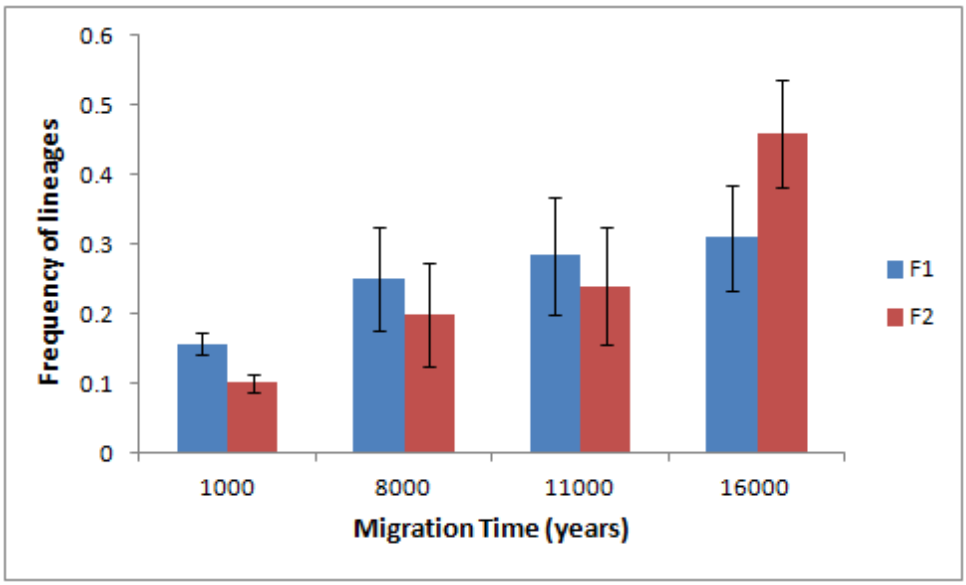

Figure 77 Probabilistic proportion of founder clusters considering different migration events.
By using both a *f1* criterion (blue bar) and a *f2* criterion (red bar) and when considering a migration from Near East and Pakistan to Arabian Peninsula for four migration periods at 1.0 ka ,8.0 ka, 11.0 ka and 16.0 ka.

Figure 78 illustrates in detail the probabilistic proportion of introduction for each lineage at each of the four migration periods, when using the *f1* criterion. This figure resembles the images from Structure-like analyses, but instead of proportion of ancestry for each individual it reflects the probability of each founder to have been introduced in certain migration events. The most frequent lineages and with high probabilities (p>75%) of having been introduced during the Late Glacial period are affiliated to the haplogroups K, U2, U3, U4, N1a1a, N1a1b and HV1. At the 11.0 ka event the most frequent and contributing lineages belong to HV and R0a, although the highest probabilities attain only 40% as there is much uncertainty in affiliation into this or the two other close migrations, at 8.0 and 16.0 ka (as happens especially for a J root lineage, reflecting its composite nature). The major contribution for the 8.0 ka event comes from J1b, T1a and M1, with probabilities higher than 60%. For the recent period at 1.0 ka, the J1d1a, K1, HV8 and N1a3 lineages reflect probabilities higher than 70%.

When using the *f2* criterion (Figure 79), the pattern of the main migrants in each period is similar, although ages tend to be overall older, as already mentioned: the first arriving haplogroups were mainly affiliated in U and N1 haplogroups; followed at 11.0 ka by R0a and HV; then at 8.0 ka dominated by J and T; and finally a mix of young and derived lineages.

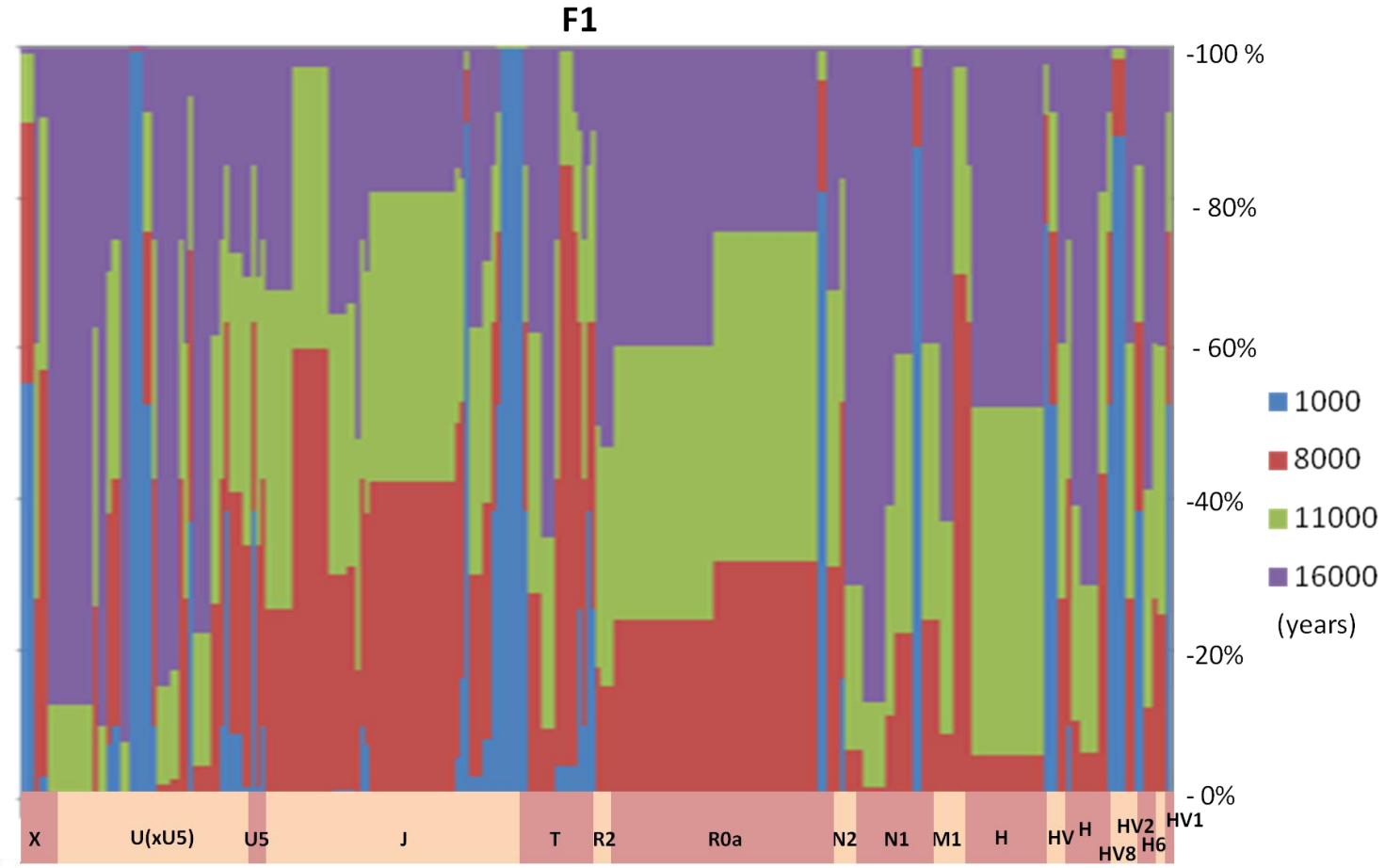

Figure 78 Probabilistic proportion of founder clusters considering four migration periods (1.0, 8.0, 11.0 and 16.0 ka).
By using a *f1* criterion and by assuming a Near East and Pakistan source for migrations into Arabian Peninsula. The haplogroup affiliations of the founders are indicated in the bottom.

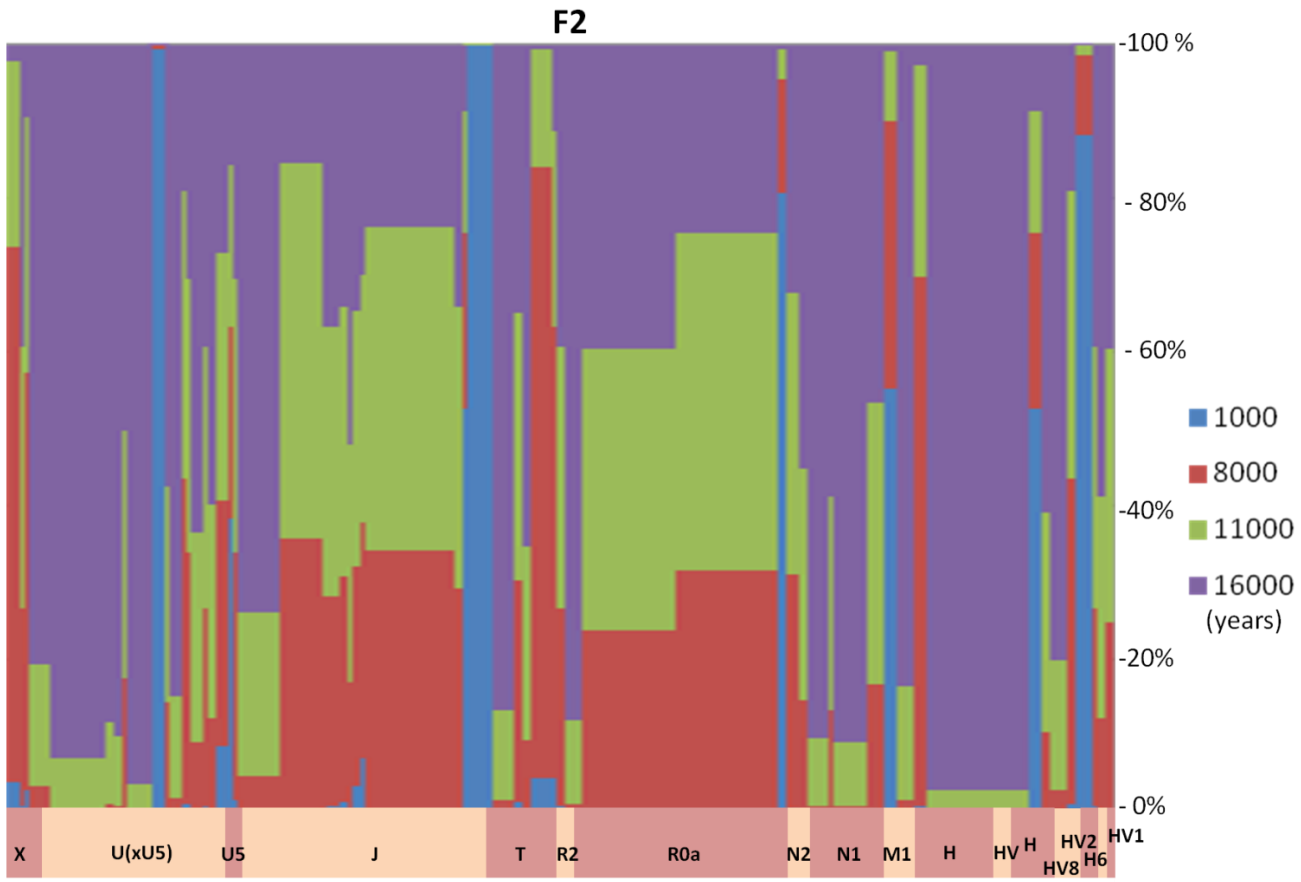

Figure 79 Probabilistic proportion of founder clusters considering four migration periods (1.0, 8.0, 11.0 and 16.0 ka).
By using a *f2* criterion and by assuming a Near East and Pakistan source for migrations into Arabian Peninsula. The haplogroup affiliations of the founders are indicated in the bottom.

Then I tried to get insights into possible heterogeneity in the introduction of the Near East and Pakistan founders in west and east Arabia. A proportion of 35% and 28% founder individuals were exclusive to west and east Arabia, respectively; from these, a proportion of 20% came necessarily from the Near East to both regions, and 25% and 41% from Iran to west and east Arabia respectively; Pakistan had a 4% input in east Arabia only. Of course these private founders overestimate low frequency lineages; so, when considering all founders, private and shared between west and east Arabia, the proportions of lineages coming necessarily from the Near East is 9% and 7.4% to west and east Arabia respectively, while from the Iran to west and east Arabia was respectively 11% and 14%, and only 1% from Pakistan to east Arabia. The high amount of founder lineages for which it was not possible to certainly affiliate to a geographical origin reflects a common ancestry between Near East and Iran.

Thus, the entrance of sequences from the Near East into Arabia took place over time, with the post-LGM period playing a dominant role. The Neolithic also saw the arrival of some sequences in Arabia, but autochthonous lineages were largely involved in the population expansions occurring during this period. The Persian influence is detectable and is slightly higher in the closer east Arabian side than in the more distant west.

9. Mitochondrial genetic exchanges across the Red Sea

The importance of the Bab al Mandab strait in genetic exchanges between eastern Africa and the Arabian Peninsula has been highlighted in many recent works, not only in the out-of-Africa migration[191] but also in more recent occurrences such as the seafaring activity across the Red Sea, during the early and middle Holocene period,[258; 260] and the Arab slave trade established between the 7th and 19th centuries.[108] I addressed this issue by characterising the mtDNA complete genome of the African L4 and L6 haplogroups, which have not been previously studied at the level of complete mtDNA genome. These are the phylogenetically closest African clades to L3, the one involved in the out-of-Africa migration. I then performed a founder analysis of an enlarged mtDNA HVS-I database for all L and Eurasian (as R0, HV, J, T, M1 and U6) haplogroups in order to quantify and date migrations across the Red Sea, and also into the neighbouring region of North Africa.

9.1. Insights from the complete sequencing of haplogroups L4 and L6

The improved phylogenetic tree for the rare haplogroups L4 and L6, including age estimates using the complete genome (using ρ and ML) and synonymous (using ρ) clocks, are presented in the figures below and an outline topology with ages scaled against the ML estimates is displayed in Figure 80. Sequences used in the phylogenetic inference are displayed in Table S1 and S2, and frequencies used in the interpolation maps are summed up in Table S10.

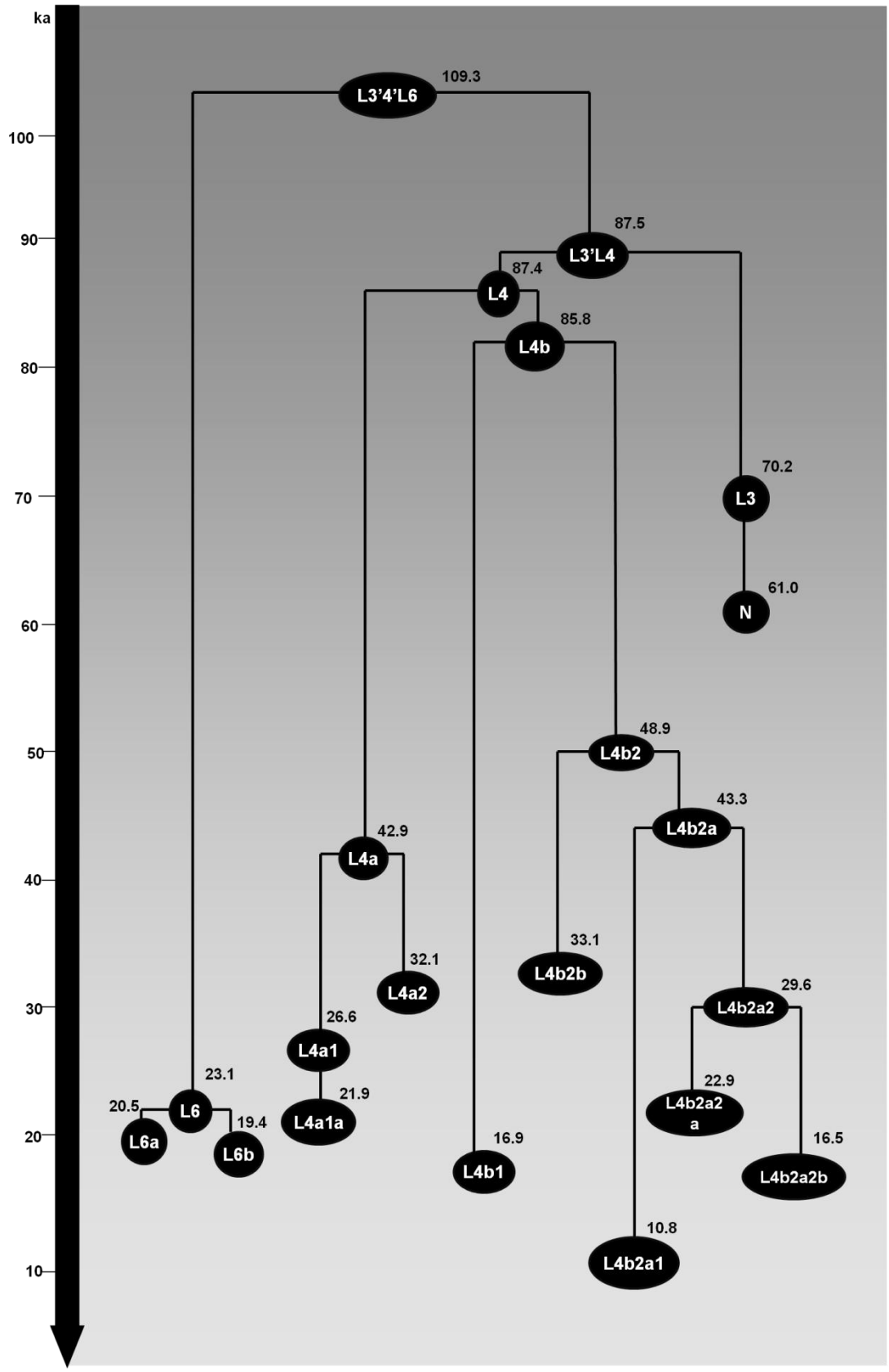

Figure 80 Schematic tree of haplogroups L4 and L6.
Ages (in ka) indicated are maximum likelihood estimates obtained with the complete mtDNA genome.

L4 the sister haplogroup of L3, seems to be more frequent in East Africa (Figure 81A) and points at a most likely East African origin with a coalescence age of 87.4 [73.6-101.4] ka. Haplogroup L4 splits into two major subclades L4a (Figure 82) dates to 42.9 [31.7-54.5] ka and L4b (Figure 83) dates to 85.8 [71.5-100.4] ka, and both exist mostly in East Africa, and the few Arabian lineages indicate possible recent exchange networks between eastern Africa and Arabia. The subclade L4b1 (defined by 27 polymorphisms and dating to 16.9 [10.6-23.5] ka) is curiously almost restricted to Burkina Faso except for one Yemen sample.

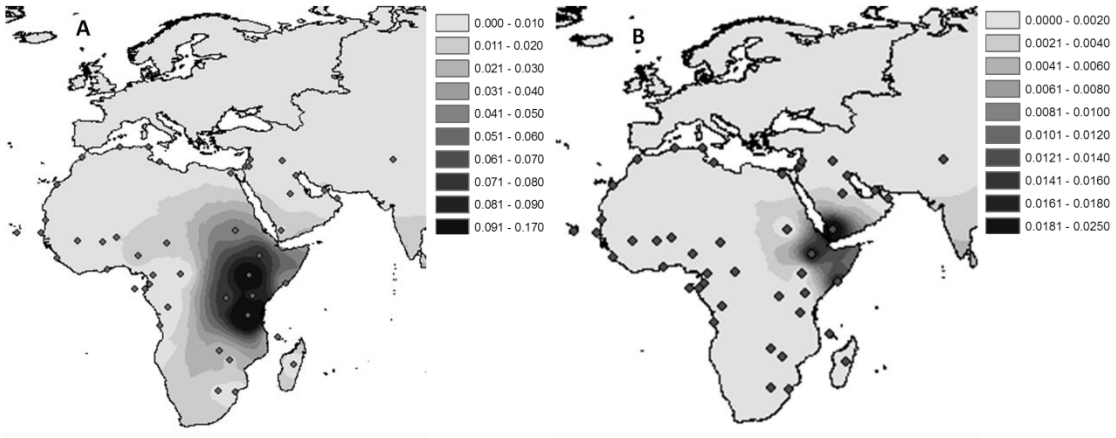

Figure 81 Frequency maps based on HVS-I data for haplogroups L4 (A) and L6 (B).

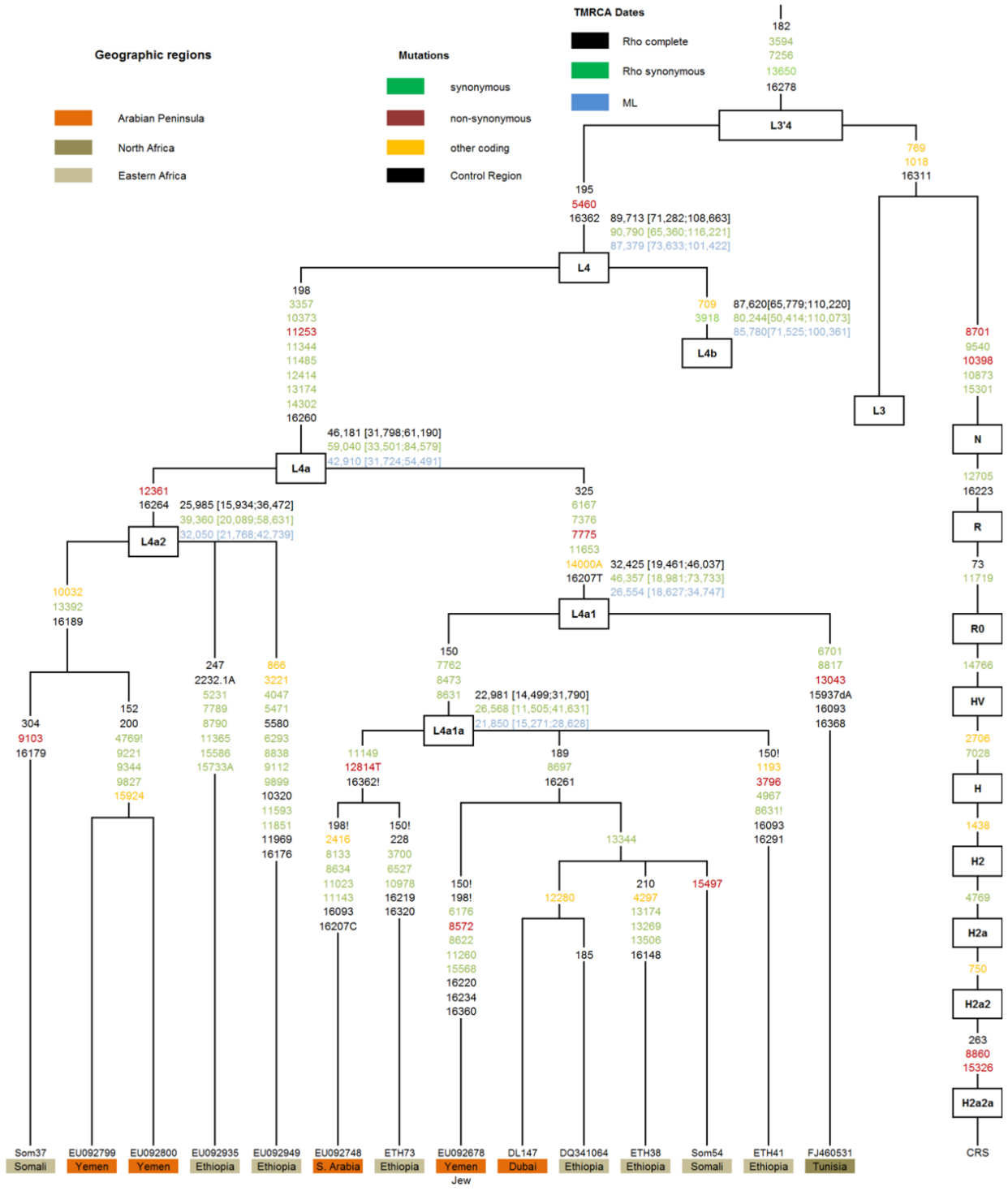

Figure 82 Phylogenetic tree of L4a.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

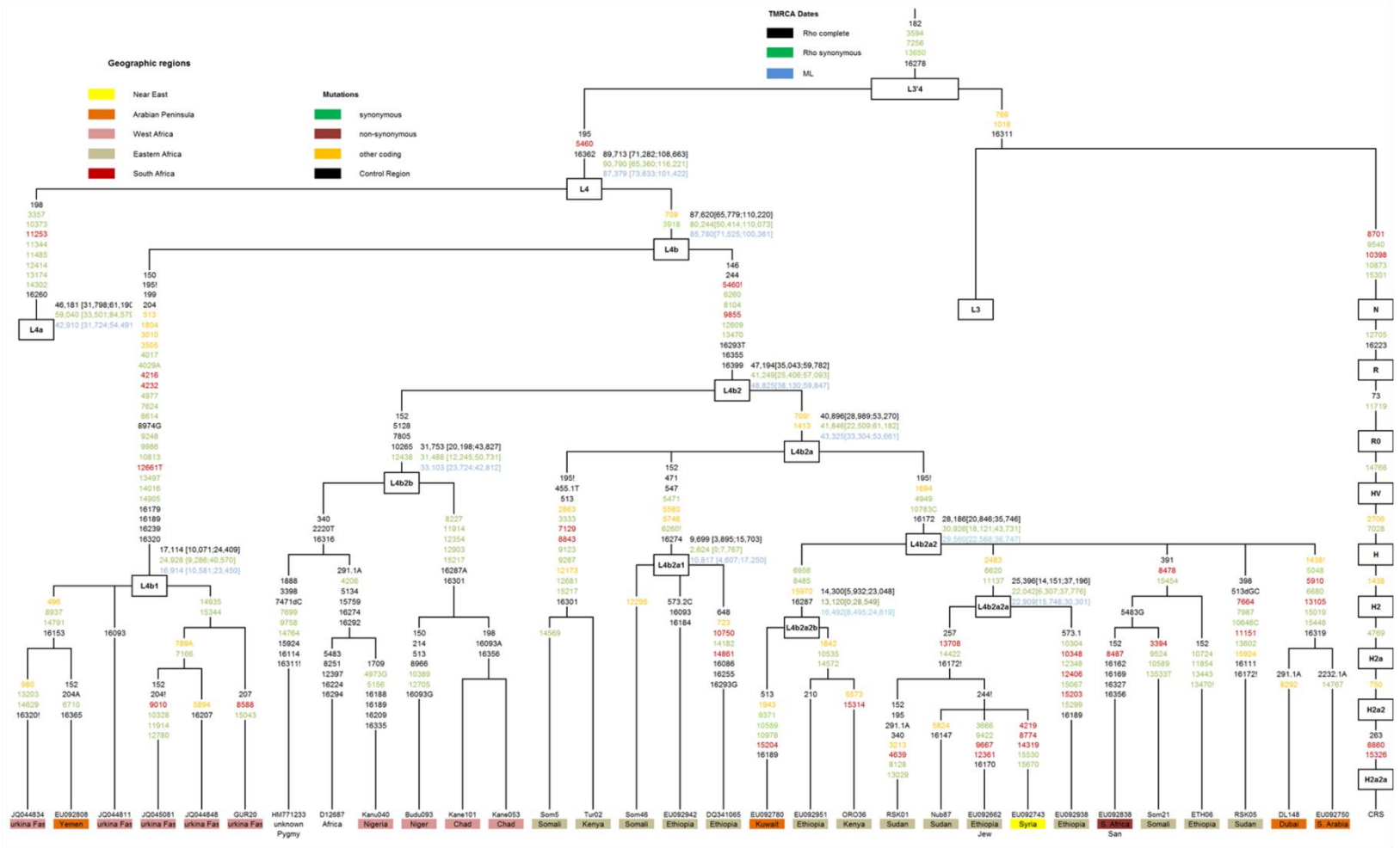

Figure 83 Phylogenetic tree of L4b.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

Haplogroup L6 (Figure 84) is much more recent and dates to 23.1 [15.8-30.5] ka. The L6 sequences in the tree are located in eastern Africa and in Yemen, the first location displaying the highest L6 diversity (Figure 81B) indicating also a probable eastern African origin, with migrations from eastern Africa into Arabia necessarily after 23 ka.

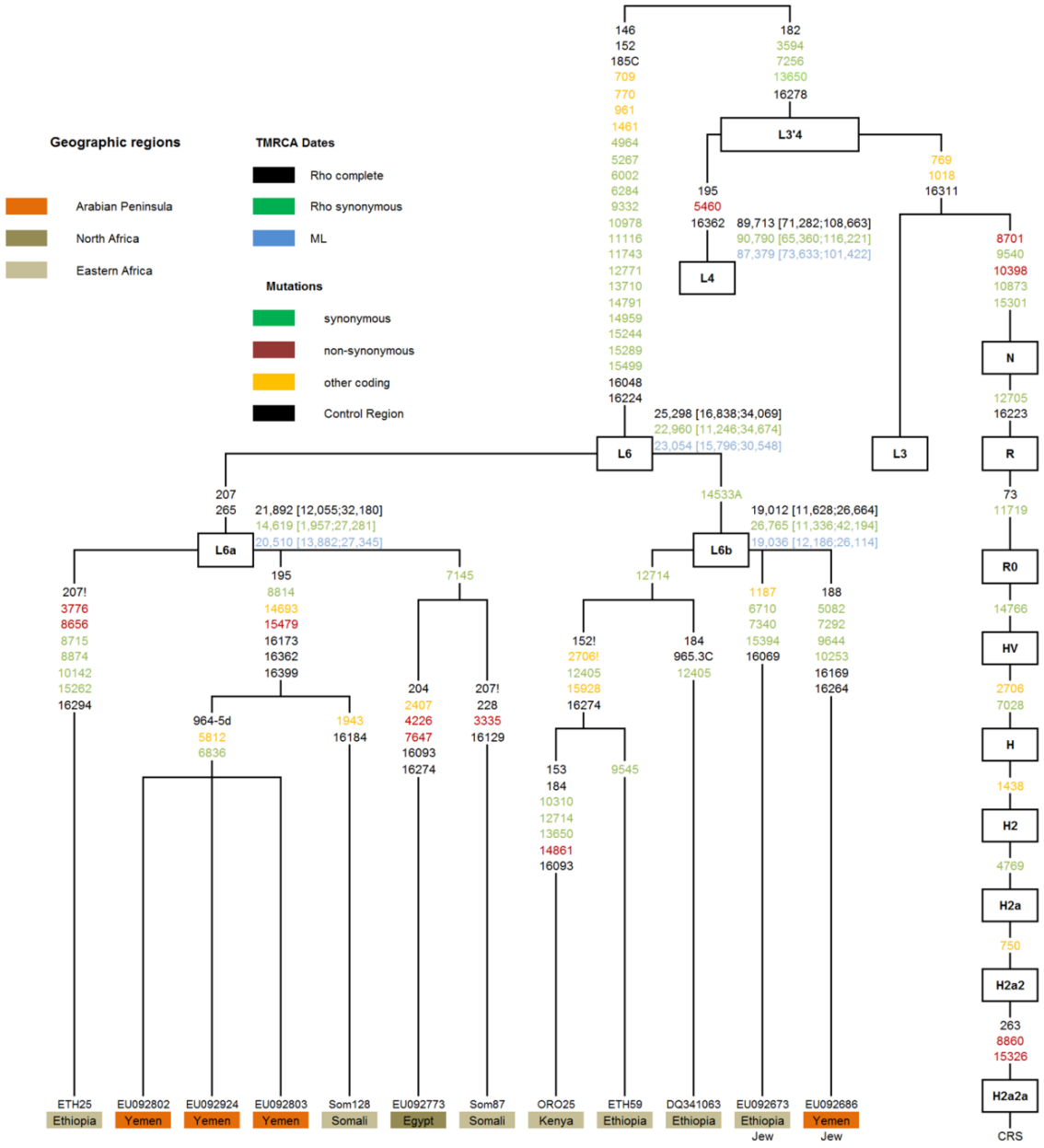

Figure 84 Phylogenetic tree of L6.
Labels on the branches represent nucleotide positions of transitions, and transversions when followed by a suffix “A,” “G,” “C,” or “T”; reversions by “!”; green indicates synonymous, brown non-synonymous, yellow other coding region, and black control region substitutions. Individual identification is indicated as well as the geographic origin when known (geographic regions are groups by a colour code according to the legend). Near the nodes, the TMRCA are indicated (mean and 95% confidence interval) for rho based on complete sequence (in black), rho based on synonymous diversity (in green) and for maximum likelihood (in blue).

The BSP using only the L4 data (Figure 85A) points to three main episodes of population growth. Two periods of increase from after 20 ka almost to the present (Table 7) are observed, with peaks around 9.8 and 18.3 ka. The last period of increase is observed at around 73.8 Ka. The BSP obtained from the overall data (L4 and L6, Figure 85B) shows similar results indicating the poor L6 signal due to the low number of samples.

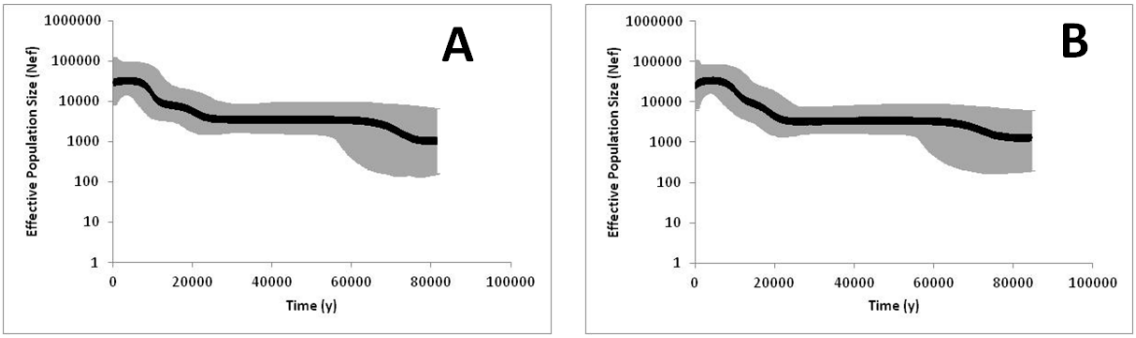

Figure 85 Bayesian Skyline Plot (BSP).
The plot indicates the median of the hypothetical effective population size through time based on data from the haplogroup L4 (A) and haplogroups L4 and L6 (B), assuming a generation time of 25 years.

Table 7 Peaks of rate of population size change through time as obtained from the BSPs and periods of time where the rate of population size increase was of at least one individual per 100 individuals in a period of 100 years.
Increment ratio corresponds to the number of times the effective population size increase during this period.

	Peak	**Range of increment**	**Increment ratio**
L4	9.4	6.9;12.4	3.25
	19.4	18.4; 22.6	1.6
	71.5	68.7; 75.6	2.1
L4_L6	9.9	7.4;12.9	2.8
	18.3	14.5; 21.1	2.4
	73.8	70; 74,6	1.5

9.2. HVS-I founder analysis from Eastern Africa to the Arabian Peninsula

I employed a founder analysis for all lineages of African descent observed in the Arabian Peninsula and Near East, assuming migrations from Africa. My first intention was to do Arabia and Near East separated, but the limited number of L sequences in these regions was statistically limitative. The founders are listed on Tables S11 and S12 for *f1* and *f2* criteria, respectively. The results (Figure 86) showed a single peak of migration at ~1.0 ka using an *f1* criterion (and at ~2.0 ka when a *f2* criterion was applied instead) in the probabilistic distribution of founder clusters across migration times scanned at 200 years intervals from 0–70 ka.

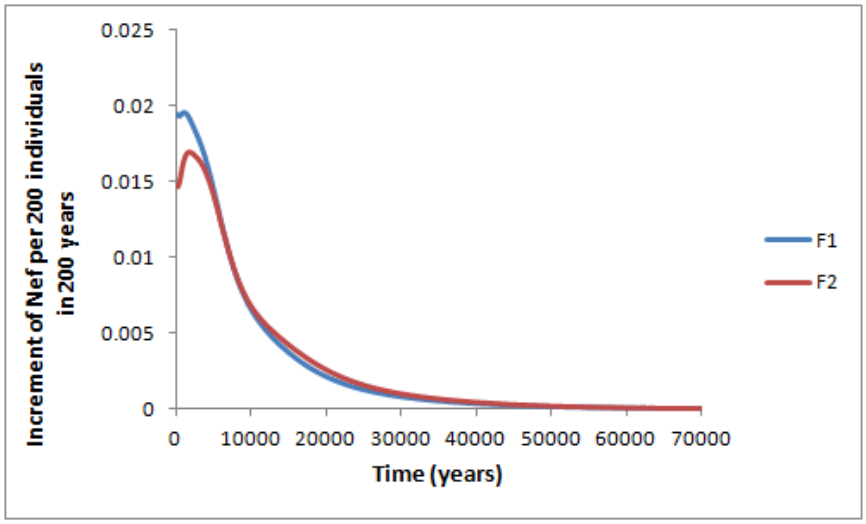

Figure 86 Probabilistic distribution of founder clusters across migration.
Times scanned at 200 years intervals from 0–70 ka, using both an *f1* criterion (blue line) and an f2 criterion (red line), when considering a migration from Africa to Arabian Peninsula plus Near East.

However these single peaks are not completely sharp, showing a slight bump between 10-20 ka. This led me to inspect more carefully the founders (Table S11), and identify some frequent ones which could have been introduced at older times (an L0* founder observed in 33 individuals with an age of introduction of 4.5 ka; an L3f one found in 13 individuals at 10.3 ka; an L3e1 found in 9 individuals at 8.3 ka; and an L0a1'4 found in 9 individuals at 11.1 ka). This could indicate that several migration events could have contributed to this input, with the large number of founders with ρ = 0 (1/4 of the total) is skewing the overall curve to the left. Thus, I investigated the proportion of founders which could have been introduced into

Arabia in four periods of time: 0.5 ka, representing the slave trade initiated by Arabs in the 6th century AD until the 19th century AD (given the peak in the general analysis, I decided to use 0.5 ka, in the middle of the slave trade period, and not 1.5 ka corresponding to the beginning of the trade); the Arabian dominance in the Red Sea trade routes at around 2.5 ka; the beginning of maritime networking in the Red Sea around 8.0 ka (an intermediate point between the 4.4 and 10.0 ka dated for those frequent founders); and the out-of-Africa migration at around 60.0 ka as the oldest possible successful migration of African lineages. Looking to the overall pattern of migrations (Figure 87) into Arabia and Near East, assuming an Africa source, it is possible to observe that the highest input of lineages is detected at 2.5 ka and 8.0 ka, with a proportion of 38% and 34%, followed by 27% at 0.5 ka and a residual 0.6% at 60.0 ka.

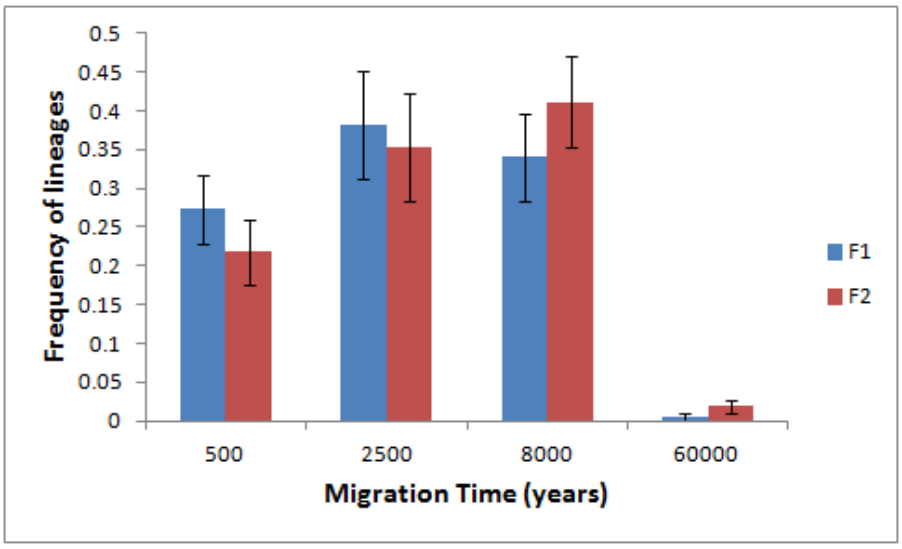

Figure 87 Probabilistic proportion of founder clusters considering four migrations periods (0.5, 2.5, 8.0 and 60.0 ka).
By using both an *f1* criterion (blue line) and an *f2* criterion (red line) when considering a migration from Africa to Arabian Peninsula and Near East.

The *f2* criterion, which leads to older ages for the introduction of lineages redistributes the lineages, increasing the proportion in the peak at 8.0 ka (41%), in detriment of the 2.5 (35%) and 0.5 (22%) ones; the value at 60.0 ka continues to be residual (1.8%).

Figure 88 displays the probabilistic proportion of introduction for each lineage at each of the four migrations periods, when using the *f1* criterion. The most frequent lineages and with probabilities higher than 50% of having been introduced during the 0.5 event are affiliated in L6, L0a1, L3b and several low frequently L2a lineages. At the 2.5 ka event the most frequent contributing lineages are several L3 ones

(L3b, L3d1a, L3e3), L2a1a2, L0a and L5, with probabilities higher than 50%. For the 8.0 ka event, with probabilities higher than 75%, L0a, L3e1, L3f1b and several low frequent L2a lineages. For the out-of-Africa event, the values are always residual for a few individuals, testifying that most probably none of the observed lineages were introduced in this event. No clear pattern of association between haplogroup and event is observable, probably reflecting the high heterogeneity in the source. The pattern for the *f2* criterion (Figure 89) is similar to the one observed for the *f1*.

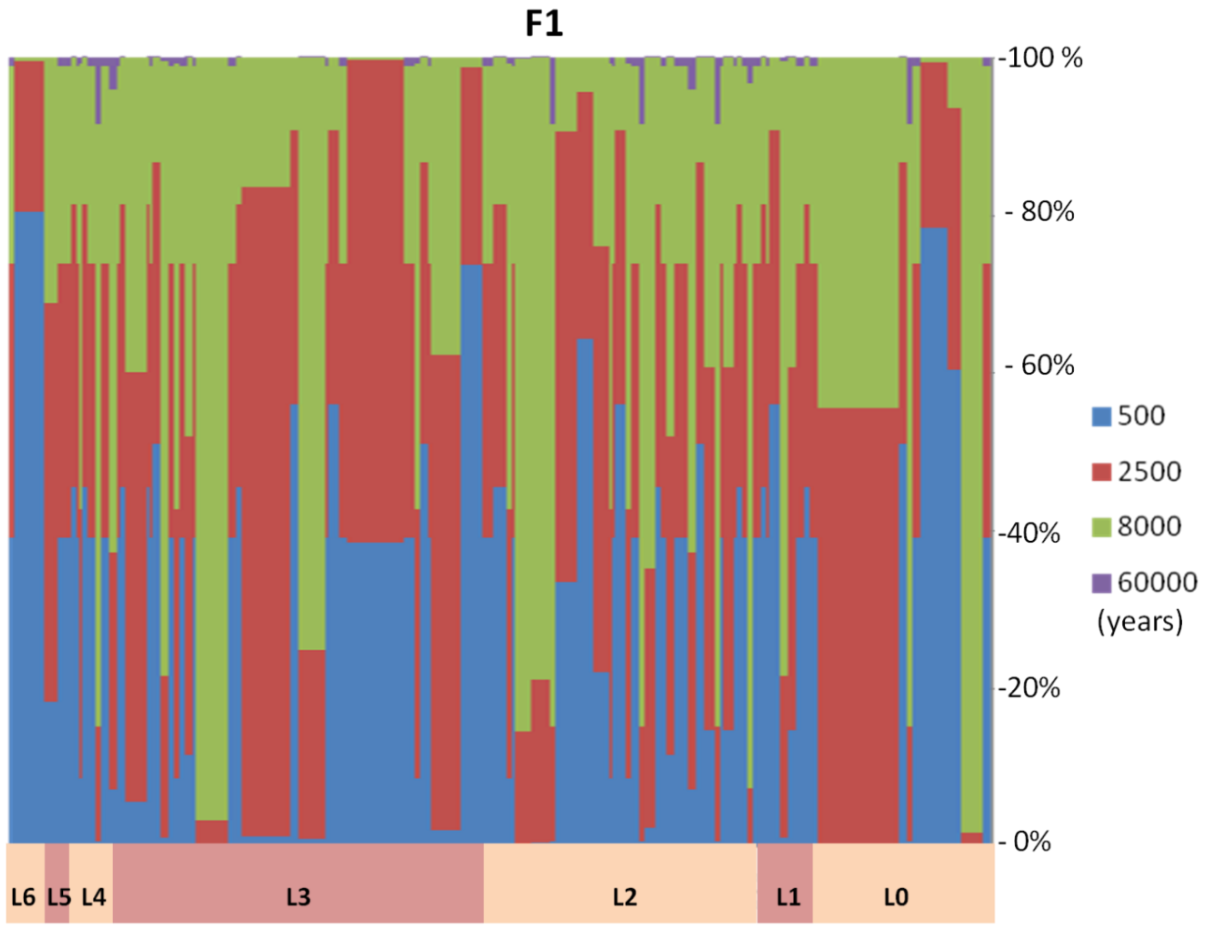

Figure 88 Probabilistic proportion of founder clusters considering four migrations periods (0.5, 2.5, 8.0 and 60.0 ka).
By using a *f1* criterion and by assuming an African source for migrations into Arabian Peninsula plus Near East. The haplogroup affiliations of the founders are indicated in the bottom.

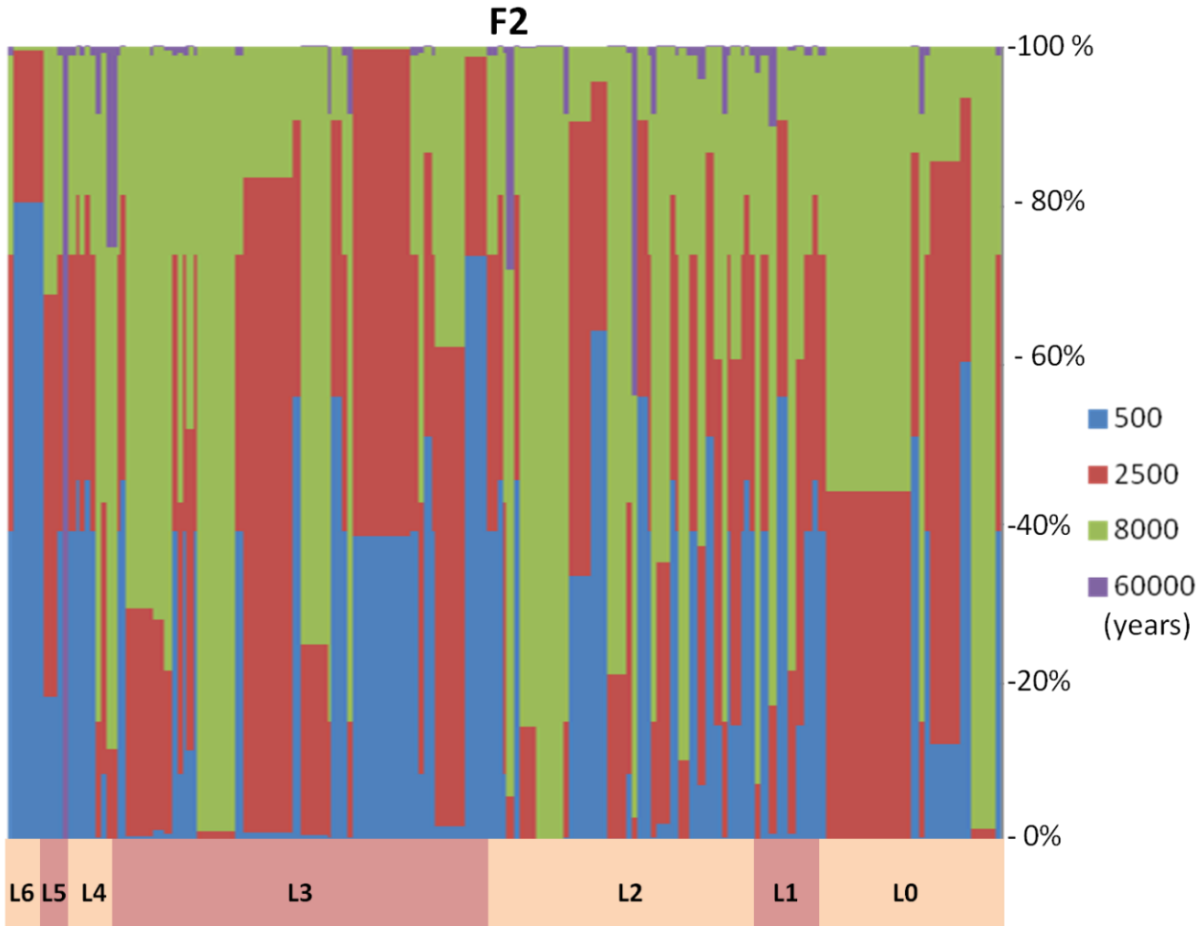

Figure 89 Probabilistic proportion of founder clusters considering four migrations periods (0.5, 2.5, 8.0 and 60.0 ka).
By using a *f2* criterion and by assuming an African source for migrations into Arabian Peninsula plus Near East. The haplogroup affiliations of the founders are indicated in the bottom.

9.3. HVS-I founder analyses from Arabian Peninsula and Near East to Eastern Africa and North Africa

By employing founder analysis, I investigated all the lineages of Eurasian descent observed in Eastern Africa, considering Arabian Peninsula and Near East as source (Figure 90A). A total of 57 and 41 founder lineages were identified when using the *f1*and *f2* criteria, respectively, observed in 268 individuals (Tables S13 and S14). One defined peak of migration was detected at ~10.2 ka for the *f1* criterion, and another at ~15.0 ka with the *f2* criterion. Likewise, when considering an Arabian Peninsula and Near East source but migrations into North Africa (Figure 90B) two well defined peaks at ~2.4 ka and 6.8 ka were obtained with the *f1* criterion and one defined peak at ~9.0 ka and a small one at 12.4 ka with the *f2* criterion. The founders are summed up in Tables S13 and S14.

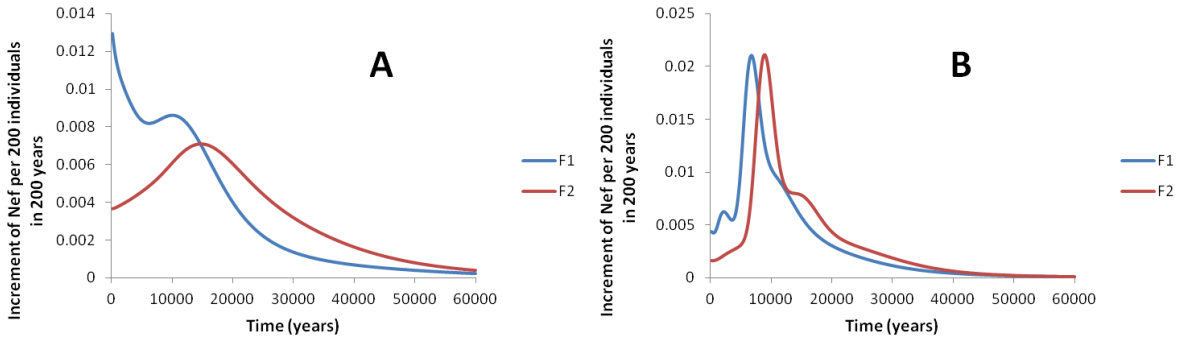

Figure 90 Probabilistic distribution of founder clusters across migration. Times scanned at 200 years intervals from 0–60 ka, using both a *f1* criterion (blue line) and a *f2* criterion (red line), when considering a migration from Arabian Peninsula and Near East to East Africa (A) and to North Africa (B).

Given those distributions, I then assayed four migration periods: at 2.0 ka for the Arabian dominance in the Red Sea trade routes and the spread of Islam; at 8.0 for the Neolithic; at 16.0 ka for the Late glacial period; and 45.0 ka that is the proposed date for the back-to-Africa migration of U6 and M1.[188] For the *f1* criterion, the proportions of lineages introduced to eastern Africa, from the Near East and Arabia, were 29% for 2.0 ka, 34% for 8.0 ka; 31% for 16.0 ka and 6% for 45.0 ka (Figure 91); proportions were higher for older dates when using the *f2* criterion, of 13%, 24%, 48% and 14%, respectively.

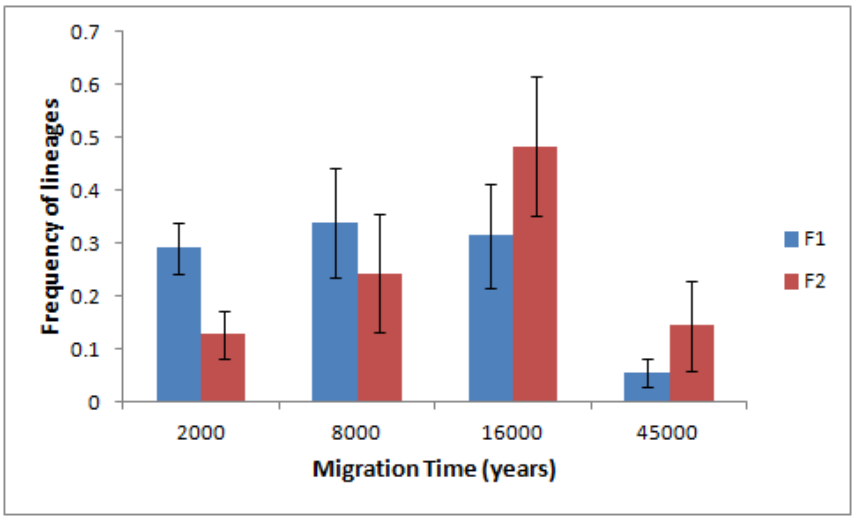

Figure 91 Probabilistic proportion of founder clusters considering four migrations periods (2.0, 8.0, 16.0 and 45.0 ka).
By using both a *f1* criterion (blue line) and a *f2* criterion (red line) when considering a migration from Arabian Peninsula and Near East to East Africa.

The probabilistic proportions for the introduction of each lineage at each of the four migrations periods, when using the *f1* criterion, are represented in Figure 92. The M1 lineages seem not to have reached East Africa in the 45.0 ka migration (and even in the f2 criterion the probability is never higher than 20%). This event led to the introduction of N2a, W (70-96% probability) and some U5 (62%) lineages. M1 had 60% of probability of having entered in the Late glacial as well as HV0 (50% probability). Around 8.0 ka, some U lineages (U6a1a), J1d1a, M1 (M1a1) and R0a had each 50% probability of entrance. At 2.0 ka, N1 (60-80%), many R0a (50-100%), T (60-70%), J (50%) and a few U and X (80%) arrived most probably in East Africa from the Near East and Arabia. The pattern for the *f2* criterion (Figure 93) is similar to the one observed for the *f1* criterion.

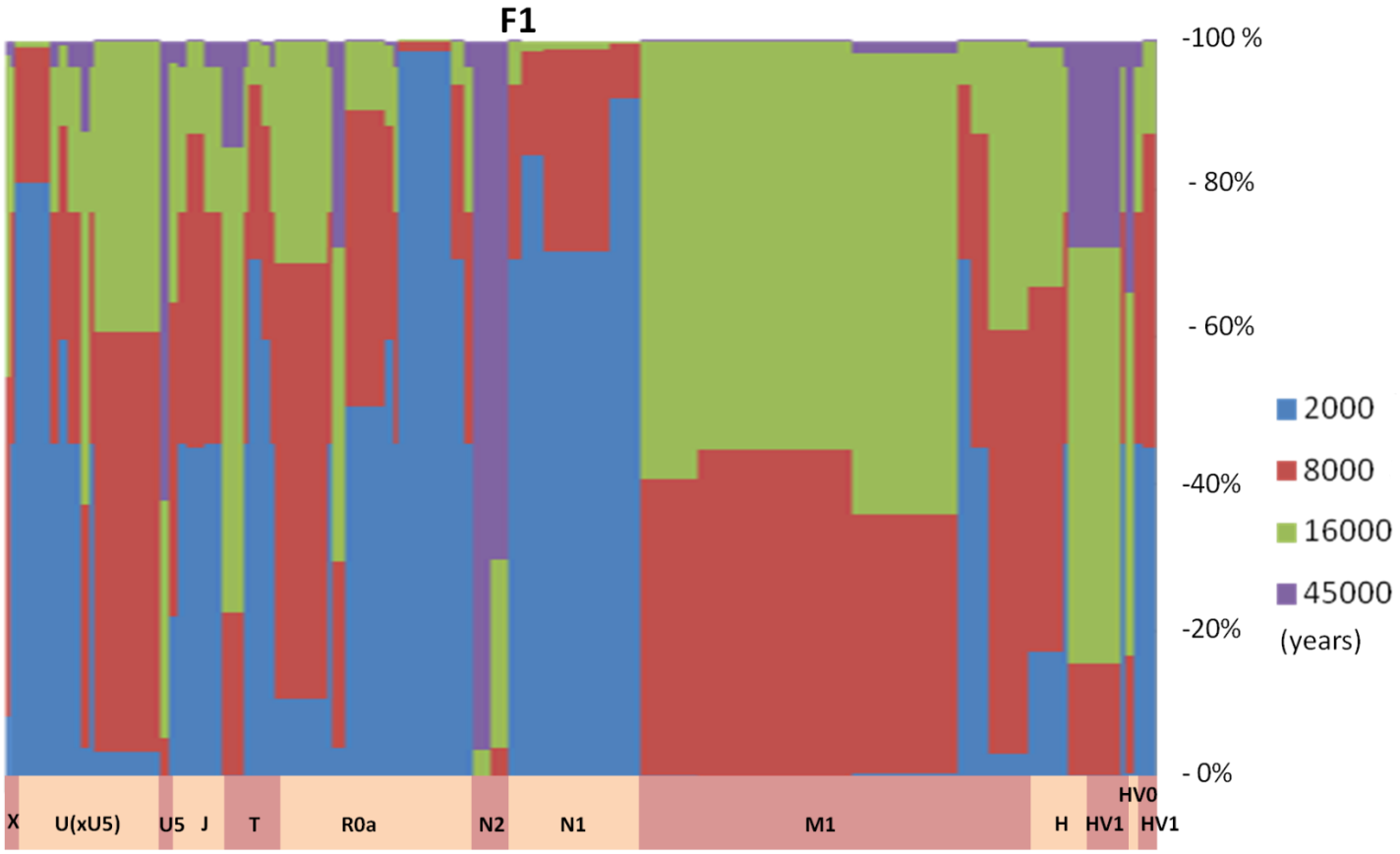

Figure 92 Probabilistic proportion of founder clusters considering four migrations periods (2.0, 8.0, 16.0 and 45.0 ka). Using a *f1* criterion and by assuming an Arabian Peninsula plus Near East and migrations into East Africa.

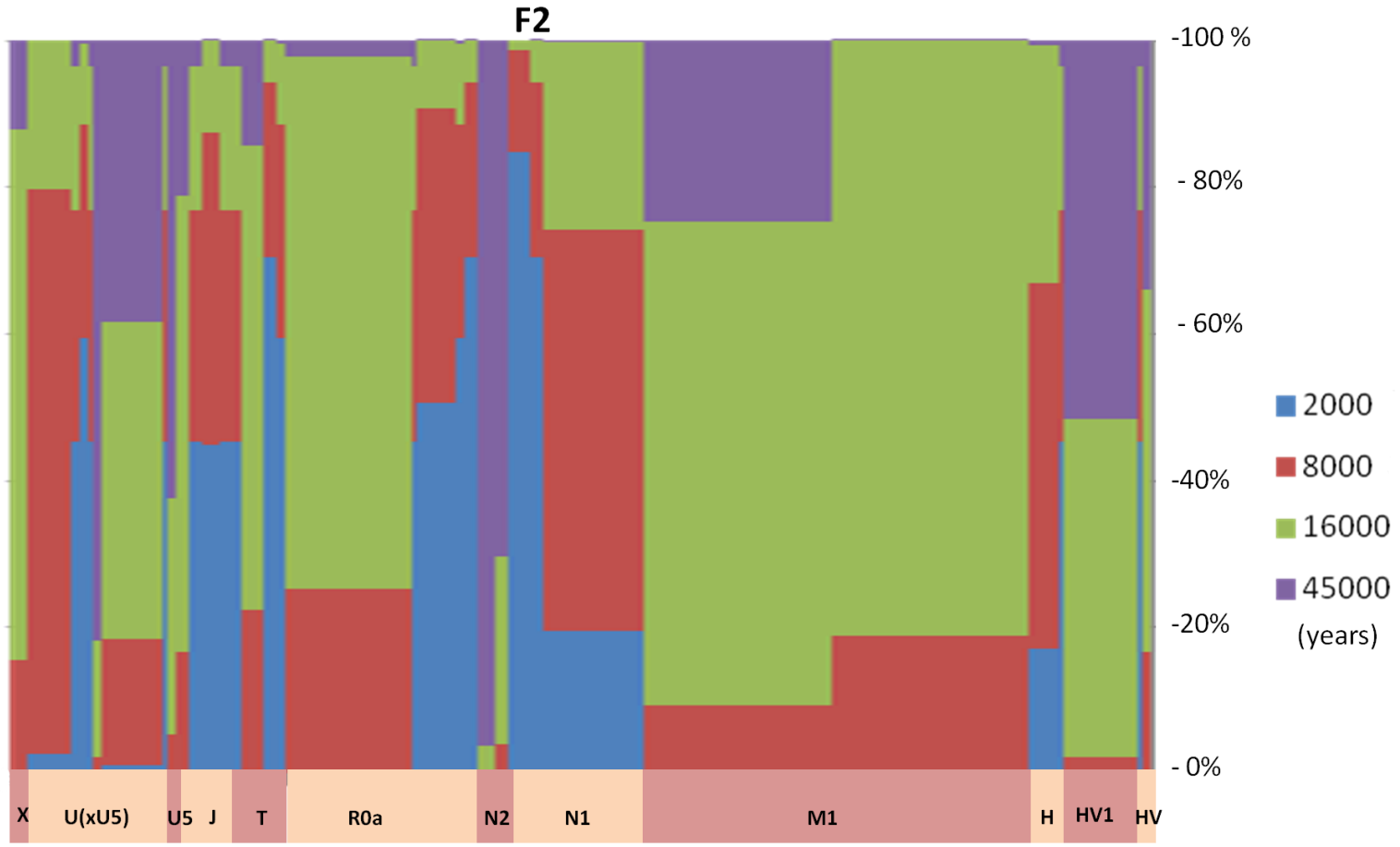

Figure 93 Probabilistic proportion of founder clusters considering four migrations periods (2.0, 8.0, 16.0 and 45.0 ka). Using a *f2* criterion and by assuming an Arabian Peninsula plus Near East and migrations into East Africa.

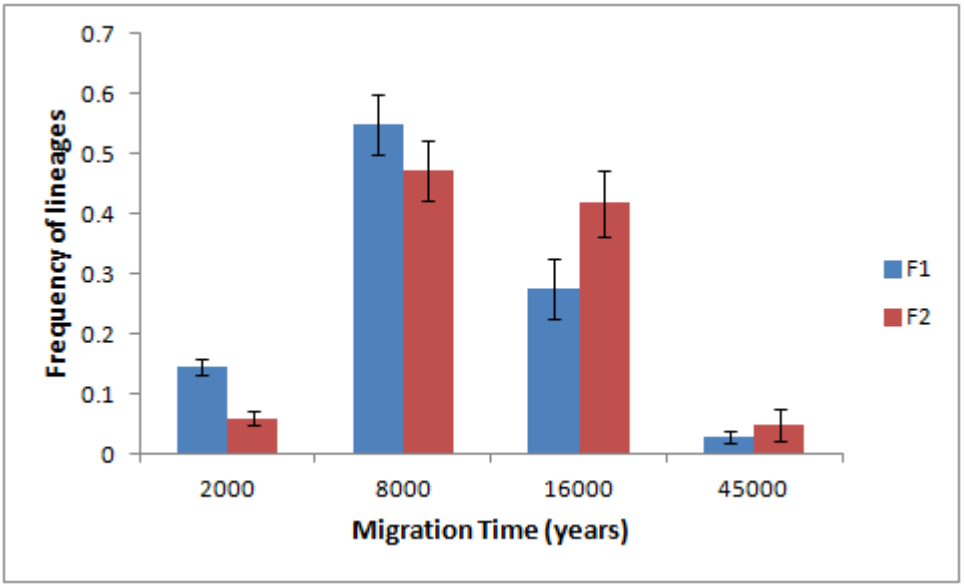

Figure 94 Probabilistic proportion of founder clusters considering four migrations periods (2.0, 8.0, 16.0 and 45.0 ka).
By using both a *f1* criterion (blue line) and a *f2* criterion (red line) when considering a migration from Arabian Peninsula and Near East to North Africa.

In the case of migration from the Near East and Arabia to North Africa (Figure 94), a total of 215 and 118 founder lineages were identified when using the *f1*and *f2* criteria, respectively, observed in 1931 individuals (Tables S13 and S14). The highest difference is in the impact of migrations around 8.0 ka, explaining 55% of the lineages, compared to 15% at 2.0 ka, 28% at 16.0 ka and 3% at 45.0 ka, for *f1* criterion. The *f2* criterion diminishes the importance of that period (47%), being closer to the 16.0 ka (42%).

Again, the probabilistic proportions of introduction for each lineage at each of the four migrations periods, when using the *f1* criterion (Figure 95) does not attribute great importance to the 45.0 ka period, with few and rare U (U3, U5, U5a1 and U6) lineages having between 70-80% probability. At 16.0 ka some HV1 and other HV lineages (45-90% probability), T (T1a, T2; 98%) and U (U3, U3a, U5b1b, U5a, U6a; 60-95%) lineages seem to have entered North Africa. At 8.0 ka, most HV (probability 50-100%), U (U5b, U5 and K; 70%), T (some T2c1 and T2b; 80%) and J (J1d1a, J2a2b and other undefined; 90%), and X (60%) moved into North Africa. While at 2.0 ka, some HV lineages, M1 and other U (U6a1, K1a1; 90-100%) arrived in the North of the African continent. The pattern for the *f2* criterion (Figure 96) favours migration at 16.0 ka for most of the lineages that not HV, which keeps the highest probability of introduction at 8.0 ka.

These results seem to show that the Neolithic was the dominant period for the introduction of maternal lineages into North Africa. It is curious that this impact was greater (double) in North Africa (47-55%) than in Arabia (24-34%). Regarding the genetic impact of the Islamisation, this seems to have been minor (6-15%) in the North Africa, at least for the female lineages.

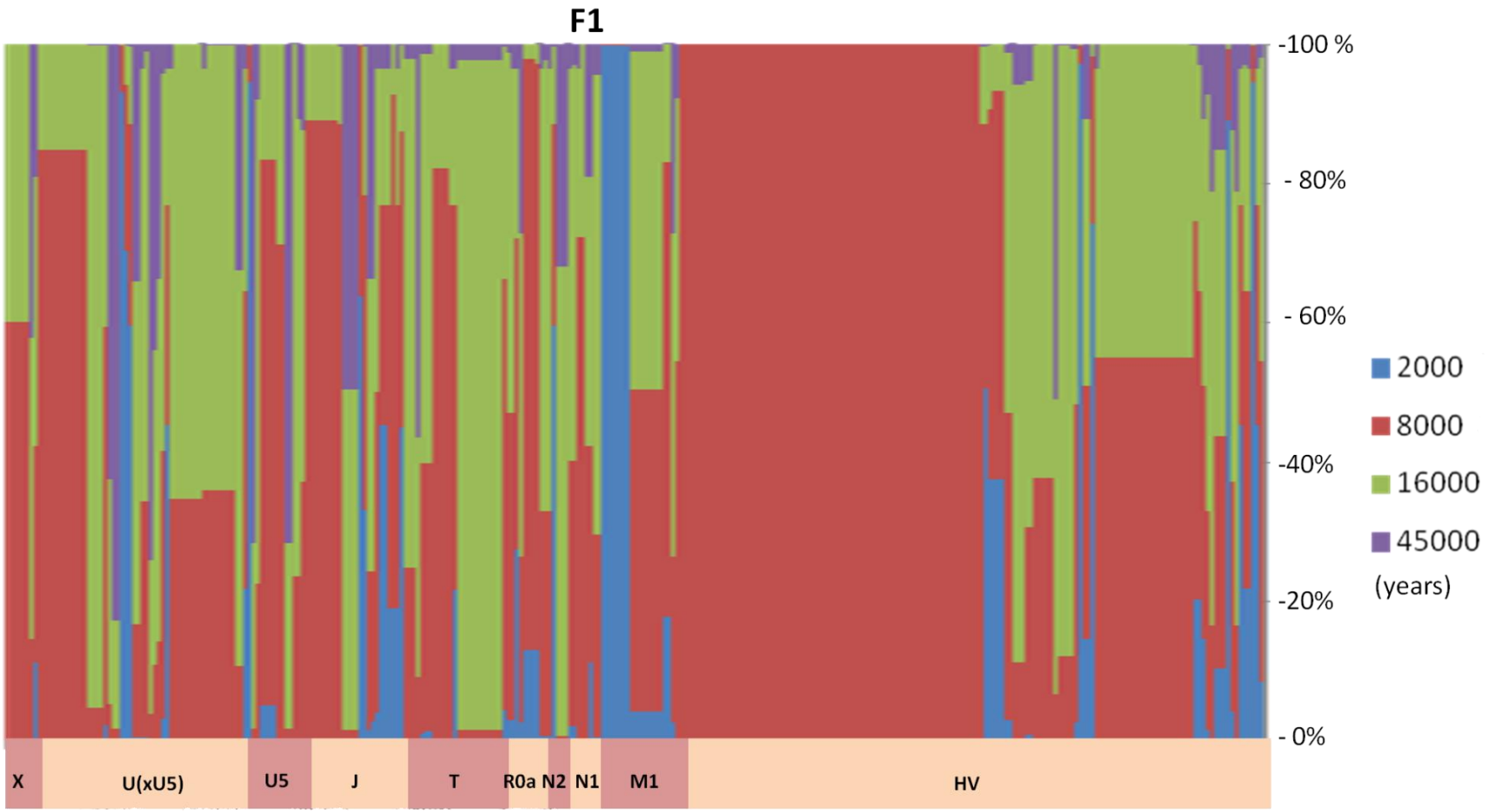

Figure 95 Probabilistic proportion of founder clusters considering four migrations periods (2.0, 8.0, 16.0 and 45.0 ka). Using a *f1* criterion and by assuming an Arabian Peninsula plus Near East and migrations into North Africa.

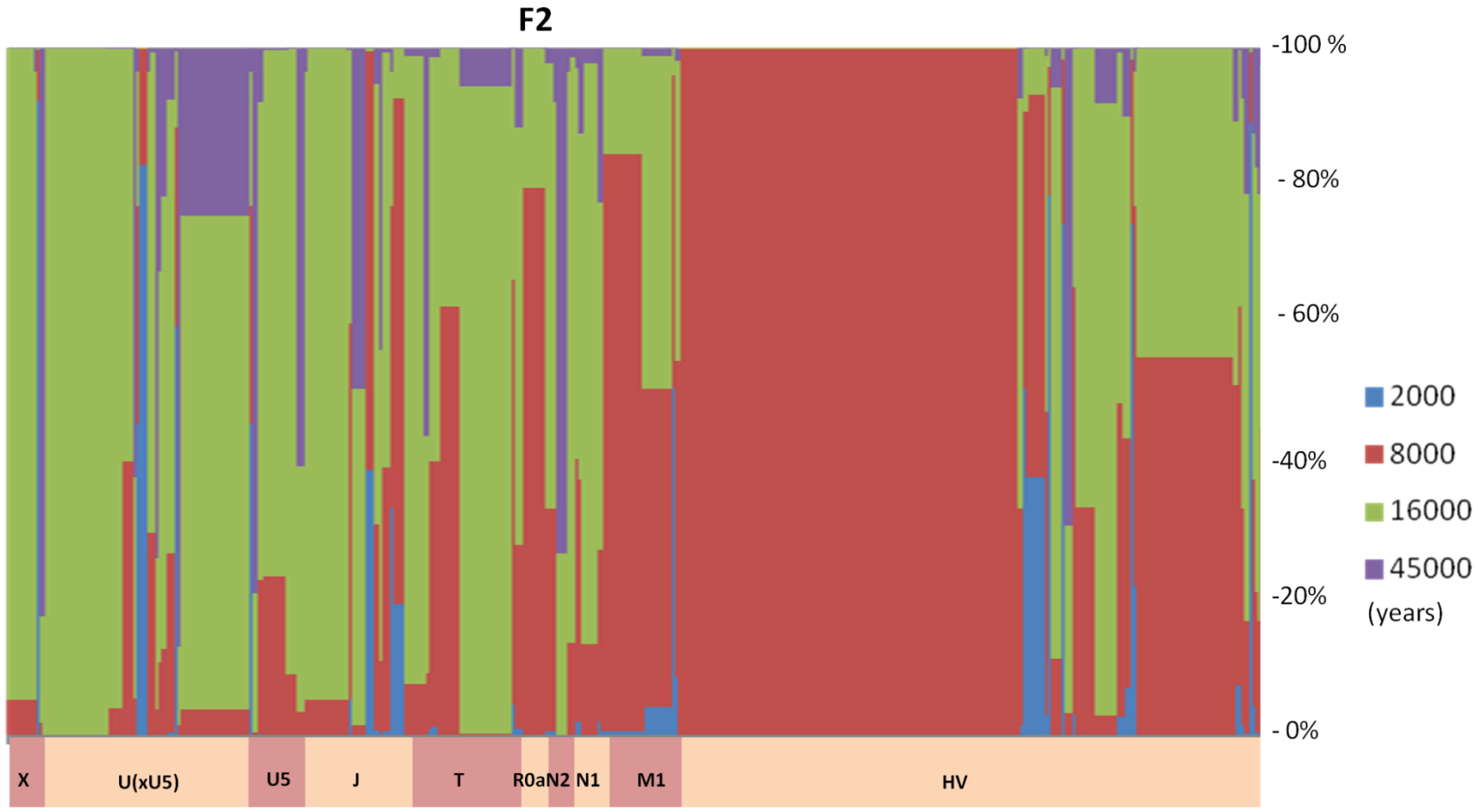

Figure 96 Probabilistic proportion of founder clusters considering four migrations periods (2.0, 8.0, 16.0 and 45.0 ka). Using a *f2* criterion and by assuming an Arabian Peninsula plus Near East and migrations into North Africa.

10. Population structure and genetic exchanges between Arabia and neighbours on genome-wide information

The PCA revealed that the two first axes explain 53.9% of the diversity, with a considerable proportion of 46.7% being displayed by axis 1 (Figure 97). This main axis splits sub-Saharan African populations from the other populations, while the 7.2% of diversity explained by PC2 arranges populations in a west-east gradient. There are two Yemeni, one Saudi and one Jordanian, as well as two Pakistani (one Sindhi and one Makrani), which are closer to the African populations testifying the higher African admixture in these individuals than the average of their population groups. When focusing attention on the Arabian Peninsula and Near Eastern populations, the Bedouin have the highest dispersion across the second axis, compatible with their nomadic life style (probably contributing to higher admixture with other populations), even higher than the one displayed by Saudi Arabia, which constitutes the bulk of the Arabian Peninsula. Yemen shows the highest dispersion along the first axis, testifying again the higher African input in the closest country to the Horn of Africa. I confirmed the clustering of Yemeni Jews with Bedouin and Saudi Arabians, already identified by Behar et al.[136], and probably indicating that they were less open to recent admixture with non-Arabian populations than their Yemen neighbours. Palestinians are closer to Egyptian samples, while Druze and Samaritans are tightly clustered and near to Syrian, Jordan and Lebanese samples. Iranians and especially Turks are closer to the European populations, while Pakistanis are clustered together farther away from the Arabian Peninsula and Near Eastern groups.

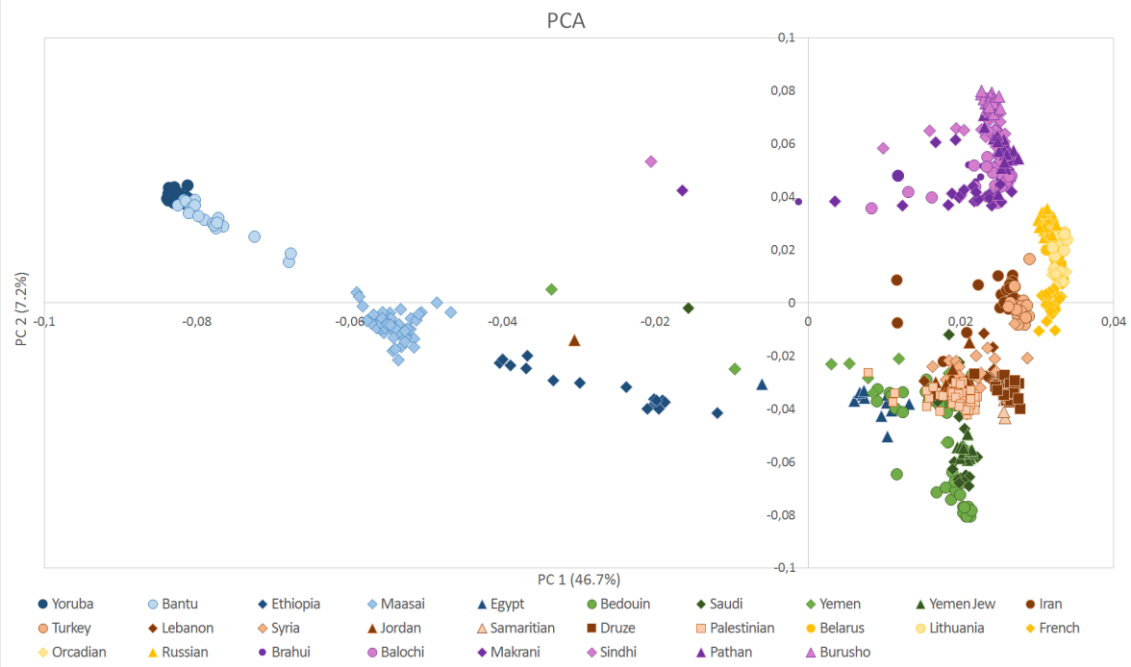

Figure 97 Principal Component Analysis (PCA).

The ADMIXTURE results indicate K=5 as the number of clusters having the lowest cross-validation error (Figure 98), meaning that this is the number of clusters that best represents the population structure of the analysed populations. This K=5 structures the populations in the following clusters (Figure 99): European; Asian; West African; East African; and a combined cluster of Arabian Peninsula and Near East. Egypt is more similar to this last cluster than to the other East African sub-Saharan populations, although Ethiopia still displays a high proportion of the Near East/Arabian Peninsula component. The proportion of the European component is higher in the Levant than in the Arabian Peninsula; and the West European populations display a higher proportion of Near East influence than the Central European ones, being replaced by an Asian component in Russians. The Asian/Pakistani component is also observable in the joined Arabian Peninsula and Near East cluster, being higher in the geographically closer populations; a similar Near East/Arabian Peninsula proportion is observed in the Pakistan populations. Interestingly, both Bedouins and Saudi Arabians (from two different publications; Li et al.[134] and Behar et al.[136], respectively), are divided in two groups, one less admixed and mainly made of the Near Eastern/Arabian component and with around 5% of East African genes, while the other more admixed also presenting European, Asian and West African genes.

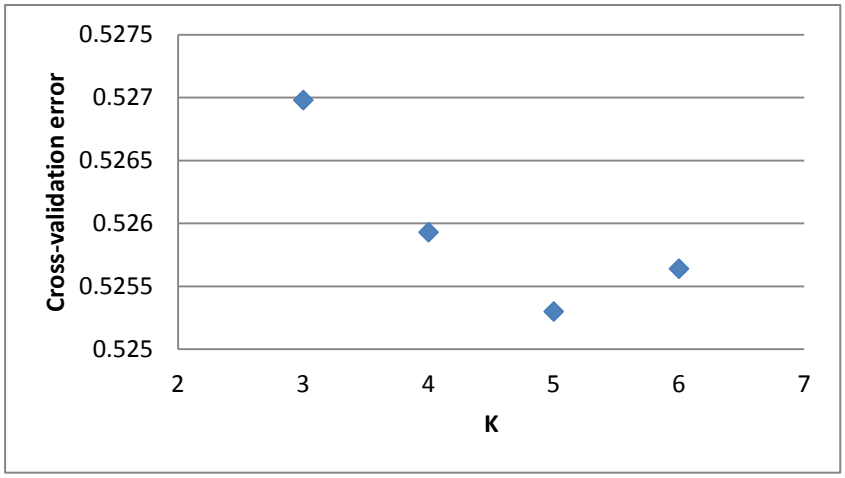

Figure 98 Cross-validation error for diverse K in the ADMIXTURE analysis.

The other *K* clusters (Figures 100-102) show that when assuming three ancestral populations, populations are grouped as African, Near Eastern/Arabian and European/Pakistan; K=4 separates Europe from Pakistan; and K=6 splits Pakistan populations in two groups, one with residual East Asian ancestry and still some African input (Brahui, Balochi and Makrani) are another with moderate levels of East Asian ancestry (Sindhi, Pathan and Burusho), as already observed by Li et al.[134]

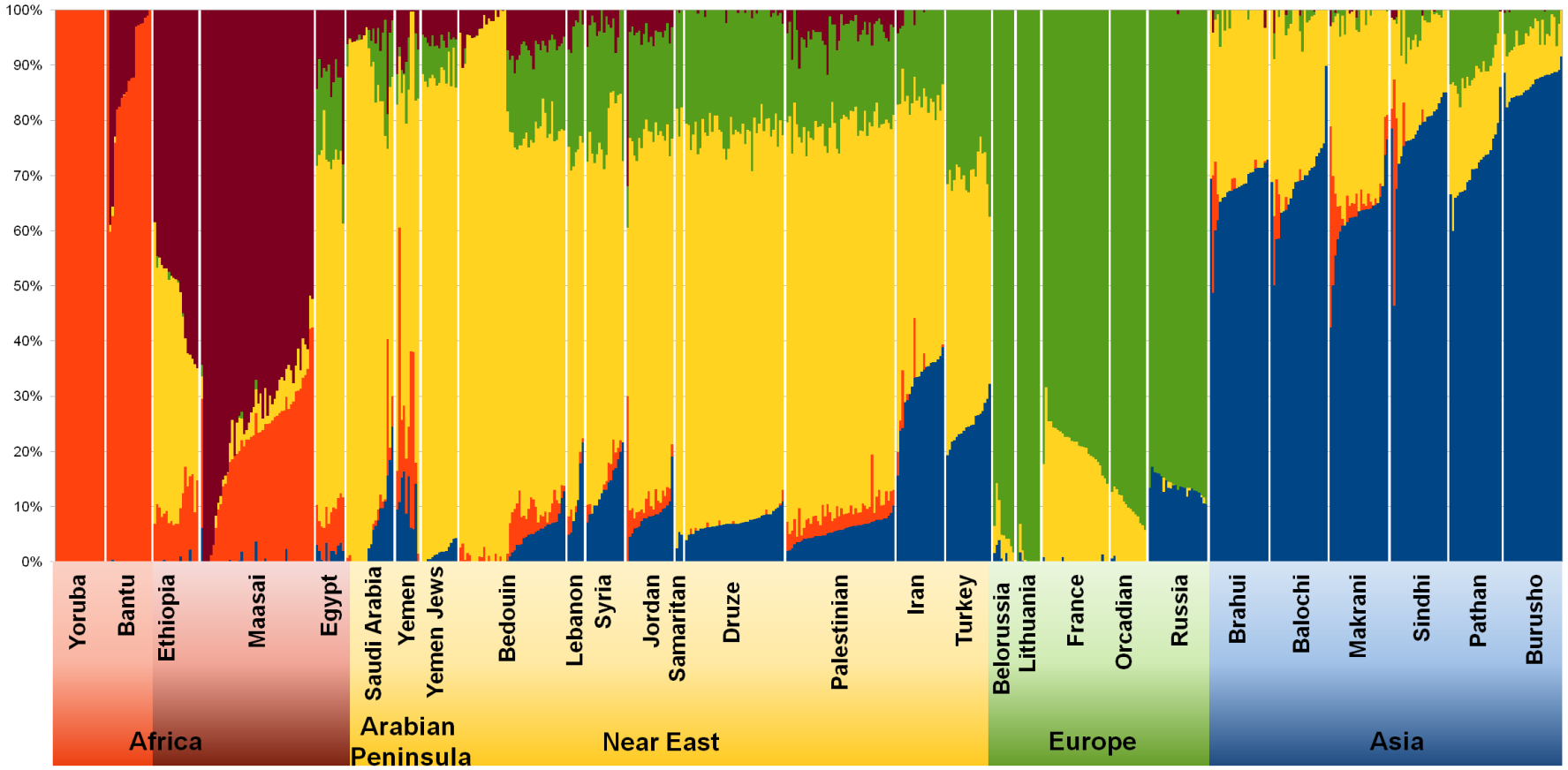

Figure 99 Population structure inferred by ADMIXTURE analysis.
Each individual is represented by a vertical (100%) stacked column of genetic components proportions shown in colour for K=5.

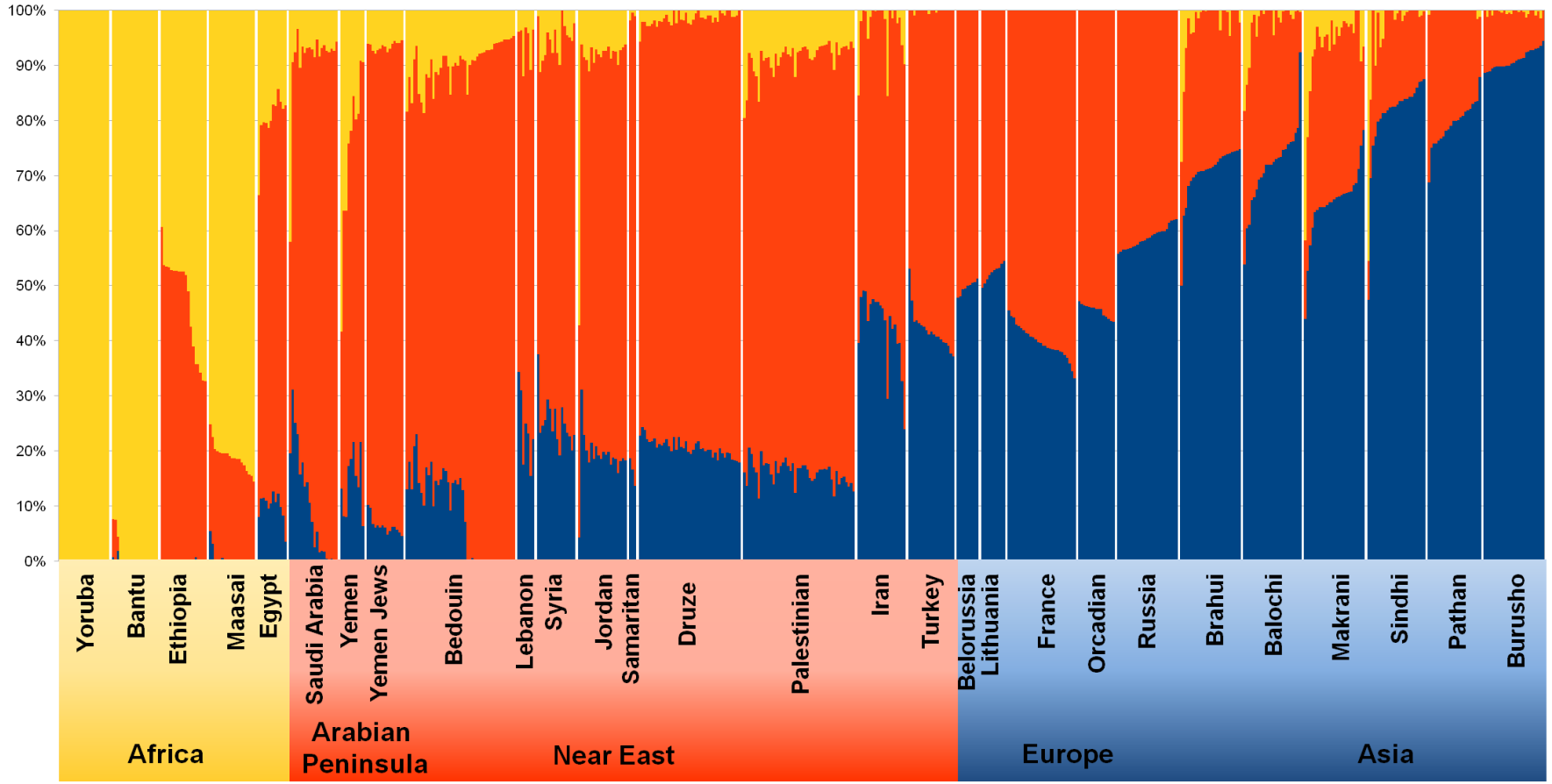

Figure 100 Population structure inferred by ADMIXTURE analysis
Each individual is represented by a vertical (100%) stacked column of genetic components proportions shown in colour for K=3.

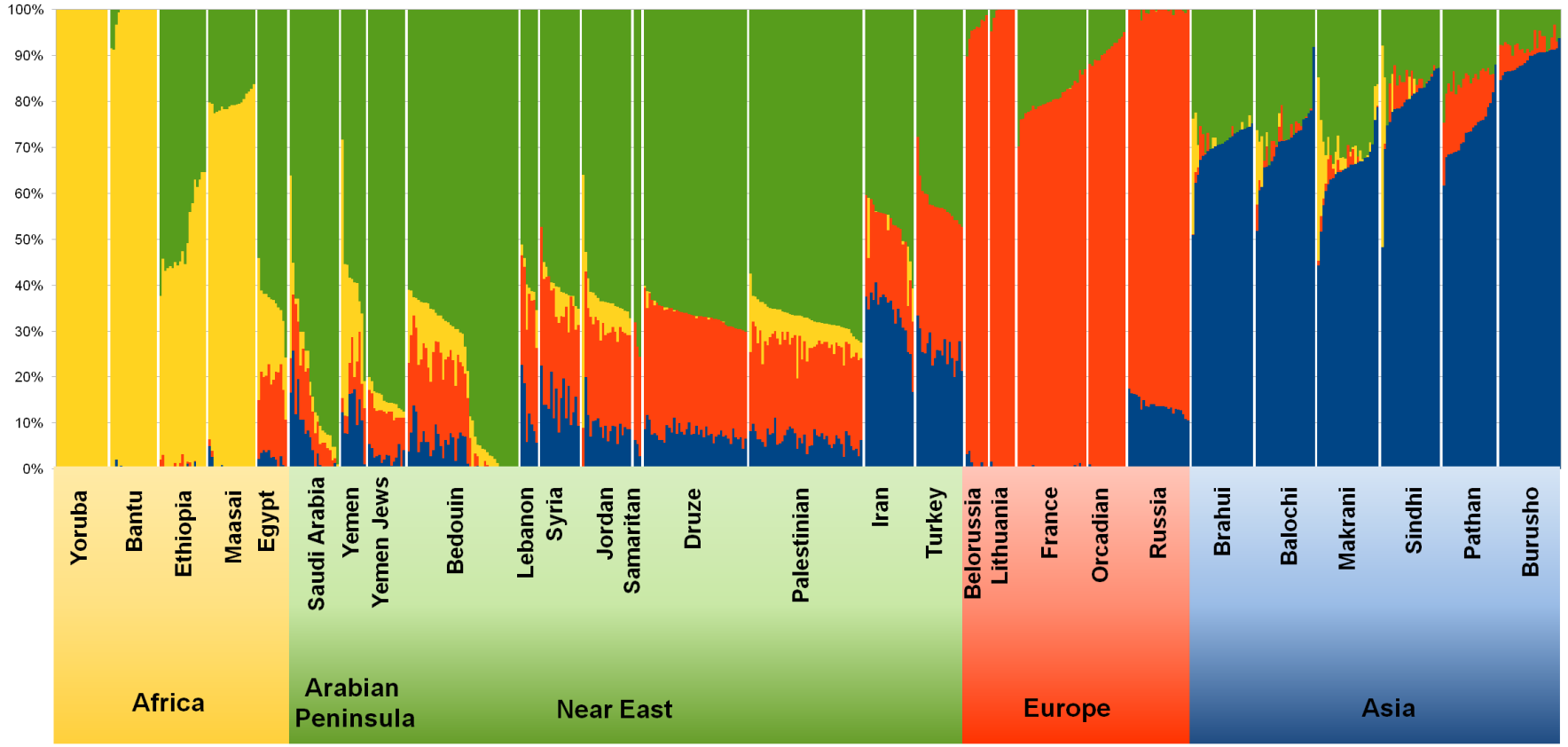

Figure 101 Population structure inferred by ADMIXTURE analysis.
Each individual is represented by a vertical (100%) stacked column of genetic components proportions shown in colour for K=4.

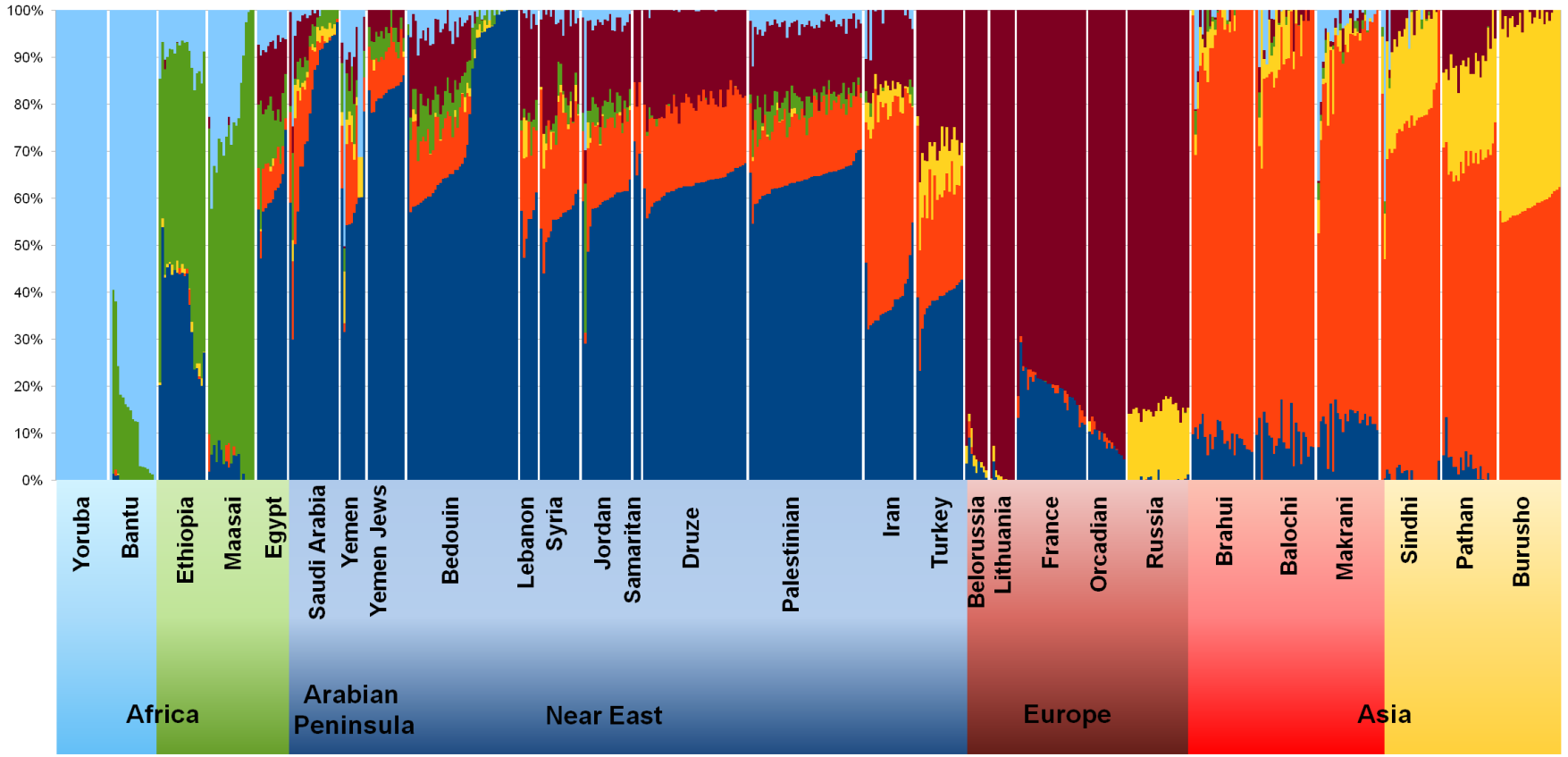

Figure 102 Population structure inferred by ADMIXTURE analysis.
Each individual is represented by a vertical (100%) stacked column of genetic components proportions shown in colour for K=6.

ADMIXTURE allows one to measure the percentage of admixture (summed up in Table 8) and ROLLOFF allows one to estimate the date of the event (as can be seen in Figures 103 and 104, the fitting of the exponential LD decay curves in our data was quite robust, hence the standard errors for the estimated date of admixture reported on Table 8 were small). Focusing on the African input into the Arabian Peninsula and the Near East (Table 8), the K=5 admixture results indicate values of about 5.3% for the East African component and 2.1% for the West African one in Saudi Arabia, and a ROLLOFF estimate for the admixture event of 31 generations ago (Figure 103C and Table 8). Notice that there was a higher heterogeneity in the western African input in Yemen (16.7% with 5.0% of standard error), probably due to the very small sample size and to recent migration of some individuals, such as the one individual also identified in the PCA, which led to the reduction of the estimate of admixture to 18 generations ago; if this individual is removed from the analysis, values are of 13.1% (3.8% of standard error) western proportion and 21 generations since the event of admixture (Figures 103A and B, and Table 8). Nonetheless, Yemen is the country with the highest African input in southwest Asia (around 20% in total). Bedouins are again very similar to Saudi Arabians, in terms of African admixture, with an estimate of ~38 generations ago for the event (Figure 103D and Table 8). Yemen Jews had an East African proportion (5.6%) similar to Yemen (6.7%), but practically no Western African influence, probably due to a less recent admixture than the Yemen population. The African input in Palestinians and Jordanians (7.5% and 9.8%, respectively) is identical to the ones observed in the Arabia, but decreases in Syrians and Lebanese (4.6% and 5.2%, respectively), with estimates for the admixture event around 30 generations ago (Figure 103E-G and Table 8). The African input is very low in Turkish (2.3%) and minimal in Iranians (0.0%), Druze (0.5%) and Samaritans (0.2%).

Concerning the Arabian/Near Eastern input into East Africa (Table 8), Ethiopians display 37.2% of this component, considerably higher than the West African proportion of 10.1%, with minimal integration of European/Pakistan genes (0.4%). Maasai from Kenya show a considerably lower Arabian/Near Eastern input, of 3.8%, counterbalanced by a higher input from West African/Bantu genes, of 21.7%. On the other hand, Egypt is quite similar to its Levantine neighbours, displaying slight, higher West (7.4%) and East (13.1%) African proportions, in detriment of the European (14.0%) and Asian (1.8%) proportions. The ROLLOFF estimates for the admixture events in these African populations were 93 generations for Ethiopia, 47 for Maasai and 30 for Egypt (Figures 104A-C and Table 8).

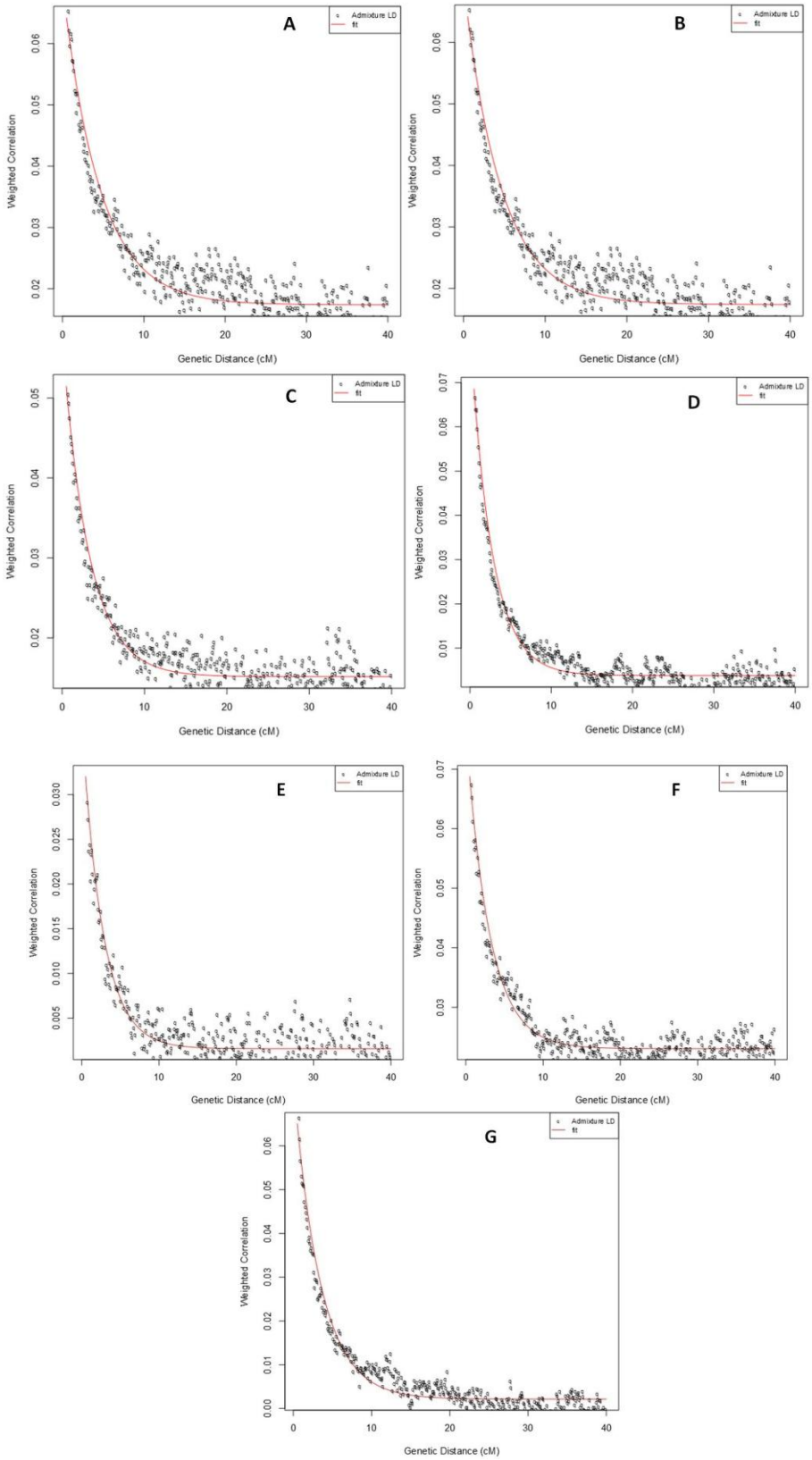

Figure 103 Admixture date estimation using ROLLOFF.
Weighted covariance as a function of genetic distance (cM) between the ancestral populations (Yoruba and Italians + Spanish) and the descendant population: A - Yemen, B - Yemen* (without the individual displaying a higher African admixture than average), C - Saudi Arabia, D - Bedouin, E - Syria, F - Jordan and G - Palestinian. Ages and confidence intervals of the admixture event are estimated by fitting an exponential distribution.

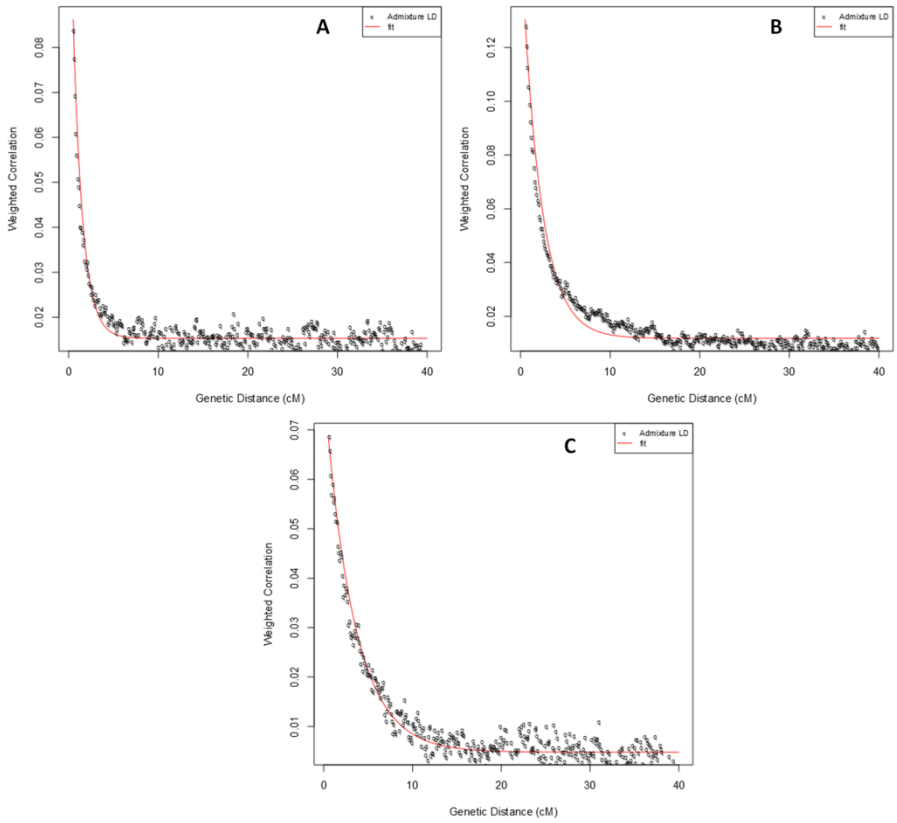

Figure 104 Admixture date estimation using ROLLOFF.
Weighted covariance as a function of genetic distance (cM) between the ancestral populations (Yoruba and Italians + Spanish) and the descendant population: A - Ethiopia, B - Maasai, and C - Egypt. Ages and confidence intervals of the admixture event are estimated by fitting an exponential distribution.

Table 8 Estimates of admixture proportions (%) and date of admixture (in generations) calculated in ROLLOFF when using Yoruba and Italians + Spanish as ancestral populations.

Population	Sample Size	West African ancestry proportion (%) ± standard error	East African ancestry proportion (%) ± standard error	Near East/Arabian/Caucasus/Iran ancestry proportion (%) ± standard error	European ancestry proportion (%) ± standard error	Asian ancestry proportion (%) ± standard error	Estimated date of admixture using ROLLOFF (generations)
Yemen	10	16.733 ± 4.977	6.715 ± 1.341	60.994 ± 3.961	5.350 ± 1.523	10.208 ± 5.253	18.377 ± 6.541
	9*	13.060 ± 3.755*	6.716 ± 1.500*	64.317 ± 2.409*	5.759 ± 1.6340*	10.147 ± 1.856*	21.019*± 7.450
Saudi Arabia	20	2.124 ± 1.231	5.304 ± 0.784	81.392 ± 3.658	5.481 ± 1.286	5.699 ± 5.504	30.762± 4.907
Yemen Jews	15	0.001 ± 0.000	5.559 ± 0.252	86.373 ± 0.581	6.362 ± 0.479	1.704 ± 0.353	n/a
Bedouin	45	2.674 ± 0.379	5.326 ± 0.439	80.592 ± 2.139	8.411 ± 1.142	2.997 ± 0.517	37.546 ± 3.104
Lebanon	7	1.662 ± 0.612	3.522 ± 1.088	61.626 ± 1.759	22.214 ± 0.942	10.977 ± 2.425	n/a
Syria	16	1.665 ± 0.414	2.928 ± 0.585	62.497 ± 1.011	19.345 ± 1.520	13.564 ± 1.070	37.334 ± 4.365
Jordan	20	4.151 ± 1.387	5.672 ± 1.408	63.786 ± 1.884	18.325 ± 0.697	8.067 ± 0.789	32.871± 4.106
Samaritan	3	0.001 ± 0.000	0.178 ± 0.177	76.278 ± 2.392	19.315 ± 1.545	4.228 ± 0.923	n/a
Druze	42	0.172 ± 0.064	0.326 ± 0.097	70.688 ± 0.441	21.669 ± 0.349	7.145 ± 0.229	n/a
Palestinian	46	3.431 ± 0.260	4.050 ± 0.358	69.778 ± 0.486	17.047 ± 0.259	5.693 ± 0.265	29.008 ± 2.194
Iran	20	1.605 ± 0.738	0.682 ± 0.306	49.970 ± 1.081	15.494 ± 0.681	32.248 ± 1.263	n/a

Turkey	19	0.001 ± 0.000	0.002 ± 0.001	45.442 ± 1.068	29.708 ± 0.759	24.847 ± 0.743	n/a
Ethiopia	19	10.113 ± 0.776	52.243 ± 1.846	37.218 ± 2.486	0.246 ± 0.136	0.179 ± 0.120	93.223± 9.678
Maasai	19	21.746 ± 1.464	74.060 ± 1.745	3.762 ± 0.342	0.118 ± 0.0627	0.314 ± 0.158	47.007± 2.933
Egypt	12	7.409 ± 0.382	13.111 ± 1.484	63.662 ± 1.728	13.995 ± 0.824	1.824 ± 0.356	30.034± 3.233

n/a – not assigned due to small sample size or low African input

* - data when removing an outlier from the analysis.

VI. DISCUSSION

The present work is, by far, the most complete characterisation of the maternal pool of the Arabian Peninsula, a place central to archaeological and population genetics research since the advancement of the southern-route hypothesis.[191; 192; 291] The mtDNA remains the most informative and secure genetic material to infer past migrations and estimate their time frame, and most of my work is focused in the analyses of this genome. I was interested in characterising important periods of the prehistory/history of this region, and for that I needed to be able to disentangle signs and influences of successive migrations, for which a good chronological dating is critical. The good overall knowledge of lineage distribution and frequencies based on the HVS-I diversity and the robust phylogeographic reconstruction through the complete sequencing of the mtDNA molecule allows us to focus in key periods of the population demography by detailed inspection of informative lineages. Some authors have been very critical of this approach, affirming that maternal lineages do not necessarily represent the entire population, being highly sensitive to drift (as discussed in Scally and Durbin),[292] but mtDNA-based conclusions for many migrations across various regions of the globe have been supported by genome-wide results (summed up in Mellars et al.).[293] In fact, the genealogical approach taken for mtDNA may overcome the effects of drift more easily than the use of genome-wide SNPs, as we recently demonstrated in the highly-drifted Ashkenazi group: the fine characterisation of full mtDNA sequences provided a detailed reconstruction of the maternal Ashkenazi pool, ascertaining that at least 80% of the lineages have a European ancestry,[288] an ancestry influence not easily identified in worldwide PCAs based on genome-wide data.[136]

To illustrate the capacity of mtDNA dating to disentangle successive migrations, a parallel can be established with archaeological dating. In archaeology, a relative chronology is ascertained through stratigraphy that consists of the careful characterisation of overlapping strata. The dating of materials in each stratum by radiocarbon methods allows us to establish a time frame as each layer can be placed in sequence and the dates interpolated. Usually, the radiocarbon estimates have short confidence intervals, but any methodology has potential artefacts such as, in this case, the inclusion of diagnostic materials in an incorrect soil layer due to the action of animals or geological events as erosion or infiltration. In genetics, the dating methodologies based on molecular clocks are more reliable for uniparental markers, as the absence of recombination does not mix portions of homologous biparental regions (although the Y-chromosome has been prone to much uncertainty due to the difficulty in estimating a reliable mutation rate).[129] Recent

technological developments have led to a considerable dataset of complete mtDNA sequences[269], enabling to attain two fundamental requirements for molecular dating: well-estimated phylogenies[266]; and reliable molecular clocks to convert genetic diversities into time (corrected for the effects of purifying selection).[116] However, the age of a haplogroup cannot be directly associated with a migration event, as the diversity arisen in the founder population and predating the migration event would be included in the measurement. This limitation is overcome in the founder analysis,[259] which picks out founder sequence types in potential source populations and dates lineage clusters deriving from them in the settlement zone of interest. In a way, the founder analysis (mostly applied only to HVS-I diversity;[99; 259] but already used with complete sequences for the typical Southeast Asian haplogroup E[294]) allows us to reconstruct the stratigraphy of the migration events responsible for making up a population genetic pool; and the complete mtDNA characterisation of the informative haplogroups allows us to estimate a more reliable time frame for the events.

The availability of genome-wide information for Arabia and neighbouring regions, so far not explored in the context of inferring local population structure and admixture,[83; 134; 136] led me to also obtain preliminary results at a genome-wide scale. These results complemented my mtDNA-based observations and will guide our group in future screenings in the region.

The first descendents of the out-of-Africa migration

My group published (and I had the opportunity to participate in) a careful dating of the out-of-Africa migration by performing the full characterisation of the mtDNA haplogroup L3, which is the MRCA of L3 lineages observed nowadays in East Africa and encompasses haplogroups N and M which include all ancient diversity observed outside Africa.[99] The results obtained from the analysis of 369 L3 sequences allowed us to estimate an age for this haplogroup, and hence the upper-bound of the out-of-Africa migration, at around 60-70 ka (with overlapping 95% confidence intervals for the various estimates, ρ total, ρ synonymous, ML and Bayesian, from 59 to 79 ka). This dating enabled us to refute some theories and to support other hypotheses, we can conclude: (1) The earlier migrations out-of-Africa during MIS 5 or in the transition to it both through the Levant[295] or via the southern route,[12; 206] as testified by archaeological remains, did not leave maternal descendents of these possible earlier migrants in the extant population; (2) The successful exit out-of-Africa took place after the Toba supereruption which occurred

at 73.5 ka; (3) The L3 expansions in eastern Africa and the exit of modern humans from Africa ~60 ka seem to have been part of a single demographic process most probably driven by the moister climate after 70 ka in eastern Africa,[19] which could have been associated with improved food resources;[182] (4) The L3 expansion was also responsible for migrations within Africa, especially into Central Africa in the time frame between 60-35 ka, but no L3 lineage reached South Africa until the Bantu introduction at ~1.8 ka, implying that the important move toward behavioural modernity, which had arisen in southern Africa by 70-80 ka,[293; 296; 297] cannot be directly linked with the L3 lineage.

A natural progressing for this line of research was to apply the same detailed mtDNA characterisation to the first descendants observed in Arabia, and contribute to address the question of where L3 evolved into the two non-African haplogroups N and M. This was the first issue investigated in my thesis. I decided to avoid the widespread and frequent haplogroup R, the phylogeny of which was highly influenced by posterior migrations and expansions (confirmed in this work, as detailed in the next section) and focus on three basal and deep N clades, haplogroups N1 (including I), N2 (including W) and X, observed at low frequencies in western Eurasians. Other N(xR) haplogroups, such as N5, N9, A, N21 and N22, seem to have dispersed towards the east of Eurasia, and were thus not addressed in the screening of Arabia. Nonetheless, this west *versus* east Eurasian dispersals of basal and rare N haplogroups is itself of interest, seeming to point to a common origin somewhere between the Indus Valley and Southwest Asia, as advanced by Metspalu et al.,[298] a good preliminary indication supporting my focus in Arabia. Also, as L3 undoubtedly originated in eastern Africa and archaeological evidence shows that modern humans arrived in the Near East only ~45–50 ka ago,[299] the most parsimonious location for the emergence of N is again in the vicinity of the Arabian Peninsula.

My phylogeographic analyses of 385 complete mtDNA sequences (85 characterised by me and 300 already published) for N1, N2 and X provided valuable insights into the origin and location of the first out-of-Africa migration descendants. I was able to better date the root of haplogroup N, at 57-65 ka (overlapping 95% confidence interval between 52.8-72.0 ka), very close to the upper bound for the out-of-Africa migration defined by the origin of L3 in eastern Africa 60–70 ka ago.[99] This suggests that haplogroup N had an origin immediately outside Africa, most likely in the Arabian Peninsula. Although I did not ascertain haplogroup R, previous evidence[88] showed that Arabia was probably also its cradle, and the time frame (~59 ka) is contemporaneous with my estimate. As N5 has been

recently described in Iran,[300] and is otherwise rare in Southern Asia, it is also possible that it arose in Southwest Asia. Further screening of Arabia and the surrounding areas may still increase the number of basal N branches present there, including for the other N(xR) branches typical nowadays of Eastern Asia, Southeast Asia and Australasia, especially the rare ones, but so far the most probable scenario is that some individuals belonging to the N(xR) root crossed into East Asia, along with lineages within R, and there gave rise to the new N clades.

If climate seems to have played such a crucial role in the L3 expansion within and out of Africa,[99] I also found evidence for its impact in the Arabian maternal pool, with arid conditions and the distribution of refugia inside the Peninsula being crucial to the migrations towards South Asia and Europe. Researchers believe that the arid conditions of MIS 4 (75–60 ka), along with the Rub' al Khali desert and Asir Mountains as major barriers, closed off the corridors from the Arabian Peninsula to the Levant until ~50 ka ago.[301; 302] This would imply a faster dispersal toward South Asia than toward Europe, as observed in the mtDNA studies as older ages for N branches are typical of South Asia rather than Europe[88] and testified by archaeological remains.[287; 303]

Archaeological and climatic research[302] is showing that even during the most hyperarid phases there were several refugia that provided food and fresh water in Arabia: the Red Sea basin and Asir-Yemeni highlands; the southwest Arabian littoral zone (including the Hadhramaut in Yemen and southern Oman); and the Gulf Oasis. The Red Sea basin and Asir-Yemeni highlands puts some difficulties into migrations between the Near East and Arabia as terrain is difficult to cross (mountains reaching 2,100 m in altitude). The Hadhramaut region/southwest Oman is not directly in contact with the Near East. But the Gulf Oasis, which lasted from 74 to 8 ka ago and included the shallow exposed basin of the Arabo-Persian Gulf, was an accessible corridor which extended considerably the southeastern tip of the Fertile Crescent. The Oasis experienced periodic contractions into coastal margins during periods of low sea level and aridity and expansions from the exposed shelf into the hinterland during pluvial episodes (from 55–24 ka ago during MIS 3 and again after the LGM, from 12–6 ka ago).[302] Further support comes from the presence of archaeological sites in the Gulf basin suggesting a long tradition of human occupation, although it was not possible to ascertain any direct cultural influence from the Levant or from Africa in the Upper Palaeolithic lithics observed in eastern Arabia, pointing to a local development of cultural techniques.[304]

The fact that time estimates, frequencies and genetic diversities observed here for N1, N2 and X are similar between the Levant and Arabia, challenges the

hypothesis of long-term isolation between these two regions. The Gulf Oasis, continually in contact with the Fertile Crescent, appears to be the most likely locus of the earliest branching of haplogroup N, including the three relict basal N(xR) haplogroups studied here, as well as the major Eurasian haplogroup R.

We can hypothesise that the Gulf Oasis was the preferential corridor to and from the Near East, especially in the most favourable climatic periods, and that the other two refugia identified in the south and southwest of the Peninsula acted as a corridor for migrations west, including back towards eastern Africa. The deep lineages in eastern Africa found here (for haplogroups I, N1a, and N1f) support the migration back to Africa, indicating also that most probably it took place a number of times beginning as early as 40 ka ago (I will come back later to this issue). This is independent evidence for back-to-Africa migrations of N lineages, which complements the migration of U6 and M1 lineages argued to have take place around 45 ka through the Sinai Peninsula.[188]

Supporting evidence for the genetic proximity from eastern Arabia to the Near East comes from Y-chromosome microsatellite diversity in the Arabian Peninsula,[253] which suggests that Dubai and Oman share genetic affinities with other Near Eastern populations, whereas Saudi Arabia and Yemen show signs of greater isolation (although for fast-evolving microsatellites, these differences might reflect more recent events).

Unfortunately, there is no genome-wide information available for Dubai and Oman, which would allow us to compare the genetic affinity between west and east Arabia and the Near East. Nonetheless, the available genome-wide data for Saudi Arabia, Yemen and Bedouin show that they form a close cluster with Near East populations (both PCA and ADMIXTURE analyses). The few differences between Near East and Arabia seem to have been introduced by geographically neighbouring populations, such as a higher European and Pakistan proportion in the Near East (between 20%-30% and 10%-30%, respectively) than in Arabia (less than 10% for both), and the opposite for the African ancestry (especially eastern African, which decreases from around 5% in Arabia to nil towards the north). Very interestingly, both Bedouin and Saudi Arabian populations include some individuals with less admixed ancestry (just the bulk Near Eastern/Arabian component and around 5% of eastern African genes), who may represent descendants of less admixed sub-populations, and could provide further insights into what characterises the respective Near Eastern/Arabian genetic pools. The back-migration to eastern Africa is testified by Ethiopia displaying a high proportion of the Near East/Arabian Peninsula component; but we will come back to this later on.

Although the main branches of the haplogroup N tree appeared between 32 and 55 ka ago during the MIS 3 pluvial stage, these ancient episodes have been, to a considerable extent, overwritten by stronger signals of expansion at the end of the last Ice Age. I addressed this issue in detail within the Arabian Peninsula and as a motor to migrations towards Africa; but Southwest Asia also acted as a reservoir for dispersals towards Europe (and indeed America) in distinct episodes, which although not being the aim of my thesis are still worth describing here.

Most of the N(xR) migrations from Southwest Asia to Europe seem to have taken place after the LGM, as shown by founder analyses, BSP and other phylogeographic evidence for N1a1, I4 and I2'3, N1b and X2+225. A possible exception is haplogroup W, within N2, which might conceivably have arrived in Europe prior to the LGM or already originated there at around 20 ka. However, many of these lineages residing in Europe since the end of the Ice Age show evidence of Neolithic expansions. For instance, two branches within N1a (N1a1a1 and N1a1a2), detected at high frequencies in LBK (Linearbandkeramik) burials, also appear to have participated in local Neolithic expansions into central Europe but seem to have subsequently dwindled dramatically in frequency.[107; 305] A similar careful phylogeographic characterisation of the JT clade[101; 289] clarified that the main period for the entrance of this clade into Europe was the Late Glacial period, ~19–12 thousand years (ka) ago, with just a few lineages coming in the Neolithic Very recently, further study of haplogroups I and W corroborated my results.[306] Thus, the fraction of lineages which arrived in Europe from the Near East during the Neolithic is quite restricted within the haplogroups N(xR), such as I1a and X2e2.

One of the most intriguing aspects of N(xR) phylogeny is the arrival of the major haplogroup X2 subclade to Central Asia and North America. It remains unclear which founder(s) arrived in America and gave rise to the characteristic subclades X2a and X2g, but so far the most probable is the root of X2+225, dating to ~21 ka. So, the same post glacial migration from the Near East into Europe also seems to have been responsible for the migration into Central Asia and North America, reaching a remarkable geographical extension. X2 had a very distinctive history relative to haplogroups A, B, C and D, its partners in the settlement of America, which took place in one or several migrations depending on accepted models (discussed in Achilli et al.[307]).

My founder analysis for the haplogroup N1b2 indicated also its most probable entrance in Europe in the postglacial period. This result is curious, as this haplogroup is found mainly in Ashkenazi Jews, and I dated a founder effect among Ashkenazi ancestors as very recent, at ~2 ka. This observation led our group to

explore this issue in further detail,[288] and we found two deep N1b2 branches of Italian origin (new coalescent age of 5.0 ka), rendering more probable that the N1b2 lineages were introduced into the Ashkenazi pool from lineages within Europe. This is a clear example of how a careful phylogeographic characterisation of mtDNA can provide powerful insights into past demographic events and help to clarify the reasons for shared ancestral components: a considerable proportion of the common European/Near East component in the Ashkenazi community is due to introduction of European females in the Diaspora at 2-3.0 ka, and not a shared Near Eastern ancestry.

Major population expansions in the Arabian Peninsula

Arabian archaeologists are still engaged in the discussion about population continuity versus replacement.[242] This can be placed at several time frames, and indeed the first one is at the level of the first successful settlement of Arabia. The absence of caves in most of Arabia has rendered secure stratigraphic reconstructions very difficult to establish, so that this issue remains largely unclear when attending only to archaeological evidence.

It is clear from the analysis of the N1, N2 and X lineages performed here that a number of ancient lineages within N(xR) have survived in Arabia, albeit at low frequencies. These lineages, which include N1a2, N1f, and possibly also N1c, N1d, and N1e, date to 15–55 ka and coalesce to the most ancient non-African mtDNA lineage ~60 ka. Thus, they are most likely relicts from the first modern-human settlement in Arabia during the earliest stage of the southern coastal dispersal from the Horn of Africa to the rest of the world.

Most of the archaeologists put an emphasis on the Neolithic period for the question of population continuity/discontinuity. Evidence that the Neolithic came to Arabia from the Fertile Crescent suggests the “Levantine hypothesis,” which states that there was a major population discontinuity between the Late Pleistocene and Early Holocene in Arabia and that this region was recolonised by Pre-Pottery Neolithic B Levantine people ~9–8 ka ago.[241] Archaeological evidence testifies the expansion of sedentary Natufian hamlets in the Levant during the wet phase between 15 and 13 ka ago, and such expansion lead to the techniques of agriculture and domestication under the harsher conditions of the Younger Dryas and ultimately to the Neolithic cultures of the early Holocene.[308]

The investigation performed here for the rare N1, N2 and X haplogroups showed that demographic expansion in Southwest Asia appears as a continuous

phenomenon from the Late Glacial period to the Neolithic period. These N(xR) expansions match signals from several haplogroup R lineages within Arabia, in which my group has been previously involved (including myself, in lab work conducted prior to my thesis).[258; 260] The sister haplogroups R0a and HV1 together account for ~16% of the Arabian pool.[102; 252; 253; 257] These haplogroups are believed to have emerged in the Near East, as the oldest lineages for the R clade are observed there,[259] in this specific case dating to around the LGM: 22.6 (20.9–24.3) ka for R0a;[258] and 22.4 (14.7–30.2) ka for HV1.[260] But the fact that several founders of R0a are present in southern Arabia, renders it possible that Arabia was an additional centre for R0 emergence.[258] These Arabian R0a founders suggest population continuity between the terminal Pleistocene and Holocene, and some of them also spread to North and eastern Africa, as observed in this work for the N(xR) lineages. Two periods of demographic expansion closely match two wet climatic periods, the end of the 35–20 ka wet period and the brief wet phase between 15-13 ka, and the latter seems to have been especially important.[258; 260] Y-chromosome diversity also suggests that the Arabian Peninsula mainly received its lineages from the Near East from the LGM onwards.[261]

So, all these data seem to suggest that important dispersals from the Near East into Arabia predated the Neolithic period, and expansions occurring during the Neolithic could have involved local lineages. My phylogeographic analyses of the mtDNA sister haplogroups J and T in Arabia, previously associated with the Neolithic in many parts of the Old World (as newcomers,[259] or as locally expanded[101]), was thus fundamental to add information to this issue. Again the frequency and diversity maps displaying similarity across the Near East and Arabian Peninsula, as well as the many basal Arabian lineages, favour at least that both regions were in close contact throughout late Pleistocene and Holocene. J assumed a much important role in Arabia overall than T, as testified by frequencies (respectively, 20.6% and 6.2% in Saudi Arabia; 14.7% and 5.8% in Yemen; 11.7% and 3.2% in Dubai; although in Southwest Oman, Dhofar, the frequencies are slighted inversed, 7.7% and 10.2%, probably reflecting local fluctuations and the many star-like J sub-branches observed mainly in Arabia and dating to around 6-7 ka). These expansions in J are also clear in the BSP analysis, for which the main increase in effective size was between 8-12 ka in Arabia, after the expansion observed in the Near East around 11-15 ka. J also shows signs of having crossed into eastern Africa, particularly the haplogroup J1d1a1, necessarily after its emergence in Arabia at ~7.1 ka.

The founder analysis performed here, by assuming migration from the Near East and Pakistan into the Arabian Peninsula, is the first attempt to quantify the genetic impact across diverse periods. The broad results seem to favour the periods 1.0, 11.0 ka and 16.0 ka, which led me to attempt imposing four migration periods at: 1.0 ka (representing recent events), at 8.0 ka for the Neolithic introduction, 11.0 ka for the Younger Dryas, and 16.0 for the Late Glacial period. The Late Glacial period was the most important, responsible for the introduction of 31%-46% of the lineages mainly belonging to the haplogroups K, U2, U3, U4, N1a1a, N1a1b, H5 and HV1. At the Younger Dryas, 24-28% of lineages, mainly unclassified HV and R0a affiliated, migrated to Arabia. The Neolithic seems to have introduced 20%-25%, mainly J1b, T1a and M1. The remaining 10-15% moved very recently, ~1.0 ka, and consist of derived lineages, including J1d1a, K1, HV8 and N1a3. It is not possible to discriminate clearly between the western Near Eastern and Iranian influences, as they share a similar mtDNA pool, so my attempt to investigate the Persian influence in Arabia was quite limited. It nevertheless suggested a higher impact in east (41%) than west (25%) Arabia for private founders, which seems to be low frequency lineages, so that values are quite similar in the overall pool (14% and 11%, respectively).

So, there seems to have been a limited entrance of lineages from the Near East into Arabia during the Neolithic, mainly for J and T. But the Neolithic in Arabia was mostly characterised by the expansion of the effective size of local N lineages, mostly R0a. These R lineages may have reached Arabia in the Late Glacial coming from the Near East, although it is still possible that Arabia (especially the Gulf Oasis region) made part of an extended centre of origin together with the Near East.

Genetic exchanges across the Red Sea

The Bab al Mandab strait and the Red Sea are important routes linking eastern Africa with the Arabian Peninsula, having highly dethroned the Sinai as the only link between Africa and Southwest Asia. This replacement is total for the successful out-of-Africa migration of individuals assigned to L3 haplogroup at ~60 ka. So far, no direct descendants of the L3 migrants have been found in the Arabian Peninsula, besides the ones affiliated in the N derived group. Nevertheless, the Arabian Peninsula carries African L(xMN) lineages in its mtDNA pool, introduced since its settlement 60 ka (making it a very interesting case for studying admixture with the sub-Saharan African gene pool). Richards et al.,[108] based on mtDNA control region information, indicated that the African maternal input in Arabia and

the Near East took place at ~2.5 ka, concordant with the Arab slave trade, and that there was an asymmetrical gender African contribution, predominantly female.

Here I investigated two relatively rare African haplogroups, phylogenetically close to L3, which are designated L4 and L6. As these lineages are very rare, they have escaped surveys of complete mtDNA sequencing to date, so their phylogeny has been poorly known. L6 is particularly interesting, as Kivisild et al.[207] detected a high proportion (12%) of these sub-Saharan lineages in their Yemeni sample, suggesting that their presence could be explained by gene flow across the Red Sea earlier than the recent movements at 2.5 ka. This frequency was not confirmed, however, in other Yemeni sample sets.[252; 254] I found that L4 is more frequent nowadays in the Near East, but the diversity indices point to a most likely eastern African origin at ~87 ka, predating the out-of-Africa dispersal (as well as its sub-branch L4b dating to ~85.8 ka). So, in theory, this sister haplogroup of L3 could have crossed into Arabia along with L3 during the initial out-of-Africa movement. Phylogenetically, the few Arabian L4 lineages are very derived, indicating recent exchange networks between eastern Africa and Arabia for its entrance. L6, similarly rare in Yemen and East Africa, dates to 23.1 [15.8-30.5] ka, and is likely to have migrated from East Africa into Arabia after that period, most probably very recently as testified by a very derived L6a sub-branch observed in three Yemenis (sharing the same lineage).

The founder analysis of the migration of sub-Saharan lineages from Africa into Arabia plus Near East (both regions have to be considered together due to the relatively low number of L sequences in Southwest Asia) showed a single peak of migration at 1.0-2.0 ka, although examination of particular frequent founders suggests that this could have been introduced earlier. This led me to impose four periods for migration, and their impact was: 27%-38% for 0.5 ka (slave trade initiated by Arabs in the 6th century AD till the 19th century; I selected the 0.5 ka to be at the middle of this time frame); 35%-38% for 2.5 ka (Arabian dominance in the Red Sea trade routes); 34%-41% for 8.0 ka (beginning of maritime routes in the Red Sea and Neolithic); and a residual 0.6%-1.8% for 60.0 ka (out-of-Africa migration). No clear pattern of association between haplogroup and event is observable, probably reflecting the high heterogeneity in the source (L6, L0a1, L3b and L2a for 0.5 ka; L3b, L3d1a, L3e3, L2a1a2, L0a and L5 for 2.5 ka; L0a, L3e1, L3f1b and L2a for 8.0 ka; and none with significant probability for 60.0 ka). Thus, the Arab maritime dominance and slave trade (0.5-2.5 ka) were the main contributors (~60-70%) to the African input into Arabia and Near East, but the entrance seems to have been initiated as early as the beginning of the maritime

networking in the Red Sea. A larger dataset of complete L(xMN) sequences in Arabia, allowing the application of founder analysis, would help to clarify this issue.

As already mentioned, the sensitivity of mtDNA to genetic drift has been used as a strong argument for caution when making inferences on migrations and admixture based solely on this marker. This issue is motivating developments of statistical methods for genome-wide analysis, which are allowing the dating of events of admixture, that according to the authors already address the biasing influence of uncertain local ancestry inference and inaccurate ancestral populations.[286] An admixture event long investigated through mtDNA information and recently addressed also in genome-wide studies is the African input into southwest European populations. First, mtDNA data indicated that the Atlantic slave trade, which began in the 15th century, was the most probable migration for this introduction into Europe.[309; 310] But the complete sequencing of African mtDNA genomes observed nowadays in Europe led to the identification of some African lineages (a proportion of 35%), which could have been introduced in Europe already at 11 ka, while the bulk of 65% could have been introduced there in the Roman and Arabic periods or during the Atlantic slave trade (the 95% confidence intervals are too large to obtain precise proportion estimates in this time frame).[311] The genome-wide analyses of 40 West Eurasian populations showed a 1-3% African inheritance in Southern Europeans and around 55 generations for this admixture event, which the authors[142] interpret as consistent with North African gene flow at the end of the Roman Empire and subsequent Arab conquest. In theory, the ROLLOFF method could detect admixture events as old as the 11 ka admixture event described by Cerezo et al.[311] in the mtDNA, but the tendency is for averaging the time of admixture from different periods. Admixture dating based on global genome analysis does not have the resolving power to distinguish between several admixture events as a careful mtDNA phylogeographic analysis has.

Although no Arabian populations have been studied for admixture, Moorjani et al.[142] applied the ROLLOFF estimates to Levantine groups, showing a 4%–15% African ancestry, and obtained a value of about 32 generations, interpreted as consistent with close political, economic, and cultural links with Egypt in the late Middle Ages. They also estimated 72 generations for the event leading to 3%–5% sub-Saharan African ancestry in diverse Jewish populations, argued as reflecting descent of these groups from a common ancestral population that already had some African ancestry prior to the Jewish Diaspora.

The results obtained in this work for the genome-wide analyses in the data available for Arabian populations provide the first estimates of African admixture in

this region, with disentanglement between western and eastern African pools. The eastern African input is around 5.3% in Saudi and Bedouin, and slightly higher (~6.7%) in Yemen (although Yemen Jews have a lower admixture of 5.6%); this input decreases beyond Jordan, being negligible in Samaritans, Druze, Turks and Iranians. The eastern African component is on average double that of the western one (from 0 to 4%), except for Yemen (13-18%) where most probably recent migration inflated the western African component, leading also to a reduction in the ROLLOFF estimate for the event of admixture to 20 generations, while in most other populations it is between 30-40 generations ago (values concordant with estimates from Moorjani et al.[142]). These date estimates are compatible with the Arab slave trade, which operated between the 7th and 19th centuries, mainly from eastern Africa (from Nubia to Zanzibar). The fact that the Druze do not have any African input also favours a recent introduction of this component into the region, as it has been suggested that the Druze have had a low migration rate with nearby populations, providing the closest snapshot of the genetic landscape of the Near East from the beginning of settlement by modern humans.[113]

It has been suggested that the African input was mainly female-mediated,[108] and my frequency of 7-20% African input in the genome-wide data agrees with the 35% female input and 0% male one. I also averaged the generations taking into account the proportion of maternal L lineages introduced at each of the three migration events considered in the mtDNA founder analysis from Africa into Arabia and Near East. Remarkably the result is 51.6 generations (20 generations for introduction of 29.5% African lineages; 100 for 34.8%; and 320 for 35.7%), very similar to the estimate of 30-40 generations from the genome-wide data. This is a nice proof of the averaging effect on the ROLLOFF and other global estimate of date of admixture based on genome-wide data. This result calls for a cautious application of this kind of estimates in complex scenarios of several migrations occurring over a long span of time, like in Arabia, and most probably in North Africa[312] and Iberia.[311] Of course, it may be quite reliable in cases of a major single migratory event like the presence of African lineages in America due to the Transatlantic slave trade.[313]

The Bab al Mandab strait and the Red Sea were also important for the opposite direction, the back-to-Africa migration. Especially, as early as the 7th millennium BC, the Red Sea was the stage for obsidian exchange networks, characterised by an extensive maritime traffic,[214] which also led to the settlement of the island Socotra (archaeologically unknown, but founder effects on R0a lineages place this between 11-3 ka).[254] Obsidian first appears on Tihama and indicates direct or

indirect contact between the central or southern highlands of Yemen, and/or the Horn of Africa. The obsidian densities are highest at sites right on the coastline and decrease at sites along the rivers, reinforcing the maritime trade assumption. These results are compatible with the emergence, in the 4 millennium BC, of major state-level civilizations of the Old World in a number of regions bordering the Arabian Peninsula, and responsible for the development of maritime activity in the Red Sea region. This intensive commerce culminated in the emergence of prosperous trading kingdoms by 2.0 ka, as the 'incense kingdoms' of Sabaea, Qataban, Hadhramaut and Ma'in in Arabia. These kingdoms ruled the Red Sea markets trading between Africa and southern Arabia mainly through the domestication and spread of the dromedary camel. Data on R0a[258] and HV1[260] demography showed that some lineages potentially passed from South Arabia to eastern Africa during the periods of intense commercial network activity began at 8.0 ka.

My phylogenetic analyses for N(xR) lineages provide insights into the back-to-Africa movements, seemingly at various periods of time. Some lineages showed some deep branches in eastern Africa: I, N1a and N1f. This is a sign of introduction in Africa beginning as early as 40 ka and extending till 15 ka. Curiously, at the level of the complete sequences available, I did not find any lineage belonging to these haplogroups that could have been introduced into eastern Africa after that period, also suggesting that by this time other haplogroups had become more frequent in Arabia, and those were the ones now taking a major role in the back-to-Africa migrations. An exception is I5a, which displays a strong founder effect in Socotra, dated at 2.0 [0.4-3.6] ka (one sequence is from Dubai), placing its entrance a little later than the time frame for the more frequent R0a, as mentioned above.[254]

I surveyed in detail the back-to-Africa migration of Eurasian lineages through founder analysis for the introduction from Arabia and the Near East into eastern Africa. The peak of migration identified dates 10-15 ka, and when imposing four migration events the proportions were: 13-29% at 2.0 ka (Arabian dominance in the Red Sea routes) for N1, R0a, T, J, K and X; 24-34% at 8.0 ka (for Neolithic and beginning of maritime trade) for U6a1a, J1d1a, M1 and R0a; 31%-48% at 16.0 (Late Glacial period) for M1 and HV1; and 6-14% at 45.0 ka (using the date suggested for the back-to-Africa migration of U6 and M1 through the Sinai)[188] for N2 and U5. Although N1 is associated here with the younger migration, when attending more carefully to the lineages involved, most probably this results from a lack of resolution when analysing only HVS-1. Two main founders are in the root of N1 sub-haplogroups, one of N1a and the other of I. Another founder in N1a could be placed in the sub-branch identified in the complete sequencing (Figure 31 in

Results) in Somalia, bearing the substitution at position 16213; but the HVS-1 data show that this is more frequent in Africa (7 individuals) than in Arabia (1 individual) so probably this Arabian individual is a recent introduction in Arabia of a N1a sub-haplogroup developed already within Africa. The migration of J1d1a into eastern Africa in the Neolithic period is confirmed in the complete sequencing complemented by frequency interpolation and founder analysis.

In the genome-wide results, it was possible to infer that this back-to-Africa migration was considerable, leading to a proportion of 37.2% of Near East/Arabian Peninsula ancestry in Ethiopia, and the ROLLOFF estimate for the date of admixture was 93 generations ago (two times higher than the time for the African admixture in Arabia and Near East). For comparison, in the Maasai from Kenya and Tanzania, the Eurasian component is an order of magnitude lower (3.8%), and the time of admixture of 47 generations, reflecting most probably later admixture events already within eastern Africa.

In parallel, I investigated the potential introduction of Eurasian lineages from Near East/Arabia into North Africa through the Sinai route. Musilova et al.[260] suggested that Egypt received some N lineages directly from the Near East after the LGM, supported in their observation of high match of HV1 lineages between Near East/Arabia and Egypt. In my founder analysis, I detected two well defined peaks at ~2.4 ka and 6.8 ka with the *f1* criterion, and two peaks at ~9.0 ka and ~12.4 ka when using the *f2* criterion. This seems to explain the role of the Neolithic period, which would support results obtained for the North African Y chromosome pool consistent with a large Neolithic origin, suggested to have been due to the demic diffusion of Afro-Asiatic-speaking pastoralists from the Near East.[314] This is confirmed by imposing migratory periods to the founders, leading to: 6-15% at 2.0 ka, mainly HV1 and other undefined HV lineages, M1 and U (U6a1, K1a1); 47-55% at 8.0 ka for most HV, U (U5b, U5 and K), T (some T2c1 and T2b) and J (J1d1a, J2a2b and other undefined J), and X; 28%-42% at 16.0 ka for some HV, T (T1a, T2) and U (U3, U3a, U5b1b, U5a, U6a) lineages; and finally 3-5% at 45.0 ka for few and rare U (U3, U5, U5a1 and U6). It seems probable that some JT lineages, especially T ones, were introduced into Northeast Africa before the Neolithic, following Late Glacial population expansions in the Near East/Arabia. Then, locally they could have been involved in population expansions in the Neolithic period, leading to signs of autochthonous founder effects, such as the one detected in the El-Hayez oasis (400 km southwest of Cairo) for sub-haplogroup T1a2a.[315] The link between U6 and M1 and the settlement of North Africa from the Near East at 45.0 ka advanced by some authors[188] has been recently put into question[190] because

their sub-haplogroups do not all display the same history, U6a is ~10.0 ka older than M1a and M1b, and sub clades of the former coalesce before or around the LGM while sub clades of the latter date to the post-LGM. In the founder analysis for North Africa, I obtained a stronger Late Glacial and a residual Upper Palaeolithic signal for U6. The fact that I did not detect a clear sign of U6 and M1 introduction at 45.0 ka was to be expected as currently no ancestors or roots of these haplogroups are detected in the Near East, although this region remains the most probable centre of origin for proto-U6 and M1.[190; 265] My founder analysis results are probably affected by local expansions of U6 in North Africa in the postglacial – the analysis is picking up U6 back-migrants into Near East at this time as ancestors (our imposition in the analysis is that Near East is the ancestor population).

At the genome-wide level, Egypt is quite similar to its Levant neighbours, displaying slightly higher West (7.4%) and East (13.1%) African proportions, and lower European (14.0%) and Asian (1.8%) proportions. The ROLLOFF estimates for the admixture events in Egypt (using Africans and Europeans as ancestral populations) was 30 generations, expectably young due to the continuous gene flow between the two regions. Here I did not include the west of North Africa in the analysis, as Behar et al.,[136] reinforced in Pennarun et al.,[190] already showed that at high *K* (*K*=10) a component very frequent in Mozabite Berbers and Moroccans and less frequent in Ethiopians, Egyptians and Near Eastern populations, could be paralleled to the U6 and M1 fraction of the mtDNA pool. Not only this frequency difference for U6 but also the major southwest European influence westwards of Cyrenaica,[264; 316; 317] renders the east of North Africa (mainly Egypt) genetically very distinct. As a careful evaluation of this issue would necessarily imply taking into account the southwest European input, far away from my aims, I caution that my results refer to a partial picture of North Africa pool, focused on the Near Eastern/Arabian input, as merely a way of contrasting with this input into East Africa.

In summary, a pattern seems to emerge: the Late Glacial period was more important for introductions of Eurasian lineages into eastern Africa, probably reflecting the higher impact of this period in the expansion of Arabian populations; while the Neolithic, especially linked to the Near East, affected to a great extent the migration pattern towards North Africa.

Future perspectives

In contrast to some other parts of the globe, where it has been possible to complement the genetic studies performed on extant populations with information from ancient DNA, no prehistoric human remains have been found yet in Arabia. It is likely to be very difficult to find well-preserved organic remains in the Arabian Peninsula in the near future, at least in sufficient numbers for accurate ancient DNA analyses. So, for the time being the only available strategy is to continue to pursue genetic studies on the extant Arabian population.

As already mentioned, we are very interested in performing genome-wide studies in Arabian Peninsula, which will allow to investigate the fine-scale population structure of this geographical paramount region for old worldwide migrations. Definitely, Dubai and Oman, leading places in the Gulf Oasis, can provide informative clues about the first descendents of the out-of-Africa migration.

Better statistical methods, allowing to evaluate admixture at a local basis along the autosomal chromosomes (such as HapMix; Price et al. 2009) are being developed, and potentially will contribute to improve genome-wide based inferences for admixture evaluation and dating of events. At the same time, these methods allow some insights into selection on phenotypes, so important for the adaptation of migrant populations to new environments and changing climatic conditions throughout time. It is not easy to demonstrate the effect of selection, as genetic drift and other factors can mimic the loss of diversity in a genomic region. Nevertheless, the long human settlement in Africa led to selection on genes related to resistance to malaria, lactose intolerance, bitter taste perception, and others. Given the admixed nature of the Arabian populations, it will be interesting to investigate selection on these and other phenotypes. Preliminary data[318] in Near Eastern and Eastern African populations have indicated convergent evolution for variants associated with the lactase persistence trait, an indication of local adaptation, supporting the need to investigate selection across diverse human populations.

As already referred to above, autosomal data will not replace uniparental markers as they provide complementary information. If one intends to evaluate the role played by gender in migrations, uniparental markers must be investigated. The results presented here for the impact of the slave trade in Arabia testifies to this. As for future perspectives, at the level of the uniparental markers, the application of founder analysis to complete mtDNA sequences could provide more precise dating of successive migrations. This would require increasing significantly the database of complete sequences in Arabia and neighbouring regions, which will become an

easier and less expensive task as next-generation sequencing techniques become more routine. Similarly, it would also be very informative to improve the resolution of Y-chromosome data in Arabia, by performing high-throughput sequencing of this chromosome.

11. Bibliography

1. Fernandes, V., Alshamali, F., Alves, M., Costa, M.D., Pereira, J.B., Silva, N.M., Cherni, L., Harich, N., Cerny, V., Soares, P., et al. (2012). The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa. Am J Hum Genet 90, 347-355.
2. Jobling, M.A., Hurles, M.E., and Tyler-Smith, C. (2003). Human Evolutionary Genetics: Origins, Peoples and Disease.(New York: Garland Science).
3. Watson, P.J. (2009). Archaeology and Anthropology: A Personal Overview of the Past Half-Century. Annual Review of Anthropology 38, 1-15.
4. Grün, R. (2006). Direct dating of human fossils. Am J Phys Anthropol 131, 2-48.
5. Reimer, P.J., Baillie, M.G.L., Bard, E., Bayliss, A., Beck, J.W., Blackwell, P.G., Ramsey, C.B., Buck, C.E., Burr, G.S., Edwards, R.L., et al. (2009). IntCal09 and Marine09 Radiocarbon Age Calibration Curves, 0-50,000 Years cal BP. Radiocarbon 51, 1111–1150.
6. Kohno, M. (2010). Applications of electron spin resonance spectrometry for reactive oxygen species and reactive nitrogen species research. J Clin Biochem Nutr 47, 1-11.
7. Foley, R. (1998). The context of human genetic evolution. Genome Res 8, 339-347.
8. Disotell, T.R. (2012). Archaic human genomics. Am J Phys Anthropol 149 Suppl 55, 24-39.
9. Clark, J.D., Beyene, Y., WoldeGabriel, G., Hart, W.K., Renne, P.R., Gilbert, H., Defleur, A., Suwa, G., Katoh, S., Ludwig, K.R., et al. (2003). Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 423, 747-752.
10. White, T.D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G.D., Suwa, G., and Howell, F.C. (2003). Pleistocene Homo sapiens from Middle Awash, Ethiopia. Nature 423, 742-747.
11. McDougall, I., Brown, F.H., and Fleagle, J.G. (2005). Stratigraphic placement and age of modern humans from Kibish, Ethiopia. Nature 433, 733-736.
12. Armitage, S.J., Jasim, S.A., Marks, A.E., Parker, A.G., Usik, V.I., and Uerpmann, H.P. (2011). The southern route "out of Africa": evidence for an early expansion of modern humans into Arabia. Science 331, 453-456.
13. deMenocal, P.B. (2011). Anthropology. Climate and human evolution. Science 331, 540-542.
14. Shackleton, N.J., Sánchez-Goñi, M.F., Pailler, D., and Lancelot, Y. (2003). Marine Isotope Substage 5e and the Eemian Interglacial. Global and Planetary Change 36, 151-155.
15. Scitable by Nature Education. (2013). In. Nature Education, a division of Nature America, Inc.
16. Cuffey, K.M., Clow, G.D., Alley, R.B., Stuiver, M., Waddington, E.D., and Saltus, R.W. (1995). Large Arctic Temperature Change at the Wisconsin-Holocene Glacial Transition. Science 270, 455-458.
17. Petit, J.R., Jouzel, J., Raynaud, D., Barkov, N.I., Barnola, J.M., Basile, I., Bender, M., Chappellaz, J., Davis, M., Delaygue, G., et al. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature 399, 429-436.
18. Metspalu, M., Kivisild, T., Bandelt, H.J., Richards, M., and Villems, R. (2006). The Pioneer Settlement of Modern Humans in Asia. In Human Mitochondrial DNA and the Evolution of Homo sapiens (Nucleic Acids and Molecular Biology), H.J. Gross, ed. (Berlin, Springer Berlin Heidelberg New York), pp 181-200.

19. Scholz, C.A., Johnson, T.C., Cohen, A.S., King, J.W., Peck, J.A., Overpeck, J.T., Talbot, M.R., Brown, E.T., Kalindekafe, L., Amoako, P.Y., et al. (2007). East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. Proc Natl Acad Sci U S A 104, 16416–16421.
20. Stewart, J.R., and Stringer, C.B. (2012). Human evolution out of Africa: the role of refugia and climate change. Science 335, 1317-1321.
21. Dunn, M., Greenhill, S.J., Levinson, S.C., and Gray, R.D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. Nature 473, 79-82.
22. Fitch, W.T. (2011). Unity and diversity in human language. Philos Trans R Soc Lond B Biol Sci 366, 376-388.
23. Baronchelli, A., Chater, N., Pastor-Satorras, R., and Christiansen, M.H. (2012). The biological origin of linguistic diversity. PLoS One 7, e48029.
24. Atkinson, Q.D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. Science 332, 346-349.
25. Stringer, C.B., and Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. Science 239, 1263-1268.
26. Cann, R.L., Stoneking, M., and Wilson, A.C. (1987). Mitochondrial DNA and human evolution. Nature 325, 31-36.
27. Brown, A.T. (2002). Genomes. In. Oxford, Wiley-Liss.
28. Watson, J.D. (1990). The human genome project: past, present, and future. Science 248, 44-49.
29. International Human Genome Sequencing, Consortium. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.
30. International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. Nature 431, 931-945.
31. Hardison, R.C. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet 16, 369-372.
32. ENCODE Project Consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816.
33. International HapMap Consortium. (2005). A haplotype map of the human genome. Nature 437, 1299-1320.
34. International HapMap Consortium. (2003). The International HapMap Project. Nature 426, 789-796.
35. International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851-861.
36. International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. Nature 467, 52-58.
37. Wilkins, J.F. (2006). Unraveling male and female histories from human genetic data. Curr Opin Genet Dev 16, 611-617.
38. Royal, C.D., Novembre, J., Fullerton, S.M., Goldstein, D.B., Long, J.C., Bamshad, M.J., and Clark, A.G. (2010). Inferring genetic ancestry: opportunities, challenges, and implications. Am J Hum Genet 86, 661-673.
39. Novembre, J., and Ramachandran, S. (2011). Perspectives on human population structure at the cusp of the sequencing era. Annu Rev Genomics Hum Genet 12, 245-274.
40. Genographic project - genetics overview. (2013). In. National Geographic.
41. Neiman, M., and Taylor, D.R. (2009). The causes of mutation accumulation in mitochondrial genomes. Proc Biol Sci 276, 1201-1209.
42. Ernster, L., and Schatz, G. (1981). Mitochondria: a historical review. J Cell Biol 91, 227s-255s.
43. Wallace, D.C. (1982). Structure and evolution of organelle genomes. Microbiol Rev 46, 208-240.

44. Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. Science 283, 1476-1481.
45. Boore, J.L. (1999). Animal mitochondrial genomes. Nucleic Acids Res 27, 1767-1780.
46. Picard, M., Shirihai, O.S., Gentil, B.J., and Burelle, Y. (2013). Mitochondrial morphology transitions and functions: implications for retrograde signaling? Am J Physiol Regul Integr Comp Physiol 304, R393-406.
47. Giles, R.E., Blanc, H., Cann, H.M., and Wallace, D.C. (1980). Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A 77, 6715-6719.
48. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. Nature 290, 457-465.
49. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23, 147.
50. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., et al. (2003). Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A 100, 171-176.
51. Meyer, S., Weiss, G., and von Haeseler, A. (1999). Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152, 1103-1110.
52. Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D., and Wilson, A.C. (1989). Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci U S A 86, 9350-9354.
53. Chinnery, P.F., and Hudson, G. (2013). Mitochondrial genetics. Br Med Bull 106, 135-159.
54. Peralta, S., Wang, X., and Moraes, C.T. (2012). Mitochondrial transcription: lessons from mouse models. Biochim Biophys Acta 1819, 961-969.
55. Brown, W.M., George, M., Jr., and Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A 76, 1967-1971.
56. Brown, W.M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci U S A 77, 3605-3609.
57. Kivisild, T., Metspalu, M., Bandelt, H.J., Richards, M., and Villems, R. (2006). TheWorld mtDNA Phylogeny. In Human Mitochondrial DNA and the Evolution of Homo sapiens (Nucleic Acids and Molecular Biology), H.J. Gross, ed. (Berlin, Springer Berlin Heidelberg New York), pp 149-180.
58. Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., and Schatten, G. (1999). Ubiquitin tag for sperm mitochondria. Nature 402, 371-372.
59. Schwartz, M., and Vissing, J. (2002). Paternal inheritance of mitochondrial DNA. N Engl J Med 347, 576-580.
60. Bandelt, H.J., Kong, Q.P., Parson, W., and Salas, A. (2005). More evidence for non-maternal inheritance of mitochondrial DNA? J Med Genet 42, 957-960.
61. Sato, M., and Sato, K. (2013). Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. Biochim Biophys Acta 1833, 1979-1984.
62. Awadalla, P., Eyre-Walker, A., and Smith, J.M. (1999). Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286, 2524-2525.
63. D'Aurelio, M., Gajewski, C.D., Lin, M.T., Mauck, W.M., Shao, L.Z., Lenaz, G., Moraes, C.T., and Manfredi, G. (2004). Heterologous mitochondrial DNA recombination in human cells. Hum Mol Genet 13, 3171-3179.

64. Kraytsberg, Y., Schwartz, M., Brown, T.A., Ebralidse, K., Kunz, W.S., Clayton, D.A., Vissing, J., and Khrapko, K. (2004). Recombination of human mitochondrial DNA. Science 304, 981.
65. Kivisild, T., Villems, R., Jorde, L.B., Bamshad, M., Kumar, S., Hedrick, P., Dowling, T., Stoneking, M., Parsons, T.J., Irwin, J.A., et al. (2000). Questioning evidence for recombination in human mitochondrial DNA. Science 288, 1931a.
66. Carelli, V., Achilli, A., Valentino, M.L., Rengo, C., Semino, O., Pala, M., Olivieri, A., Mattiazzi, M., Pallotti, F., Carrara, F., et al. (2006). Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees. Am J Hum Genet 78, 564-574.
67. Yao, Y.G., Bandelt, H.J., and Young, N.S. (2007). External contamination in single cell mtDNA analysis. PLoS One 2, e681.
68. Schapira, A.H. (2012). Mitochondrial diseases. Lancet 379, 1825-1834.
69. Bandelt, H.J., Yao, Y.G., Salas, A., Kivisild, T., and Bravi, C.M. (2007). High penetrance of sequencing errors and interpretative shortcomings in mtDNA sequence analysis of LHON patients. Biochem Biophys Res Commun 352, 283-291.
70. Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A.C., Mohyuddin, A., Mehdi, S.Q., Rosser, Z., Stoneking, M., Jobling, M.A., et al. (2004). A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74, 1183-1197.
71. Alessandrini, F., Turchi, C., Onofri, V., Buscemi, L., Pesaresi, M., and Tagliabracci, A. (2005). Multiplex PCR development of Y-chromosomal biallelic polymorphisms for forensic application. J Forensic Sci 50, 519-525.
72. Roewer, L., Croucher, P.J., Willuweit, S., Lu, T.T., Kayser, M., Lessig, R., de Knijff, P., Jobling, M.A., Tyler-Smith, C., and Krawczak, M. (2005). Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116, 279-291.
73. Fan, H., and Chu, J.Y. (2007). A brief review of short tandem repeat mutation. Genomics Proteomics Bioinformatics 5, 7-14.
74. Kimura, M., and Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci U S A 75, 2868-2872.
75. Gusmao, L., Sanchez-Diz, P., Calafell, F., Martin, P., Alonso, C.A., Alvarez-Fernandez, F., Alves, C., Borjas-Fajardo, L., Bozzo, W.R., Bravo, M.L., et al. (2005). Mutation rates at Y chromosome specific microsatellites. Hum Mutat 26, 520-528.
76. Burgarella, C., and Navascues, M. (2011). Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. Eur J Hum Genet 19, 70-75.
77. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E., and de Knijff, P. (1997). Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6, 799-803.
78. Ravid-Amir, O., and Rosset, S. (2010). Maximum likelihood estimation of locus-specific mutation rates in Y-chromosome short tandem repeats. Bioinformatics 26, i440-445.
79. Forster, P., Rohl, A., Lunnemann, P., Brinkmann, C., Zerjal, T., Tyler-Smith, C., and Brinkmann, B. (2000). A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67, 182-196.
80. Zhivotovsky, L.A., Underhill, P.A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., et al. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74, 50-61.

81. Walsh, B. (2001). Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. Genetics 158, 897-912.
82. Nachman, M.W., and Crowell, S.L. (2000). Estimate of the mutation rate per nucleotide in humans. Genetics 156, 297-304.
83. The, Genomes, Project, Consortium. (2010). A map of human genome variation from population-scale sequencing. Nature 467, 1061-1073.
84. Onofri, V., Alessandrini, F., Turchi, C., Pesaresi, M., Buscemi, L., and Tagliabracci, A. (2006). Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci Int 157, 23-35.
85. Edwards, A.W., and Cavalli-Sforza, L.L. (1965). A method for cluster analysis Biometrics 21, 362-375.
86. Cavalli-Sforza, L.L., Menozzi, P., and Piazza, A. (1996). The History and Geography of Human Genes.(Princeton: Princeton University Press).
87. Torroni, A., Achilli, A., Macaulay, V., Richards, M., and Bandelt, H.J. (2006). Harvesting the fruit of the human mtDNA tree. Trends Genet 22, 339-345.
88. Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.J., Torroni, A., and Richards, M.B. (2010). The archaeogenetics of Europe. Current biology : CB 20, R174-183.
89. Underhill, P.A., and Kivisild, T. (2007). Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 41, 539-564.
90. Ingman, M., Kaessmann, H., Paabo, S., and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. Nature 408, 708-713.
91. Howell, N., Elson, J.L., Turnbull, D.M., and Herrnstadt, C. (2004). African Haplogroup L mtDNA sequences show violations of clock-like evolution. Mol Biol Evol 21, 1843-1854.
92. Ho, S.Y., Phillips, M.J., Cooper, A., and Drummond, A.J. (2005). Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol 22, 1561-1568.
93. Rosenberg, N.A., and Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat Rev Genet 3, 380-390.
94. Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., and Torroni, A. (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64, 232-249.
95. Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E., et al. (2002). Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70, 1152-1171.
96. Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M., and Wallace, D.C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53, 563-590.
97. Richards, M.B., Macaulay, V.A., Bandelt, H.J., and Sykes, B.C. (1998). Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62, 241-260.
98. Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C., and Cabrera, V.M. (2001). Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2, 13.
99. Soares, P., Alshamali, F., Pereira, J.B., Fernandes, V., Silva, N.M., Afonso, C., Costa, M.D., Musilova, E., Macaulay, V., Richards, M.B., et al. (2012). The

expansion of mtDNA haplogroup L3 within and out of Africa. Mol Biol Evol 29, 915-927.

100. Richards, M., Bandelt, H.J., Kivisild, T., and Oppenheimer, S. (2006). A Model for the Dispersal ofModern Humans out of Africa. In Human Mitochondrial DNA and the Evolution of Homo sapiens (Nucleic Acids and Molecular Biology), H.J. Gross, ed. (Berlin, Springer Berlin Heidelberg New York), pp 225-266.
101. Pala, M., Olivieri, A., Achilli, A., Accetturo, M., Metspalu, E., Reidla, M., Tamm, E., Karmin, M., Reisberg, T., Hooshiar, K.B., et al. (2012). Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. Am J Hum Genet 90, 915-924.
102. Abu-Amero, K.K., Larruga, J.M., Cabrera, V.M., and Gonzalez, A.M. (2008). Mitochondrial DNA structure in the Arabian Peninsula. BMC evolutionary biology 8, 45.
103. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5, 26.
104. Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H.J., Torroni, A., and Richards, M.B. (2010). The archaeogenetics of Europe. Curr Biol 20, R174-183.
105. Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67, 1251-1276.
106. Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., Dorzhu, C., Luzina, F., Lee, H.K., Vanecek, T., et al. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian populations. Am J Hum Genet 81, 1025-1041.
107. Palanichamy, M.G., Zhang, C.L., Mitra, B., Malyarchuk, B., Derenko, M., Chaudhuri, T.K., and Zhang, Y.P. (2010). Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. BMC evolutionary biology 10, 304.
108. Richards, M., Rengo, C., Cruciani, F., Gratrix, F., Wilson, J.F., Scozzari, R., Macaulay, V., and Torroni, A. (2003). Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. Am J Hum Genet 72, 1058-1064.
109. Behar, D.M., Metspalu, E., Kivisild, T., Achilli, A., Hadid, Y., Tzur, S., Pereira, L., Amorim, A., Quintana-Murci, L., Majamaa, K., et al. (2006). The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. Am J Hum Genet 78, 487-497.
110. Feder, J., Blech, I., Ovadia, O., Amar, S., Wainstein, J., Raz, I., Dadon, S., Arking, D.E., Glaser, B., and Mishmar, D. (2008). Differences in mtDNA haplogroup distribution among 3 Jewish populations alter susceptibility to T2DM complications. BMC Genomics 9, 198.
111. Behar, D.M., Metspalu, E., Kivisild, T., Rosset, S., Tzur, S., Hadid, Y., Yudkovsky, G., Rosengarten, D., Pereira, L., Amorim, A., et al. (2008). Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 3, e2062.
112. Reidla, M., Kivisild, T., Metspalu, E., Kaldma, K., Tambets, K., Tolk, H.V., Parik, J., Loogvali, E.L., Derenko, M., Malyarchuk, B., et al. (2003). Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet 73, 1178-1190.
113. Shlush, L.I., Behar, D.M., Yudkovsky, G., Templeton, A., Hadid, Y., Basis, F., Hammer, M., Itzkovitz, S., and Skorecki, K. (2008). The Druze: a population genetic refugium of the Near East. PLoS One 3, e2105.

114. Nachman, M.W., Brown, W.M., Stoneking, M., and Aquadro, C.F. (1996). Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142, 953-963.
115. Elson, J.L., Turnbull, D.M., and Howell, N. (2004). Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection. Am J Hum Genet 74, 229-238.
116. Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V., and Richards, M.B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet 84, 740-759.
117. Forster, P., Harding, R., Torroni, A., and Bandelt, H.J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59, 935-945.
118. Gerber, A.S., Loggins, R., Kumar, S., and Dowling, T.E. (2001). Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Annu Rev Genet 35, 539-566.
119. Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A., et al. (2006). The role of selection in the evolution of human mitochondrial genomes. Genetics 172, 373-387.
120. Endicott, P., and Ho, S.Y. (2008). A Bayesian evaluation of human mitochondrial substitution rates. Am J Hum Genet 82, 895-902.
121. Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., and Oefner, P.J. (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7, 996-1005.
122. Trombetta, B., Cruciani, F., Sellitto, D., and Scozzari, R. (2011). A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. PLoS One 6, e16073.
123. The Y Chromosome Consortium. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12, 339-348.
124. Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18, 830-838.
125. Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon Lahr, M., Foley, R.A., Oefner, P.J., and Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65, 43-62.
126. Gonzalez, M., Gomes, V., Lopez-Parra, A.M., Amorim, A., Carracedo, A., Sanchez-Diz, P., Arroyo-Pardo, E., and Gusmao, L. (2013). The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88. Eur J Hum Genet 21, 324-331.
127. Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., and Scozzari, R. (2011). A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. Am J Hum Genet 88, 814-818.
128. Balaresque, P., Bowden, G.R., Adams, S.M., Leung, H.Y., King, T.E., Rosser, Z.H., Goodwin, J., Moisan, J.P., Richard, C., Millward, A., et al. (2010). A predominantly neolithic origin for European paternal lineages. PLoS Biol 8, e1000285.
129. Busby, G.B., Brisighelli, F., Sanchez-Diz, P., Ramos-Luis, E., Martinez-Cadenas, C., Thomas, M.G., Bradley, D.G., Gusmao, L., Winney, B., Bodmer, W., et al. (2012). The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc Biol Sci 279, 884-892.

130. Wei, W., Ayub, Q., Chen, Y., McCarthy, S., Hou, Y., Carbone, I., Xue, Y., and Tyler-Smith, C. (2013). A calibrated human Y-chromosomal phylogeny based on resequencing. Genome Res 23, 388-395.
131. Mendez, F.L., Krahn, T., Schrack, B., Krahn, A.M., Veeramah, K.R., Woerner, A.E., Fomine, F.L., Bradman, N., Thomas, M.G., Karafet, T.M., et al. (2013). An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. Am J Hum Genet 92, 454-459.
132. Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Wong, W.S., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature 488, 471-475.
133. Francalacci, P., Morelli, L., Angius, A., Berutti, R., Reinier, F., Atzeni, R., Pilu, R., Busonero, F., Maschio, A., Zara, I., et al. (2013). Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science 341, 565-569.
134. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100-1104.
135. Auton, A., Bryc, K., Boyko, A.R., Lohmueller, K.E., Novembre, J., Reynolds, A., Indap, A., Wright, M.H., Degenhardt, J.D., Gutenkunst, R.N., et al. (2009). Global distribution of genomic diversity underscores rich complex history of continental human populations. Genome Res 19, 795-803.
136. Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. Nature 466, 238-242.
137. 1000 Genomes A Deep Catalog of Human Genetic Variation. (2012). In. (The European Bioinformatics Institute. EMBL-EBI.
138. Ebersberger, I., Metzler, D., Schwarz, C., and Paabo, S. (2002). Genomewide comparison of DNA sequences between humans and chimpanzees. Am J Hum Genet 70, 1490-1497.
139. Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics 155, 945-959.
140. Tang, H., Peng, J., Wang, P., and Risch, N.J. (2005). Estimation of individual admixture: analytical and study design considerations. Genet Epidemiol 28, 289-301.
141. Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., and Myers, S. (2009). Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet 5, e1000519.
142. Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmon, G., Burns, E., Ostrer, H., Price, A.L., and Reich, D. (2011). The history of African gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet 7, e1001373.
143. Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., et al. (2008). Genes mirror geography within Europe. Nature 456, 98-101.
144. Bryc, K., Auton, A., Nelson, M.R., Oksenberg, J.R., Hauser, S.L., Williams, S., Froment, A., Bodo, J.M., Wambebe, C., Tishkoff, S.A., et al. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc Natl Acad Sci U S A 107, 786-791.
145. Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodriguez-Botigue, L., Ramachandran, S., Hon, L., Brisbin, A., et al. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci U S A 108, 5154-5162.

146. Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Guldemann, T., Kure, B., Mpoloka, S.W., Nakagawa, H., Naumann, C., et al. (2012). The genetic prehistory of southern Africa. Nat Commun 3, 1143.
147. Henn, B.M., Botigue, L.R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J.K., Fadhlaoui-Zid, K., Zalloua, P.A., Moreno-Estrada, A., Bertranpetit, J., et al. (2012). Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet 8, e1002397.
148. The, H., Pan-Asian, SNP, Consortium. (2009). Mapping human genetic diversity in Asia. Science 326, 1541-1545.
149. Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature 488, 370-374.
150. Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C.D., and Ostrer, H. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. Proc Natl Acad Sci U S A 107 Suppl 2, 8954-8961.
151. Price, A.L., Patterson, N., Yu, F., Cox, D.R., Waliszewska, A., McDonald, G.J., Tandon, A., Schirmer, C., Neubauer, J., Bedoya, G., et al. (2007). A genomewide admixture map for Latino populations. Am J Hum Genet 80, 1024-1036.
152. Paabo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., and Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annu Rev Genet 38, 645-679.
153. Goodwin, W., and Ovchinnikov, I. (2006). Ancient DNA and the Neanderthals. In Human Mitochondrial DNA and the Evolution of Homo sapiens (Nucleic Acids and Molecular Biology), H.J. Gross, ed. (Berlin, Springer Berlin Heidelberg New York), pp 201-224.
154. Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., and Wilson, A.C. (1984). DNA sequences from the quagga, an extinct member of the horse family. Nature 312, 282-284.
155. Paabo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. Nature 314, 644-645.
156. Clark, A.G. (2008). Genome sequences from extinct relatives. Cell 134, 388-389.
157. Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., and Paabo, S. (1997). Neandertal DNA sequences and the origin of modern humans. Cell 90, 19-30.
158. Ovchinnikov, I.V., Gotherstrom, A., Romanova, G.P., Kharitonov, V.M., Liden, K., and Goodwin, W. (2000). Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404, 490-493.
159. Krause, J., Orlando, L., Serre, D., Viola, B., Prufer, K., Richards, M.P., Hublin, J.J., Hanni, C., Derevianko, A.P., and Paabo, S. (2007). Neanderthals in central Asia and Siberia. Nature 449, 902-904.
160. Green, R.E., Malaspinas, A.S., Krause, J., Briggs, A.W., Johnson, P.L., Uhler, C., Meyer, M., Good, J.M., Maricic, T., Stenzel, U., et al. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell 134, 416-426.
161. Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., et al. (2006). Analysis of one million base pairs of Neanderthal DNA. Nature 444, 330-336.
162. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft sequence of the Neandertal genome. Science 328, 710-722.

163. Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Paabo, S., Pritchard, J.K., et al. (2006). Sequencing and analysis of Neanderthal genomic DNA. Science 314, 1113-1118.
164. Ermini, L., Olivieri, C., Rizzi, E., Corti, G., Bonnal, R., Soares, P., Luciani, S., Marota, I., De Bellis, G., Richards, M.B., et al. (2008). Complete mitochondrial genome sequence of the Tyrolean Iceman. Curr Biol 18, 1687-1693.
165. Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., et al. (2012). New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun 3, 698.
166. Gilbert, M.T., Kivisild, T., Gronnow, B., Andersen, P.K., Metspalu, E., Reidla, M., Tamm, E., Axelsson, E., Gotherstrom, A., Campos, P.F., et al. (2008). Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320, 1787-1789.
167. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463, 757-762.
168. Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Jane Adler, C., Richards, S.M., Sarkissian, C.D., Ganslmeier, R., Friederich, S., et al. (2013). Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nat Commun 4, 1764.
169. Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., and Paabo, S. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature 464, 894-897.
170. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468, 1053-1060.
171. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science 338, 222-226.
172. Blum, M.G., and Jakobsson, M. (2011). Deep divergences of human gene trees and models of human origins. Mol Biol Evol 28, 889-898.
173. Wolpoff, M.H., Hawks, J., and Caspari, R. (2000). Multiregional, not multiple origins. Am J Phys Anthropol 112, 129-136.
174. Relethford, J.H. (2008). Genetic evidence and the modern human origins debate. Heredity 100, 555-563.
175. Stringer, C. (2002). Modern human origins: progress and prospects. Philos Trans R Soc Lond B Biol Sci 357, 563-579.
176. Barham, L., and Mitchell, P. (2008). The first Africans: African archaeology from the earliest tool makers to most recent foragers.(Cambridge University Press Cambridge).
177. Smith, T.M., Tafforeau, P., Reid, D.J., Grun, R., Egginsll, S., Boutakiout, M., and Hublin, J.J. (2007). Earliest evidence of modern human life history in North African early Homo sapiens. P Natl Acad Sci USA 104, 6128-6133.
178. Stringer, C. (2011). The origin of our species.(Penguin).
179. Millard, A.R. (2008). A critique of the chronometric evidence for hominid fossils: I. Africa and the Near East 500–50ka. J Hum Evol 54, 848-874.
180. Rightmire, G.P., Deacon, H., Schwartz, J.H., and Tattersall, I. (2006). Human foot bones from Klasies River main site, South Africa. J Hum Evol 50, 96-103.
181. Chiaroni, J., Underhill, P.A., and Cavalli-Sforza, L.L. (2009). Y chromosome diversity, human expansion, drift, and cultural evolution. Proc Natl Acad Sci U S A 106, 20174-20179.
182. Mellars, P. (2006). Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science 313, 796-800.

183. Beyin, A. (2011). Upper Pleistocene Human Dispersals out of Africa: A Review of the Current State of the Debate. Int J Evol Biol 2011, 615094.
184. Tchernov, E. (1992). The Afro-Arabian component in the Levantine Mammalian fauna - A short biogeographical review. Israel Journal of Zoology 38, 155-192.
185. Shea, J.J. (2007). The boulevard of broken dreams: evolutionary discontinuity in the Late Pleistocene Levant. In The human revolution revisited, P. Mellars, K. Boyle, O. Bar-Yosef, andC. Stringer, eds. (Cambridge, McDonald Institute Archaeological Publications), pp 219–232.
186. Stringer, C.B., Grun, R., Schwarcz, H.P., and Goldberg, P. (1989). ESR dates for the hominid burial site of Es Skhul in Israel. Nature 338, 756-758.
187. Forster, P., and Matsumura, S. (2005). Evolution. Did early humans go north or south? Science 308, 965-966.
188. Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., Scozzari, R., Cruciani, F., Behar, D.M., Dugoujon, J.M., et al. (2006). The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. Science 314, 1767-1770.
189. Forster, P., and Romano, V. (2007). Timing of a back-migration into Africa. Science 316, 50-53.
190. Pennarun, E., Kivisild, T., Metspalu, E., Metspalu, M., Reisberg, T., Moisan, J.P., Behar, D.M., Jones, S.C., and Villems, R. (2012). Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa. BMC Evol Biol 12, 234.
191. Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308, 1034-1036.
192. Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A., and Singh, L. (2005). Reconstructing the origin of Andaman Islanders. Science 308, 996.
193. Petraglia, M.D., and Alsharekh, A. (2003). The Middle Palaeolithic of Arabia: Implications for modern human origins, behaviour and dispersals Antiquity 77, 671–684
194. Siddall, M., Rohling, E.J., Almogi-Labin, A., Hemleben, C., Meischner, D., Schmelzer, I., and Smeed, D.A. (2003). Sea-level fluctuations during the last glacial cycle. Nature 423, 853-858.
195. Sirocko, F. (2003). Ups and downs in the Red Sea. Nature 423, 813-814.
196. Bailey, G. (2009). The Red Sea, Coastal Landscapes, and Hominin Dispersals. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 15–38.
197. Abu-Amero, K.K., Gonzalez, A.M., Larruga, J.M., Bosley, T.M., and Cabrera, V.M. (2007). Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC evolutionary biology 7, 32.
198. Rose, J.I. (2004). The Question of Upper Pleistocene Connections between East Africa and South Arabia. Current Anthropology 45, 551-555.
199. Rose, J., and Petraglia, M.D. (2009). Tracking the Origin and Evolution of Human Populations in Arabia. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 1-14.
200. Masry, A.H. (1997). Prehistory In Northeastern Arabia - The problem of Interregional Interaction.(London and New York: Kegan Paul International), pp 256.
201. Hogarth, D.G. (1904). The Penetration of Arabia: A Record of the Development of Western Knowledge Concerning the Arabian Peninsula (New York: Cambridge Library Collection), pp 359.

202. Parker, A.G. (2009). Pleistocene Climate Change in Arabia: Developing a Framework for Hominin Dispersal over the Last 350 ka. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 39-50.
203. Thompson, A. (2000). Origins of Arabia (London: Stacey International London), pp 108.
204. Wilkinson, T.J. (2009). Environment and Long-Term Population Trends in Southwest Arabia. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 51-68.
205. Ambrose, S.H. (1998). Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. J Hum Evol 34, 623-651.
206. Petraglia, M., Korisettar, R., Boivin, N., Clarkson, C., Ditchfield, P., Jones, S., Koshy, J., Lahr, M.M., Oppenheimer, C., Pyle, D., et al. (2007). Middle Paleolithic assemblages from the Indian subcontinent before and after the Toba super-eruption. Science 317, 114-116.
207. Kivisild, T., Reidla, M., Metspalu, E., Rosa, A., Brehm, A., Pennarun, E., Parik, J., Geberhiwot, T., Usanga, E., and Villems, R. (2004). Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 75, 752-770.
208. Lovejoy, P.E. (1983). Transformations inSlavery - A history of slavery in Africa. In. (New York, Cambridge University Press), pp 200.
209. Segal, R. (2002). Islam's Black Slaves - A history of Africa's other black diaspora (London: Atlantic Books), pp 288.
210. Faber, A. (1997). Genetic subgrouping of the Semitic languages. In The Semitic Languages, R. Hetzron, ed. (London Routledge).
211. Hetzron, R. (1976). Two principles of genetic reconstruction. In http://dxdoiorg/101016/jbbr201103031. pp 89–104.
212. Kitchen, A., Ehret, C., Assefa, S., and Mulligan, C.J. (2009). Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. Proc Biol Sci 276, 2703-2710.
213. Lovell, A., Moreau, C., Yotova, V., Xiao, F., Bourgeois, S., Gehl, D., Bertranpetit, J., Schurr, E., and Labuda, D. (2005). Ethiopia: between Sub-Saharan Africa and western Eurasia. Ann Hum Genet 69, 275-287.
214. Boivin, N., Blench, R., and Fuller, D.Q. (2009). Archaeological, Linguistic and Historical Sources on Ancient Seafaring: A Multidisciplinary Approach to the Study of Early Maritime Contact and Exchange in the Arabian peninsula. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 251-278.
215. Crassard, R. (2009). The Middle Paleolithic of Arabia: The View from the Hadramawt Region, Yemen. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 151-168.
216. Whalen, N.M., and Pease, D.W. (1992). Early mankind in Arabia. 43, 16–23.
217. Whalen, N., Killick, A., James, N., Morsi, G., and Kamal, M. (1981). Saudi Arabian archaeological reconnaissance 1980: B. Preliminary report on the Western Province survey. Atlal 5, 43–58.
218. Whalen, N.M., Siraj-Ali, J., Sindi, H.O., Pease, D.W., and Badein, M.A. (1988). A complex of sites in the Jeddah–Wadi Fatimah area. Atlal 11, 77–85.
219. Zarins, J., Whalen, N., Ibrāham, M., Jawad Mursi, A.A., and Khan, M. (1980). Comprehensive archeological survey program, preliminary report on the Central and Southwestern Provinces survey. Atlal 4, 9–36.
220. Whalen, N., and Pease, D.W. (1991). Archaeological Survey in Southwest Yemen. Paléorient 17, 127-131.

221. Amirkhanov, H. (1994). Research on the Palaeolithic and Neolithic of Hadramaut and Mahra. Arab Archaeol Epigr 5, 217–228.
222. Bulgarelli, G.M. (1987). Evidence of Palaeolithic industries in Northern Yemen. In Yemen: 3,000 years of art and civilization in Arabia Felix, W. Daum, ed. (Frankfurt, Umschau-Verlag), pp 32–33.
223. de Maigret, A. (2002). Arabia Felix: an exploration of the Archaeological History of Yemen.(London: Stacey International).
224. Petraglia, M.D., Drake, N., and Alsharekh, A. (2009). Acheulean Landscapes and Large Cutting Tool Assemblages in the Arabian peninsula. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 103-116.
225. Burns, S.J., Fleitmann, D., Matter, A., Kramers, J., and Al-Subbary, A.A. (2003). Indian Ocean climate and an absolute chronology over Dansgaard/Oeschger events 9 to 13. Science 301, 1365-1367.
226. Delagnes, A., Tribolo, C., Bertran, P., Brenet, M., Crassard, R., Jaubert, J., Khalidi, L., Mercier, N., Nomade, S., Peigne, S., et al. (2012). Inland human settlement in southern Arabia 55,000 years ago. New evidence from the Wadi Surdud Middle Paleolithic site complex, western Yemen. J Hum Evol 63, 452-474.
227. Rose, J.I., Usik, V.I., Marks, A.E., Hilbert, Y.H., Galletti, C.S., Parton, A., Geiling, J.M., Cerny, V., Morley, M.W., and Roberts, R.G. (2011). The Nubian Complex of Dhofar, Oman: an African middle stone age industry in Southern Arabia. PLoS One 6, e28239.
228. Crassard, R., and Hilbert, Y.H. (2013). A Nubian Complex Site from Central Arabia: Implications for Levallois Taxonomy and Human Dispersals during the Upper Pleistocene. PLoS One 8, e69221.
229. Petraglia, M.D. (2011). Archaeology: Trailblazers across Arabia. Nature 470, 50-51.
230. Petraglia, M.D., Alsharekh, A., Breeze, P., Clarkson, C., Crassard, R., Drake, N.A., Groucutt, H.S., Jennings, R., Parker, A.G., Parton, A., et al. (2012). Hominin Dispersal into the Nefud Desert and Middle Palaeolithic Settlement along the Jubbah Palaeolake, Northern Arabia. PLoS One 7, e49840.
231. Crassard, R., Petraglia, M.D., Drake, N.A., Breeze, P., Gratuze, B., Alsharekh, A., Arbach, M., Groucutt, H.S., Khalidi, L., Michelsen, N., et al. (2013). Middle Palaeolithic and Neolithic Occupations around Mundafan Palaeolake, Saudi Arabia: Implications for Climate Change and Human Dispersals. PLoS One 8, e69665.
232. Tosi, M. (1985). Archaeological activities in the Yemen Arab Republic. Tihamah coastal archaeological survey. East and West 35, 363–369.
233. Tosi, M. (1986). Archaeological activities in the Yemen Arab Republic. Neolithic and protohistoric cultures. Survey and excavation on the coastal plain (Tihamah). East and West 36, 400–414.
234. Zarins, J., and Zahrani, A. (1985). Recent archaeological investigations in the southern Tihama Plain. The sites of Athar and Sihi, 1404/1984. Atlal 9, 65–107.
235. Zarins, J., and Badr, H. (1986). Archaeological investigations in the Tihama plain II 1405/1985. Atlal 10, 36–57.
236. Khalidi, L. (2005). The prehistoric and early historic settlement patterns of the Tihamah coastal plain (Yemen): preliminary findings of the Tihamah Coastal Survey 2003. Proceedings of the Seminar for Arabian Studies 35, 115–127.
237. Fedele, F.G., and Zaccara, D. (2005). Wadi al-Tayyila 3: a mid-Holocene site on the Yemen plateau and its lithic collection. In Sabaean studies Archaeological, epigraphical and historical studies in honour of Yusuf M ‘Abdullah, Alessandro de Maigret, and Christian J Robin A.M. Sholan, S.

Antonini, andB. Arbach, eds. (Naples, Naples and San ‘a: University of Naples).
238. Zeuner, F.E. (1954). “Neolithic” sites from the Rub al-Khali, southern Arabia. Man 54, 133–136.
239. McCorriston, J. (2000). Early settlement in Hadramawt: preliminary report on prehistoric occupation at Shi’b Munayder. Arabian Archaeology and Epigraphy 11, 129–153.
240. Crassard, R., and Philipp, D. (2013). Invited editors’ preface. R. Crassard and D. Philipp, eds. (Arab. arch. epig), pp 1–2.
241. Uerpmann, H.-P., Potts, D.T., and Uerpmann, M. (2009). Holocene (Re-) Occupation of Eastern Arabia. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 205–214.
242. Fedele, F.G. (2009). Early Holocene in the Highlands: Data on the Peopling of the Eastern Yemen Plateau, with a Note on the Pleistocene Evidence. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 215-236.
243. Crassard, R., Petraglia, M.D., Parker, A.G., Parton, A., Roberts, R.G., Jacobs, Z., Alsharekh, A., Al-Omari, A., Breeze, P., Drake, N.A., et al. (2013). Beyond the levant: first evidence of a pre-pottery neolithic incursion into the nefud desert, saudi arabia. PLoS One 8, e68061.
244. McCorriston, J., and Martin, L. (2009). Southern Arabia’s Early Pastoral Population History: Some Recent Evidence. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 237-250.
245. Marshall, F., and Hildebrand, E. (2002). Cattle before crops: the beginnings of food production in Africa. Journal of World Prehistory 16, 99–143.
246. Fuller, D.Q. (2006). Agricultural origins and frontiers in South Asia: a working synthesis. Journal of World Prehistory 20, 1–86.
247. Zohary, D., and Hopf, M. (2000). Domestication of plants in the Old World.(Oxford: Oxford University Press), pp 328.
248. Biagi, P., Torke, W., Tosi, M., and H.-P., U. (1984). Qurum: a case study of coastal archaeology in northern Oman. World Archaeology 16, 43.
249. Mitchell, P. (2005). African connections: archaeological perspectives on Africa and the wider world.(Walnut Creek:Altamira Press), pp 328.
250. Ray, P.H. (2003). The Archaeology of seafaring in ancient South Asia.(Cambridge: Cambridge University Press), pp 350.
251. Cadenas, A.M., Zhivotovsky, L.A., Cavalli-Sforza, L.L., Underhill, P.A., and Herrera, R.J. (2008). Y-chromosome diversity characterizes the Gulf of Oman. Eur J Hum Genet 16, 374-386.
252. Cerny, V., Mulligan, C.J., Ridl, J., Zaloudkova, M., Edens, C.M., Hajek, M., and Pereira, L. (2008). Regional differences in the distribution of the sub-Saharan, West Eurasian, and South Asian mtDNA lineages in Yemen. Am J Phys Anthropol 136, 128-137.
253. Alshamali, F., Brandstatter, A., Zimmermann, B., and Parson, W. (2008). Mitochondrial DNA control region variation in Dubai, United Arab Emirates. Forensic Sci Int Genet 2, e9-10.
254. Cerny, V., Pereira, L., Kujanova, M., Vasikova, A., Hajek, M., Morris, M., and Mulligan, C.J. (2009). Out of Arabia-the settlement of island Soqotra as revealed by mitochondrial and Y chromosome genetic diversity. Am J Phys Anthropol 138, 439-447.
255. Luis, J.R., Rowold, D.J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., Underhill, P.A., Cavalli-Sforza, L.L., and Herrera, R.J. (2004). The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Am J Hum Genet 74, 532-544.

256. Rowold, D.J., Luis, J.R., Terreros, M.C., and Herrera, R.J. (2007). Mitochondrial DNA geneflow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. J Hum Genet 52, 436-447.
257. Al-Abri, A., Podgorna, E., Rose, J.I., Pereira, L., Mulligan, C.J., Silva, N.M., Bayoumi, R., Soares, P., and Cerny, V. (2012). Pleistocene-Holocene boundary in Southern Arabia from the perspective of human mtDNA variation. Am J Phys Anthropol 149, 291-298.
258. Cerny, V., Mulligan, C.J., Fernandes, V., Silva, N.M., Alshamali, F., Non, A., Harich, N., Cherni, L., El Gaaied, A.B., Al-Meeri, A., et al. (2011). Internal diversification of mitochondrial haplogroup R0a reveals post-last glacial maximum demographic expansions in South Arabia. Mol Biol Evol 28, 71-78.
259. Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., et al. (2000). Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67, 1251-1276.
260. Musilova, E., Fernandes, V., Silva, N.M., Soares, P., Alshamali, F., Harich, N., Cherni, L., Gaaied, A.B., Al-Meeri, A., Pereira, L., et al. (2011). Population history of the Red Sea--genetic exchanges between the Arabian Peninsula and East Africa signaled in the mitochondrial DNA HV1 haplogroup. Am J Phys Anthropol 145, 592-598.
261. Abu-Amero, K.K., Hellani, A., Gonzalez, A.M., Larruga, J.M., Cabrera, V.M., and Underhill, P.A. (2009). Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC genetics 10, 59.
262. Tofanelli, S., Ferri, G., Bulayeva, K., Caciagli, L., Onofri, V., Taglioli, L., Bulayev, O., Boschi, I., Alu, M., Berti, A., et al. (2009). J1-M267 Y lineage marks climate-driven pre-historical human displacements. Eur J Hum Genet 17, 1520-1524.
263. Hunter-Zinck, H., Musharoff, S., Salit, J., Al-Ali, K.A., Chouchane, L., Gohar, A., Matthews, R., Butler, M.W., Fuller, J., Hackett, N.R., et al. (2010). Population genetic structure of the people of Qatar. Am J Hum Genet 87, 17-25.
264. Cerny, V., Fernandes, V., Costa, M.D., Hajek, M., Mulligan, C.J., and Pereira, L. (2009). Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. BMC Evol Biol 9, 63.
265. Pereira, L., Silva, N.M., Franco-Duarte, R., Fernandes, V., Pereira, J.B., Costa, M.D., Martins, H., Soares, P., Behar, D.M., Richards, M.B., et al. (2010). Population expansion in the North African late Pleistocene signalled by mitochondrial DNA haplogroup U6. BMC Evol Biol 10, 390.
266. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30, E386-394.
267. Kloss-Brandstatter, A., Pacher, D., Schonherr, S., Weissensteiner, H., Binna, R., Specht, G., and Kronenberg, F. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat 32, 25-32.
268. Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J., and Allen, R.C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am J Hum Genet 48, 137-144.
269. Pereira, L., Freitas, F., Fernandes, V., Pereira, J.B., Costa, M.D., Costa, S., Maximo, V., Macaulay, V., Rocha, R., and Samuels, D.C. (2009). The diversity present in 5140 human mitochondrial genomes. Am J Hum Genet 84, 628-640.
270. van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30, E386-394.

271. Alexe, G., Satya, R.V., Seiler, M., Platt, D., Bhanot, T., Hui, S., Tanaka, M., Levine, A.J., and Bhanot, G. (2008). PCA and clustering reveal alternate mtDNA phylogeny of N and M clades. J Mol Evol 67, 465-487.
272. Bandelt, H.J., Forster, P., Sykes, B.C., and Richards, M.B. (1995). Mitochondrial portraits of human populations using median networks. Genetics 141, 743-753.
273. Forster, M., and Forster, P. Phylogenetic Network software (1999). In. Fluxus Technology Ltd.
274. Bandelt, H.J., Macaulay, V., and Richards, M. (2000). Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol Phylogenet Evol 16, 8-28.
275. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.J., and Norby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67, 718-726.
276. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13, 555-556.
277. Aitkenhead, M.J., and Aalders, I.H. (2009). Predicting land cover using GIS, Bayesian and evolutionary algorithm methods. J Environ Manage 90, 236-250.
278. Drummond, A.J., Rambaut, A., Shapiro, B., and Pybus, O.G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 22, 1185-1192.
279. Drummond, A.F., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7, 214.
280. Drummond, A.F., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29, 1969-1973.
281. Atkinson, Q.D., Gray, R.D., and Drummond, A.J. (2009). Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. Proc Biol Sci 276, 367-373.
282. Thomson, N. (2009). Bayesian mixture modelling of migration by founder analysis. Doctor of Philosophy, University of Glasgow, Glasgow.
283. Alves, M., Alves, J., Camacho, R., Soares, P., and Pereira, L. (2012). From Networks to Trees. In PACBB, Advances in Intelligent and Soft Computing, Springer, ed. (Salamanca-Spain), pp 129-136.
284. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19, 1655-1664.
285. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet 2, e190.
286. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics 192, 1065-1093.
287. Barker, G., Barton, H., Bird, M., Daly, P., Datan, I., Dykes, A., Farr, L., Gilbertson, D., Harrisson, B., Hunt, C., et al. (2007). The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). J Hum Evol 52, 243–261.
288. Costa, M.D., Pereira, J.B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., Perego, U.A., Rychkov, S., Naumova, O., Hatina, J., et al. (2013). A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. Nat Commun 4, 2543.
289. Pereira, J.B. (2013). Genetic characterisation of modern human dispersals in the Greater Mediterranean. Doctor of Philosophy, University of Leeds, Leeds.
290. Palanichamy, M.G., Sun, C., Agrawal, S., Bandelt, H.J., Kong, Q.P., Khan, F., Wang, C.Y., Chaudhuri, T.K., Palla, V., and Zhang, Y.P. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966-978.

291. Lahr, M.M., and Foley, R. (1994). Multiple dispersals and modern human origins. Evolutionary Anthropology: Issues, News, and Reviews 3, 48-60.
292. Scally, A., and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. Nat Rev Genet 13, 745-753.
293. Mellars, P., Gori, K.C., Carr, M., Soares, P.A., and Richards, M.B. (2013). Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. Proc Natl Acad Sci U S A 110, 10699-10704.
294. Soares, P., Trejaut, J.A., Loo, J.H., Hill, C., Mormina, M., Lee, C.L., Chen, Y.M., Hudjashov, G., Forster, P., Macaulay, V., et al. (2008). Climate change and postglacial human dispersals in southeast Asia. Mol Biol Evol 25, 1209-1218.
295. Bar-Yosef, O. (1992). The role of western Asia in modern human origins. Philos Trans R Soc Lond B Biol Sci 337, 193-200.
296. Henshilwood, C.S., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G.A., Mercier, N., Sealy, J.C., Valladas, H., Watts, I., et al. (2002). Emergence of modern human behavior: Middle Stone Age engravings from South Africa. Science 295, 1278-1280.
297. Texier, P.J., Porraz, G., Parkington, J., Rigaud, J.P., Poggenpoel, C., Miller, C., Tribolo, C., Cartwright, C., Coudenneau, A., Klein, R., et al. (2010). From the Cover: A Howiesons Poort tradition of engraving ostrich eggshell containers dated to 60,000 years ago at Diepkloof Rock Shelter, South Africa. Proc Natl Acad Sci U S A 107, 6180-6185.
298. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D.M., Gilbert, M.T., et al. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet 5, 26.
299. Shea, J.J. (2008). The Lower and Middle Paleolithic in the Middle East and neighboring regions. Evol Anthropol 17, 205–207.
300. Ashrafian-Bonab, M., Lawson Handley, L.J., and Balloux, F. (2007). Is urbanization scrambling the genetic structure of human populations? A case study. Heredity (Edinb) 98, 151-156.
301. Van Andel, T.H., and Tzedakis, P.C. (1996). Palaeolithic landscapes of Europe and environs: 150,000-25,000 years ago: an overview. Quaternary Science Reviews 15, 481-500.
302. Rose, J.I. (2010). New Light on Human Prehistory in the Arabo-Persian Gulf Oasis. Current Anthropology 51, 849-883.
303. Turney, C.S.M., Bird, M.I., Fifield, L.K., Roberts, R.G., Smith, M.A., Dortch, C.E., Grün, R., Lawson, E., Ayliffe, L.K., Miller, G.H., et al. (2001). Early human occupation at Devil' Lair, southwestern Australia 50,000 years ago. Quaternary Research 55, 3–13.
304. Maher, L.A. (2009). The Late Pleistocene of Arabia in Relation to the Levant. In The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics, M.D. Petraglia and J. Rose, eds. (The Netherlands, Springer), pp 187-204.
305. Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Tanzer, M., Villems, R., Renfrew, C., Gronenborn, D., Alt, K.W., et al. (2005). Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. Science 310, 1016-1018.
306. Olivieri, A., Pala, M., Gandini, F., Hooshiar Kashani, B., Perego, U.A., Woodward, S.R., Grugni, V., Battaglia, V., Semino, O., Achilli, A., et al. (2013). Mitogenomes from two uncommon haplogroups mark late glacial/postglacial expansions from the near east and neolithic dispersals within Europe. PLoS One 8, e70492.
307. Achilli, A., Perego, U.A., Lancioni, H., Olivieri, A., Gandini, F., Hooshiar Kashani, B., Battaglia, V., Grugni, V., Angerhofer, N., Rogers, M.P., et al.

(2013). Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proc Natl Acad Sci U S A 110, 14308-14313.
308. Bar-Yosef, O. (1998). The Natufian culture in the Levant, threshold to the origins of agriculture. Evolutionary Anthropology: Issues, News, and Reviews 6, 159-177.
309. Pereira, L., Prata, M.J., and Amorim, A. (2000). Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet 64, 491-506.
310. Pereira, L., Cunha, C., Alves, C., and Amorim, A. (2005). African female heritage in Iberia: a reassessment of mtDNA lineage distribution in present times. Hum Biol 77, 213-229.
311. Cerezo, M., Achilli, A., Olivieri, A., Perego, U.A., Gomez-Carballa, A., Brisighelli, F., Lancioni, H., Woodward, S.R., Lopez-Soto, M., Carracedo, A., et al. (2012). Reconstructing ancient mitochondrial DNA links between Africa and Europe. Genome Res 22, 821-826.
312. Harich, N., Costa, M.D., Fernandes, V., Kandil, M., Pereira, J.B., Silva, N.M., and Pereira, L. (2010). The trans-Saharan slave trade - clues from interpolation analyses and high-resolution characterization of mitochondrial DNA lineages. BMC Evol Biol 10, 138.
313. Thomas, H. (1998). The Slave Trade: A History of the Atlantic Slave Trade 1440-1870.(London: Papermac), pp 912.
314. Arredi, B., Poloni, E.S., Paracchini, S., Zerjal, T., Fathallah, D.M., Makrelouf, M., Pascali, V.L., Novelletto, A., and Tyler-Smith, C. (2004). A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. Am J Hum Genet 75, 338-345.
315. Kujanova, M., Pereira, L., Fernandes, V., Pereira, J.B., and Cerny, V. (2009). Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert. Am J Phys Anthropol 140, 336-346.
316. Ennafaa, H., Cabrera, V.M., Abu-Amero, K.K., Gonzalez, A.M., Amor, M.B., Bouhaha, R., Dzimiri, N., Elgaaied, A.B., and Larruga, J.M. (2009). Mitochondrial DNA haplogroup H structure in North Africa. BMC Genet 10, 8.
317. Pereira, L., Cerny, V., Cerezo, M., Silva, N.M., Hajek, M., Vasikova, A., Kujanova, M., Brdicka, R., and Salas, A. (2010). Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel. Eur J Hum Genet 18, 915-923.
318. Scheinfeldt, L.B., and Tishkoff, S.A. (2013). Recent human adaptation: genomic approaches, interpretation and insights. Nat Rev Genet 14, 692-702.

VII. APPENDIX

Table S 1 Mitochondrial haplotypes for the complete sequences that were fully characterised in this study and the corresponding geographic region.

Sample Id	Geographic region	Haplogroup	Haplotype
AZB28	Azerbaijan	X2	73 153 195 198 249dA 263 315.1C 750 794 1438 1719 2706 4769 6221 6371 6635 7028 7205 8860 9078 11719 12447 12705 13884T 13966 14178 14470 14766 15119 15326 16183C 16189 16223 16261 16278 16519
AZB43	Azerbaijan	X2	73 153 195 198 249dA 263 315.1C 750 794 1438 1719 2706 4769 6221 6371 6635 7028 7205 8860 9078 11719 12447 12705 13884T 13966 14178 14470 14766 15119 15326 16183C 16189 16223 16261 16278 16519
AZB5	Azerbaijan	W3	73 119 194 195 204 207 263 315.1C 709 750 1243 1406 1438 2706 3505 4769 5046 5460 7028 7853 8860 8994 11017 11674 11719 11947 12414 12705 14766 15326 15884C 16223 16292 16295 16304 16519
BUL3847	Bulgaria	N1b1	73 152 195 263 309.1C 315.1C 750 769 1438 1598 1703 1719 2639 2706 3921A 4461 4769 4960 5471 7028 8251 8472 8836 8860 9116 9335 10238 11362 11719 12501 12705 12771 12822 14766 15326 16126 16145 16176G 16223 16390 16519
Bul7062	Bulgaria	I5a1	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 4529T 4769 5074 6593 7028 8251 8860 10034 10238 10398 11719 12501 12705 12771 12961 13780 14233 14766 15043 15326 15924 16129 16148 16223 16391 16519
BUL7063	Bulgaria	W3	73 189 194 195 199 204 207 210 263 309.1C 315.1C 709 750 1243 1406 1438 2706 3505 4769 5046 5460 7028 8251 8860 8994 11674 11719 11947 12414 12705 12923T 14766 14767 15326 15884C 16172 16223 16231 16292
CHE13	Chechnya	I2	73 152 199 204 207 250 263 309.1C 315.1C 503 573.1C 750 1438 1719 2706 4529T 4769 6227 7028 8251 8860 10034 10238 10398 11719 12303 12501 12603 12705 13780 14766 15043 15326 15758 15924 16129 16223 16391 16519
CHE14	Chechnya	X2e2	73 153 195 200 263 309.1C 315.1C 750 1438 1719 2706 3948 4769 6221 6371 7028 8860 11719 12084 12705 13966 14470 14766 15310 15326 16183C 16189 16223 16278 16519
CHV11	Chuvash	I1a	73 199 204 250 263 309.1C 315.1C 455.1T 573.1C 750 1438 1719 2706 3447 4529T 4769 4964 6734 7028 8251 8616T 8860 9966 10034 10238 10398 11719 12501 12705 13780 14182 14766 15043 15326 15924 16129 16172 16223 16255 16311 16391 16519
CZ303	Czech Republic	I2	73 152 199 204 207 250 263 309.1C 315.1C 573.1C 750 1438 1719 2706 4529T 4769 6480 7028 8251 8860 10034 10238 10398 11719 12501 12705 13780 14766 15043 15326 15758 15924 16129 16223 16391 16519
CZ317	Czech Republic	I1a1	73 199 203 204 250 263 309.1C 315.1C 455.1T 573.1C 750 1438 1719 2706 3447 3834 3990 4454 4529T 4769 6734 7028 7075C 8251 8616T 8860 9053 9947 10034 10238 10398 10915 11719 12501 12705 13780 14766 15043 15326 15758 15924 16129 16172 16223 16311 16319 16391 16519
CZ333	Czech Republic	I1a1	73 199 203 204 250 263 309.1C 315.1C 573.1C 750 1438 1719 2706 3447 3990 4529T 4769 6734 7028 8251 8616T 8860 9053 9947 10034 10238 10398 10915 11719 12084 12501 12705 13443 13780 14766 15043 15326 15924 16129 16172 16223 16274 16311 16319 16391 16519
DL175	Dubai	N1b1b	73 152 263 315.1C 750 1438 1598 1703 1719 2639 2706 3921A 4769 4904 4960 5237 5471 6272 7028 8251 8472 8836 8860 9335 10238 11362 11719 12501 12705 12822 14766 15326 16093 16145 16176G 16223 16390 16519

DL244	Dubai	W6	73 189 194 195 204 207 263 309.1C 315.1C 709 750 1243 1438 2706 3505 4093 4646 4769 5046 5460 6297 7028 7269 8251 8614 8705 8860 8994 11674 11719 11947 12414 12705 14766 15326 15884C 16223 16292 16325 16519
DL247	Dubai	N1c	73 189 195 204 207 210 263 315.1C 750 1438 1719 2706 4769 7028 8222 8860 10238 11025 11437 11719 11914 12501 12705 13637 13780 14766 15326 16093 16192 16201 16223 16519
DL249	Dubai	X2	73 153 195 225 263 315.1C 750 1438 1655dA 1719 2706 4769 6221 6371 7028 7244 8507 8860 9840A 11719 12705 13656 13966 14470 14766 15326 16183C 16189 16223 16278 16519
DL250	Dubai	X2	73 153 195 225 263 309.1C 315.1C 750 1438 1719 2706 4769 6221 6371 7028 8860 11290 11719 12705 13966 14470 14766 15052 15074 15326 15930 16183C 16189 16223 16248 16278 16519
DL58	Dubai	I5a	73 199 204 250 263 315.1C 573.1C 750 1438 1719 1793 2706 3615 3705 4529T 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519
DL60	Dubai	N1e'I	73 143 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 4529T 4769 7028 8251 8860 10238 10388 10398 11719 12361 12501 12705 13748 13780 14766 15043 15326 15924 16093 16223 16309 16311 16391 16519
DL63	Dubai	N1e'I	73 143 199 204 250 263 315.1C 573.1C 710 750 1438 1719 2706 3645 4529T 4769 7028 8251 8860 10238 10388 10398 10790 11719 12361 12501 12705 13748 13780 14766 15043 15326 15924 16223 16309 16311 16391 16519
ETH45	Ethiopia	N1a	73 199 204 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 2758 4769 5315 6893 7028 8860 8901 10238 10398 10586 11719 12501 12705 13780 14766 15043 15326 16147G 16172 16223 16248 16355 16519
ETH52	Ethiopia	N1a	73 151 199 203 204 263 309.2C 315.1C 573.1C 669 750 789A 1438 1719 2702 2706 2758 4769 5315 7028 8860 8901 10238 10398 11719 12501 12705 13681 13780 14766 15043 15100 15326 16147G 16172 16189 16223 16248 16355
ETH55	Ethiopia	N1a	73 150 152 235 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 2758 4769 5315 7028 8860 8901 10238 10398 11719 12501 12705 13681 13780 14634 14766 15043 15326 16147G 16172 16223 16248 16260 16355
HEL51	Greece	X2d	73 195 204 263 315.1C 709 750 1438 1719 2706 4527 4529 4769 5186 6221 6371 6791 7028 8503 8860 11719 11878 12705 13708 13966 14470 14766 15326 16093 16183C 16189 16223 16278 16519
HEL59	Greece	I3	73 151 152 199 204 207 239 250 263 309.1C 315.1C 499 521.2CA 573.1C 750 1438 1719 2706 4529T 4769 6323 7028 8251 8860 10034 10238 10398 11719 12501 12705 13780 14766 15043 15326 15565 15924 16086 16129 16223 16391 16519
J44	Morocco	X2b	73 150 152 153 195 225 235 263 309.1C 315.1C 750 1438 1719 2706 4769 6221 6371 6509 7028 8393 8860 10238 10398 11719 12705 13708 13966 14470 14766 15326 15927 16173 16183C 16189 16223 16278
J59	Morocco	N1b1	73 150 152 235 263 309.1C 315.1C 750 1438 1598 1703 1719 2639 2706 3921A 4769 4960 5471 7028 8251 8472 8836 8860 9335 10238 11362 11719 12501 12705 12822 14766 15326 16145 16176G 16223 16311 16390
JHA114	Yemen	N1a	73 199 204 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 4769 5315 7028 7382 8860 8901 10238 10398 11719 12501 12705 13780 14766 15043 15326 16147G 16172 16223 16248 16355 16399 16519
JHG230	Yemen	X2e2	73 146 152 153 195 263 309.1CC 315.1C 750 1438 1719 2706 3948 4769 6221 6371 7028 7775 8856 8860 11719 12084 12705 13966 14470 14766 15310 15326 16183C 16189 16223 16278 16325 16519
JSO153	Socotra	I5a	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 3615 3705 4529T 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519

JSO164	Socotra	I5a	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 3615 3705 4529T 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519
JSO180	Socotra	I5a	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 3615 3705 4529T 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14007 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519
JSO196	Socotra	I5a	73 199 204 250 263 315.1C 489 573.1C 750 1438 1719 2706 3615 3705 4529T 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519
JSO205	Socotra	I5a	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 3615 3705 4529T 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519
JSO211	Socotra	I5a	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 3615 3705 4529T 4688 4742 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519
JSO215	Socotra	I5a	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 3615 3705 4529T 4769 5074 5096 5773 7028 8251 8562 8742 8860 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 15941 16086 16129 16148 16223 16391 16519
JTI96	Yemen	W	73 189 194 195 204 207 263 309.1C 315.1C 553 709 750 1243 1438 2706 4769 5046 5147 5460 7028 8251 8697 8860 8994 9972 10724 11674 11719 11947 12414 12705 14766 15326 15884C 16223 16292 16519
KAB13	Karbardia	X2	73 153 195 225 263 315.1C 750 1438 1719 2403 2706 4769 6221 6371 7028 8860 8966 9716 11719 11928 12705 13966 14470 14766 15326 16189 16223 16248 16278
KAB16	Kabardia	N1a	73 152 199 204 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 4769 5315 7028 8452 8460 8860 8901 10238 10398 11719 12501 12705 13780 14766 15043 15326 16147G 16172 16223 16248 16295 16297 16355 16519
KAB36	Kabardia	W1c	73 189 195 204 207 263 309.1C 315.1C 709 750 1243 1438 1700 2706 3505 4769 5046 5460 7028 7864 8251 8860 8994 9804 11674 11719 11947 12414 12705 12984 14148 14766 15326 15884C 16223 16257 16292 16519
KUR82	Kurd (Turkey)	N1b1	73 152 263 309.1C 315.1C 522dCA 750 1438 1598 1703 1719 2639 2706 3921A 4769 4960 5471 7010 7028 7337 8251 8469 8472 8836 8860 9133 10238 11362 11719 12501 12705 12822 14690 14766 15326 16145 16176G 16223 16390
Kur84	Kurd (Turkey)	W3	73 189 194 195 199 204 207 263 315.1C 709 750 1243 1406 1413 1438 1693 2258 2706 3505 4769 5046 5460 7028 8251 8860 8950 8994 11467 11674 11719 11947 12063 12414 12705 12923T 14557 14634 14766 15326 15884C 16223 16292 16519
Kur91	Kurd (Turkey)	I1	73 199 204 250 263 291.1A 294.1T 309.1C 315.1C 455.1T 573.1C 750 1438 1719 2244 2706 4529T 4769 6734 7028 7705 8251 8860 9966 10034 10238 10398 11719 12501 12705 13780 14757 14766 15043 15326 15924 16129 16223 16311 16391 16519
KUR92	Kurd (Turkey)	W4	73 119 143 146 189 194 195 196 204 263 309.1C 315.1C 522dCA 709 750 1243 1438 2706 3505 4769 5046 5460 6977 7028 8251 8860 8994 11674 11719 11947 12414 12705 13011 13145 14766 15308 15326 15884C 16082 16223 16292 16519
KUR93	Kurd (Turkey)	W	73 189 194 195 204 207 263 309.1C 315.1C 522dCA 709 750 1243 1438 2706 3505 4769 5046 5128 5460 7028 8251 8860 8994 11674 11719 11947 12414 12705 14097 14766 15326 15884C 16129 16223 16292 16519

KUR94	Kurd (Turkey)	W3a1	73 189 194 195 204 207 263 309.1C 315.1C 709 750 1243 1406 1438 1709 2706 3505 4769 5046 5460 7028 8251 8860 8994 10845 11674 11719 11947 12414 12705 13263 14766 15326 15784 15884C 16223 16244 16292 16368 16519
NE1	Iraq	X2e	73 153 195 263 309.1C 315.1C 750 1438 1719 2706 4547 4769 6221 6371 7028 7853 8860 11719 11875 12084 12705 13966 14470 14766 15310 15326 16189 16223 16265 16278 16519
NE2	Iraq	N1b1	73 152 263 315.1C 750 1438 1598 1703 1719 2639 2706 3921A 4769 4960 5471 5987 7028 8251 8472 8836 8860 9335 9921 10238 11362 11719 12501 12705 12822 14766 15326 16093 16145 16176G 16223 16390 16519
NOS10	North Ossetia	W6	73 179 189 194 195 204 207 263 309.1C 315.1C 709 750 1243 1438 2706 3505 4093 4769 5046 5460 7028 8251 8614 8860 8994 11339 11674 11719 11947 12414 12705 14766 15326 15884C 16192 16223 16292 16325
NOS105	North Ossetia	X2f	73 153 195 257 263 315.1C 750 1438 1719 2392 2706 4769 6221 6371 7028 8860 10398 10849 11719 12705 13966 14470 14766 15326 16183C 16189 16223 16278 16311 16519
NOS126	North Ossetia	X2f	73 153 195 257 263 315.1C 750 1438 1719 2706 4769 6221 6371 7028 8860 11719 12705 13966 14470 14766 15326 16183C 16189 16223 16239 16278 16298 16318 16519
NOS2	North Ossetia	I4	73 199 204 250 263 309.1C 315.1C 573.1C 750 1438 1719 2706 4529T 4769 7028 7874 8251 8519 8860 10034 10238 10398 10819 11719 12501 12705 13780 14766 15043 15323 15326 15924 16129 16223 16294 16391
NOS40	North Ossetia	N1b1	73 152 263 315.1C 750 1438 1598 1703 1719 2263 2639 2706 3921A 4769 4960 5471 7028 8251 8472 8836 8860 9335 10238 10497 11362 11719 12501 12705 12822 14766 15326 16145 16176G 16223 16390
NOS71	North Ossetia	I5	73 199 204 250 263 309.1C 315.1C 573.1C 750 1438 1719 2706 4529T 4769 7028 8251 8860 9554 10034 10238 10398 11719 12501 12705 13780 14233 14766 15043 15326 15924 16129 16223 16391
NOS84	North Ossetia	W6	73 189 194 195 204 207 263 309.1C 315.1C 709 750 1243 1438 2171G 2706 3505 3606 4093 4769 5046 5460 7028 8251 8614 8860 8994 9139 11674 11719 11947 12411 12414 12705 14766 15326 15884C 16192 16223 16292 16304 16325 16519
PAL1024	Israel	W6	73 189 194 195 204 207 263 309.1C 315.1C 316 709 750 1243 1438 2706 3505 4093 4646 4769 5046 5460 6297 7028 8251 8614 8705 8860 8994 11674 11719 11947 12414 12705 14766 15326 15884C 16124 16192 16223 16325
PAL516	Israel	W6	73 152 185 189 194 195 204 207 263 309.1C 315.1C 709 750 1243 1438 2706 3505 4093 4769 5046 5460 7028 8251 8614 8860 8994 9211 11674 11719 11947 12414 12705 14766 15326 15884C 16192 16223 16292 16325
PAL706	Israel	X2b	73 153 195 200 225 226 251 263 309.1C 315.1C 750 1438 1719 2706 3316 4769 6221 6371 7028 8393 8860 11719 12705 13708 13966 14470 14766 15326 15927 16189 16223 16278 16316 16519
QL32	Tunisia	X3	73 146 153 256 263 309.1C 315.1C 750 1438 2706 3531 4769 6221 6371 7028 8860 11719 12705 13785 13879 13966 14470 14560 14766 15326 15672 16183C 16189 16223 16278 16519
R503	Tunisia	X2b	73 153 195 225 226 249dA 260 263 309.1C 315.1C 750 1438 1719 2706 4769 6221 6371 7028 8393 8860 10388 11719 12705 13708 13966 14470 14766 15148 15326 15927 16183C 16189 16223 16278 16519
R553	Tunisia	N1b1c	73 152 263 271 309.1C 315.1C 522dCA 750 1438 1598 1703 1719 2639 2706 3921A 4769 4960 5471 7028 8251 8472 8836 8860 9335 9957 10238 10909 11362 11719 12501 12705 12822 14766 15326 16145 16176G 16223 16390 16519
R574	Tunisia	I1a1	73 189 199 203 204 250 263 315.1C 455.1T 573.1C 750 1438 1719 2706 3447 3990 4529T 4769 6734 7028 8251 8616T 8860 9947 10034 10238 10398 10915 11719 12501 12705 13780 14766 15043 15326 15924 16129 16172 16223 16311 16391 16519

SICTO569	Sicily	N1b1	73 152 263 309.1C 315.1C 750 867 1438 1598 1703 1719 2639 2706 3308 3921A 4769 4960 5471 7028 8251 8472 8477 8836 8860 9335 10238 11362 11719 12501 12705 12822 13851 14560 14766 15326 16129 16145 16176G 16223 16249 16390
SK03	Tunisia	X2	73 153 195 263 309.1C 315.1C 750 1438 1719 1822 2706 4769 6221 6371 7028 8281dCCCCCTCTA 8860 8908 11719 12705 13966 14470 14766 15326 16183C 16189 16223 16248 16278
Som126	Somali	N1a	73 199 204 207 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 3535 4769 4924 5315 7028 8860 8901 9729 10238 10398 11719 12501 12630 12705 13780 14766 15043 15326 16147G 16172 16213 16223 16248 16355 16519
Som133	Somali	N1a	73 199 204 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 2758 3576 4769 5315 7028 8860 8901 10238 10398 11641 11719 11914 12501 12705 13928 14766 15043 15319 15326 16147G 16172 16223 16248 16355 16519
Som24	Somali	N1b Other	73 152 263 315.1C 522dCA 750 1438 1598 2639 2706 4769 5471 5507A 6905 7013 7028 8251 8836 8860 10238 11719 12026 12501 12705 13681 13780 13928 14766 15043 15326 15784 16176G 16223 16258C 16390 16519
Som61	Somali	N1a	73 179 199 204 207 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 3535 4769 4924 5315 7028 8860 8901 9729 10238 10398 11719 12501 12630 12705 13780 14766 15043 15326 16147G 16172 16213 16223 16248 16355 16519
Som66	Somali	N1b Other	73 152 263 522dCA 750 1438 1598 2639 2706 4769 5471 5507A 7028 8251 8836 8860 9325 10238 11719 12026 12501 12705 14766 15326 15784 16176G 16223 16258C 16390 16519
Som69	Somali	I	73 199 204 250 263 309.1C 315.1C 573.1C 750 1438 1719 2706 4529T 4769 7028 8251 8860 8882 10034 10238 10398 11719 12501 12705 13780 14766 15043 15326 15511 15924 16129 16223 16391 16519
Som8	Somali	N1a	73 199 204 207 263 309.1C 315.1C 573.1C 669 750 1438 1719 2702 2706 3535 4769 4924 5315 7028 8860 8901 9729 10238 10398 11719 12501 12630 12705 13780 14766 15043 15326 16147G 16172 16213 16223 16248 16355 16519
Som89	Somali	I	73 199 250 263 315.1C 573.1C 750 1438 1719 2706 3447 4529T 4769 7028 8251 8860 10034 10238 10398 11719 11929 12501 12705 13780 14766 15043 15326 15924 16129 16223 16391 16519
Som9	Somali	N1a	73 199 204 263 315.1C 573.1C 669 750 1438 1719 2702 2706 2758 4769 5315 7028 8860 8901 8954 10238 10398 10586 11719 12501 12705 13780 14766 15043 15326 16147G 16172 16223 16248 16355 16519
TURAT10	Turkey	X4	73 143 152 195 263 315.1C 466 750 1438 2706 4769 6221 6227 6371 7028 7278 8860 11719 12406 12705 13966 14470 14766 15326 16182C 16183C 16189 16223 16266 16274 16278 16390
TURAT13	Turkey	N1b	73 152 263 315.1C 750 1438 1598 1703 1719 2639 2706 3921A 4769 4960 5471 5480 7028 7142 8251 8472 8836 8860 9185G 10238 11719 12501 12705 12822 13350 14766 15326 16124 16145 16176G 16223 16256 16390 16400 16519
TURAT37	Turkey	I1a1	73 199 203 204 250 263 309.1C 315.1C 455.1T 573.1C 750 1438 1719 2706 3447 3990 4529T 4769 6734 7028 8251 8616T 8860 9053 9254 9947 10034 10238 10398 10915 11719 12501 12705 13780 14766 15043 15326 15924 16129 16172 16223 16311 16391 16519
TURAT4	Turkey	W1c	73 119 152 189 195 204 207 263 309.1C 315.1C 709 750 1243 1438 2706 3505 4769 5046 5460 7028 7864 8251 8860 8994 11674 11719 11947 12414 12705 14148 14766 15326 15884C 16193 16223 16246 16292 16519
TURAT50	Turkey	W6	73 189 194 195 204 207 211 263 309.1C 315.1C 709 750 1243 1438 2706 3505 4093 4769 5046 5460 7028 8271 8557 8614 8860 8994 11674 11719 11947 12414 12705 13638 14766 15326 15884C 16192 16223 16292 16325 16519
TURAT8	Turkey	X2	73 146 153 195 263 315.1C 750 1438 1655dA 1719 2706 4769 5147 6221 6371 6713 7028 8860 9840A 11719 12705 13035A 13656 13966 14470 14766 15326 16183C 16189 16223 16278 16519

TURAT82	Turkey	I2	73 152 199 204 207 250 263 315.1C 573.1C 750 1438 1719 2706 4529T 4769 7028 8251 8860 10034 10238 10398 11353 11719 12501 12705 13780 14766 15043 15326 15758 15924 16129 16172 16223 16391 16519
TURAT91	Turkey	I5a	73 150 152 199 204 250 263 309.1C 315.1C 573.1C 750 1438 1719 2706 4529T 4769 5074 5231 6278 7028 8251 8860 9190 10034 10238 10398 11719 12073 12501 12705 13269 13780 14233 14766 15043 15052 15326 15924 16129 16148 16223 16301 16311 16391 16519
TURAT96	Turkey	I	73 199 204 250 263 315.1C 573.1C 750 1438 1719 2706 3645 4529T 4769 7028 8251 8860 10034 10238 10398 10955 11719 12501 12705 13780 14766 15043 15326 15924 16129 16185 16186 16188 16223 16327 16391 16519
TURATB12	Turkey	N1b1	73 152 263 315.1C 750 1438 1598 1703 1719 2639 2706 3921A 4769 4960 5471 7028 8251 8472 8836 8860 9335 10238 10373 11362 11719 12501 12705 12822 13419 14766 15317 15326 16126 16145 16176G 16209 16223 16390 16519
TURATC4	Turkey	X2e	73 153 195 200 263 315.1C 750 1438 1719 2706 4769 6221 6371 7028 8347 8860 9245 10338 11719 12084 12705 13440 13966 14470 14766 15310 15326 15721 15942 16182C 16183C 16189 16223 16265 16278 16311 16519
YBA055	Yemen	X2e2	73 152 153 195 263 315.1C 750 1438 1719 2706 3948 4769 6221 6371 7028 8856 8860 11719 12084 12705 13966 14470 14766 15310 15326 16183C 16189 16223 16278 16519
Budu093	Niger	L4b2b	73 146 150 152 195 214 244 263 315.1C 513 709 750 769 1018 1438 2706 3918 4769 5128 6260 7028 7805 8104 8227 8701 8860 8966 9540 9855 9855 10265 10389 10398 10873 11719 11914 12354 12438 12609 12705 12609 12705 12903 13470 14766 15217 15301 15326 16093G 16223 16287A 16293T 16301 16311 16355 16362 16399
DL147	Dubai	L4a1a	73 150 189 195 198 263 315.1C 325 750 769 1018 1438 2706 3357 4769 5460 6167 7028 7376 7762 7775 8473 8631 8697 8701 8860 9540 10373 10398 10873 11253 11344 11485 11653 11719 12280 12414 12705 13174 13344 14000A 14302 14766 15301 15326 16207T 16223 16260 16261 16311 16362 16519
DL148	Dubai	L4b2a2	73 146 244 263 291.1A 315.1C 750 769 1018 1413 1694 2706 3918 4769 4949 5048 5910 6260 6680 7028 8104 8292 8701 8860 9540 9855 10398 10783C 10873 11719 12609 12705 13105 13470 14766 15019 15301 15326 15448 16172 16223 16293T 16311 16319 16355 16362 16399
ETH06	Ethiopia	L4b2a2	73 146 152 244 263 391 315.1C 750 769 1018 1413 1438 1694 2706 3918 4769 4949 6260 7028 8104 8478 8701 8860 9540 9855 10398 10724 10783C 10873 11719 11854 12609 12705 13443 14766 15301 15326 15454 16172 16223 16293T 16311 16355 16362 16399 16519
ETH25	Ethiopia	L6a	73 146 152 182 185C 263 265 309.1C 315.1C 709 750 769 770 961 1018 1438 1461 2706 3594 3776 4769 4964 5267 6002 6284 7028 7256 8656 8701 8715 8860 8874 9332 9540 10142 10398 10873 10978 11116 11719 11743 12705 12771 13650 13710 14766 14791 14959 15244 15262 15289 15301 15326 15499 16048 16223 16224 16278 16294 16311 16519
ETH38	Ethiopia	L4a1a	73 150 189 195 198 210 263 309.1C 315.1C 325 750 769 1018 1438 2706 3357 4297 4769 5460 6167 7028 7376 7762 7775 8473 8631 8697 8701 8860 9540 10373 10398 10873 11253 11344 11485 11653 11719 12414 12705 13174 13269 13174 13269 13344 13506 14000A 14302 14766 15301 15326 16148 16207T 16223 16260 16261 16311 16362 16519
ETH41	Ethiopia	L4a1a	73 195 198 263 315.1 325 750 769 1018 1193 1438 2706 3357 3796 4769 4967 5460 6167 7028 7376 7762 7775 8473 8701 8860 9540 10373 10398 10873 11253 11344 11485 11653 11719 12414 12705 13174 14000A 14302 14766 15301 15326 16093 16207T 16223 16260 16291 16311 16362 16519
ETH59	Ethiopia	L6b	73 146 182 185C 263 309.1C 315.1C 709 750 769 770 961 1018 1438 1461 3594 4769 4964 5267 6002 6284 7028 7256 8701 8860 9332 9540 9545 10398 10873 10978 11116 11719 11743 12405 12705 12714 12771 13650 13710 14533A 14766 14791 14959 15244 15289 15301 15326 15499 15928 16048 16223 16224 16274 16278 16311 16519

ETH73	Ethiopia	L4a1a	73 195 198 228 263 315.1C 325 750 769 1018 1438 2706 3357 3700 4769 5460 6167 6527 7028 7376 7762 7775 8473 8631 8701 8860 9540 10373 10398 10873 10978 11149 11253 11344 11485 11653 11719 12414 12705 12814T 13174 14000A 14302 14766 15301 15326 16207T 16219 16223 16260 16311 16320 16519
GUR20	Burkina Faso	L4b1	73 150 199 204 207 263 309.1C 315.1C 513 709 750 769 1018 1438 1804 2706 3010 3505 3918 4017 4029A 4216 4232 4769 4977 5460 7028 7624 8588 8614 8701 8860 8974G 9248 9540 9986 10398 10813 10873 11719 12661T 12705 13497 14016 14766 14905 14935 15043 15301 15326 15344 16179 16189C 16223 16239 16311 16320 16362 16519
Kane101	Chad	L4b2b	73 146 152 195 198 244 263 315.1C 514dGC 709 750 769 1018 1438 2706 3918 4769 5128 6260 7028 7805 8104 8227 8701 8860 9540 9855 10265 10398 10873 11719 11914 12354 12438 12609 12705 12903 13470 14766 15217 15301 15326 16093A 16223 16287A 16293T 16301 16311 16355 16356 16362 16399 16519
KANE53	Chad	L4b2b	73 146 152 195 198 244 263 315.1C 514dGC 709 750 769 1018 1438 2706 3918 4769 5128 6260 7028 7805 8104 8227 8701 8860 9540 9855 10265 10398 10873 11719 11914 12354 12438 12609 12705 12903 13470 14766 15217 15301 15326 16093A 16223 16287A 16293T 16301 16311 16355 16356 16362 16399 16519
Kanu040	Nigeria	L4b2b	73 146 152 195 244 263 291.1A 315.1C 340 523dCA 709 750 769 1018 1438 1709 2220T 2706 3918 4206 4769 4973G 5128 5134 5156 6260 7028 7805 8104 8701 8860 9540 9855 10265 10398 10873 11719 12438 12609 12705 13470 14766 15301 15326 15759 16188 16189 16209 16223 16274 16292 16293T 16311 16316 16335 16355 16362 16399 16519
NUB087	Nubia	L4b2a2a	73 146 257 263 315.1C 750 769 1018 1413 1438 1694 2483 2706 3918 4769 4949 5824 6260 6620 7028 8104 8701 8860 9540 9855 10398 10783C 10873 11137 11719 12609 12705 13470 13708 14422 14766 15301 15326 16147 16223 16293T 16311 16355 16362 16399 16519
ORO25	Kenya	L6b	73 146 153 182 184 185C 263 315.1C 709 750 769 770 961 1018 1438 1461 3594 4769 4964 5267 6002 6284 7028 7256 8701 8860 9332 9540 10310 10398 10873 10978 11116 11719 11743 12405 12705 12714 12771 13650 13650 13710 14533A 14766 14791 14861 14959 15244 15289 15301 15326 15499 15928 16048 16093 16223 16224 16274 16278 16311 16519
ORO36	Kenya	L4b2a2b	73 146 244 263 309.1C 315.1C 750 769 1018 1413 1438 1694 1842 2706 3918 4769 4949 5573 6260 6956 7028 8104 8485 8701 8860 9540 9855 10398 10535 10783C 10873 11719 12609 12705 13470 14572 14766 15301 15314 15326 15970 16172 16223 16287 16293T 16311 16355 16362 16399
RSK_01		L4b2a2a	73 146 152 195 244 263 257 291.1A 315.1C 340 750 769 1018 1413 1438 1694 2483 2706 3213 3918 4639 4769 4949 6260 6620 7028 8104 8128 8701 8860 9540 9855 10398 10783C 10873 11137 11719 12609 12705 13029 13470 13708 14422 14766 15301 15326 16223 16293T 16311 16355 16362 16399
RSK_05		L4b2a2	73 146 244 263 309.1C 315.1C 398 513dGC 750 769 1018 1413 1438 1694 2706 3918 4769 4949 6260 7028 7664 7987 8104 8701 8860 9540 9855 10398 10646C 10783C 10873 11151 11719 12609 12705 13470 13602 14766 15301 15326 15924 16111 16223 16293T 16311 16355 16362 16399
Som128	Somali	L6a	73 146 152 182 185C 195 207 263 265 315.1C 709 750 769 770 961 1018 1438 1461 1943 2706 3594 4769 4964 5267 6002 6284 7028 7256 8701 8814 8860 9332 9540 10398 10873 10978 11116 11719 11743 12705 12771 13650 13710 14693 14766 14791 14959 15244 15289 15301 15326 15479 15499 16048 16173 16184 16223 16224 16278 16311 16362 16399 16519
Som21	Somali	L4b2a2	73 146 244 263 315.1C 391 750 769 1018 1413 1438 1694 2706 3394 3918 4769 4949 5483G 6260 7028 8104 8478 8701 8860 9524 9540 9855 10398 10589 10783C 10873 11719 12609 12705 13470 13533T 14766 15301 15326 15454 16172 16223 16293T 16311 16355 16362 16399 16519

Som37	Somali	L4a2	73 195 198 263 304 309.1C 523dCA 750 769 1018 1438 2706 3357 4769 5460 7028 8701 8860 9103 9540 10032 10373 10398 10873 11253 11344 11485 11719 12361 12414 12705 13174 13392 14302 14766 15301 15326 16179 16189 16223 16260 16264 16311 16362 16519
Som46	Somali	L4b2a1	73 146 152 195 244 263 315.1C 471 547 750 769 1018 1413 1438 2706 3918 4769 5471 5580 5746 7028 8104 8701 8860 9540 9855 10398 10873 11719 12295 12609 12705 13470 14766 15301 15326 16223 16274 16293T 16311 16355 16362 16399 16519
Som5	Somali	L4b2a	73 146 244 263 315.1C 455.1T 513 750 769 1018 1413 1438 2706 2863 3333 3918 4769 6260 7028 7129 8104 8701 8843 8860 9123 9287 9540 9855 10398 10873 11719 12173 12609 12681 12705 13470 14569 14766 15217 15301 15326 16301 16223 16293T 16311 16355 16362 16399 16519
Som54	Somali	L4a1a	73 150 189 195 198 263 315.1C 325 750 769 1018 1438 2706 3357 4769 5460 6167 7028 7376 7762 7775 8473 8631 8697 8701 8860 9540 10373 10398 10873 11253 11344 11485 11653 11719 12414 12705 13174 13344 14000A 14302 14766 15301 15326 15497 16207T 16223 16260 16261 16311 16362 16519
Som87	Somali	L6a	73 146 152 182 185C 228 263 265 315.1C 709 750 769 770 961 1018 1438 1461 2706 3335 3594 4769 4964 5267 6002 6284 7028 7145 7256 8701 8860G 9332 9540 10398 10873 10978 11116 11719 11743 12705 12771 13650 13710 14766 14791 14959 15244 15289 15301 15326 15499 16048 16129 16223 16224 16278 16311 16519
Tur_02	Kenya	L4b2a	73 146 244 263 315.1C 455.1T 513 750 769 1018 1413 1438 2706 2863 3333 3918 4769 6260 7028 7129 8104 8701 8843 8860 9123 9287 9540 9855 10398 10873 11719 12173 12609 12681 12705 13470 14766 15217 15301 15326 16223 16293T 16301 16311 16355 16362 16399 16519
JAC-1	Yemen	J1b1b1	73 263 271 295 309.1C 315.1C 462 489 523d 750 1438 2706 3010 4216 4769 5460 7028 8860 8269 10398 10410A 11251 11719 12246 12612 13708 13879 14766 15326 15452A 16069 16126 16145 16261 16399 16519
JAC-4	Yemen	J1b3	73 146 263 295 309.1C 315.1C 462 489 750 1438 2706 3010 4216 4769 7028 8231 8269 8460 8642 8659 8860 10398 11251 11710 11719 12612 13708 13782 14766 15326 15452A 16069 16126 16145 16222 16261 16362
JAC-7	Yemen	J1d1a1	73 152 263 295 315.1C 370 462 489 750 1007 1438 2706 3010 4216 4769 5147 6546 7028 7789 7963 8860 9380 10398 11251 11719 12612 13392 13708 14766 15326 15452A 16069 16086 16126 16193 16309
JAC-9	Yemen	T2a1a	73 263 309.1C 315.1C 709 750 1438 1719 1888 2706 2850 4216 4721 4769 4917 7022 7028 8251 8697 8860 10463 11251 11719 11812 13359 13368 13965 14233 14687 14766 14905 15326 15452A 15607 15928 16093 16126 16183C 16189 16294 16296 16519
JAC-12	Yemen	T2a1a	73 263 309.2C 709 750 1438 1719 1888 2706 2850 4216 4721 4769 4917 7022 7028 8251 8697 8860 9055 10463 11251 11719 11812 13359 13368 13965 14233 14687 14766 14905 15326 15452A 15607 15928 16126 16294 16296 16519
JAC-30	Yemen	T2g1a1a	73 200 263 315.1C 709 750 789 1438 1888 1977 2706 3834 4216 4769 4917 7028 8697 8860 10463 10576 11251 11719 11812 13368 14233 14524 14766 14798 14839 14905 15326 15452A 15607 15928 16114A 16126 16294 16519
JTI-57	Yemen	T2c1b2	73 146 152 263 309.1C 315.1C 523d 709 750 1438 1888 2380 2706 3221 4216 4769 4917 6261 6521 7028 8697 8860 10289 10398 10463 10822 11251 11719 11812 13368 14233 14766 14905 15326 15452A 15607 15928 16126 16288 16292 16294 16296 16311
JTI-74	Yemen	J1d1a1	73 152 263 295 315.1C 462 489 750 1007 1438 2706 3010 4216 4703 4769 7028 7202 7789 7963 8131 8860 9380 10398 11251 11719 12612 13392 13708 14766 15326 15452A 16069 16126 16193 16309

JTI-97	Yemen	J1d1a	73 152 263 295 315.1C 462 489 514d 516 750 1007 1438 2706 3010 4216 4769 7028 7789 7963 8622 8860 10398 11251 11719 12612 13708 14766 14996 15326 15452A 16069 16126 16193 16287 16300 16309
JHA-134	Yemen	T2e3	73 146 150 195 263 309.1C 315.1C 709 750 1438 1888 2706 4216 4769 4917 7028 8697 8860 9698 10463 11251 11719 11812 12477 13020 13368 13962 14233 14766 14905 15326 15452A 15607 15928 16126 16153 16233C 16257 16294 16325 16399 16519
JSO-152	Yemen	J2a2a1	73 150 195 235 263 295 309.1C 315.1C 489 750 1438 1888 2706 3316 4216 4769 6671 7028 7476 8386 8860 10398 10499 11251 11377 11719 12171 12570 12612 13708 14766 15257 15326 15452A 15679 16069 16126 16179
JSO-155	Yemen	T2i	73 263 315.1C 709 750 1438 1888 2706 4216 4769 4917 7028 8155 8697 8860 9422C 10463 11251 11719 11812 13368 14233 14766 14905 15397 15452A 15607 15928 16126 16294 16296 16362 16519
JSO-160	Yemen	T2c1c	73 263 315.1C 709 750 1438 1888 2706 3010 4216 4769 4823 4917 6261 7028 8697 8860 9177 10463 10822 11251 11719 11812 13368 14233 14319 14766 14905 15326 15452A 15607 15928 16126 16146 16183C 16189 16294 16296 16519
JSO-208	Yemen	J2a2	73 150 195 263 295 315.1C 489 709 750 1438 2706 4216 4769 6671 6908 7028 7476 7830 8860 9189 10398 10499 11002 11251 11377 11719 12570 12612 13708 14577 14766 15257 15326 15452A 15679 16069 16126
JHG-222	Yemen	J2a2a1a	73 150 195 235 263 295 309.2C 315.1C 489 750 1438 2706 3316 4216 4769 6671 7028 7476 7585 8386 8860 10398 10499 11002 11251 11377 11719 12171 12570 12612 13708 14766 15257 15326 15452A 15679 16069 16126 16179
JHG-234	Yemen	J1d1a1	73 152 263 295 315.1C 462 489 750 1007 1438 2706 3010 4216 4769 7028 7789 7963 8269 8860 10398 11251 11719 12612 13392 13420 13708 14766 15326 15452A 16069 16126 16145 16193 16300 16309
JHG-237	Yemen	J1b2a	73 263 295 315.1C 462 489 750 1438 1733 2706 3010 4216 4769 7028 8269 8860 8962 10398 11251 11719 12612 13708 14766 15326 15452A 15466 16069 16126 16136 16145 16218 16222 16261
JHG-245	Yemen	J2a2b1	73 150 185 189 195 263 295 309.1C 315.1C 489 522 750 1438 2706 4216 4769 6671 7028 7476 8712 8860 10398 10499 11002 11251 11377 11719 12570 12612 13708 14766 15257 15326 15452A 15679 16069 16126 16241
YBA044	Yemen	T1a	73 114A 263 315.1C 709 750 1438 1888 2706 4216 4769 4917 5123 6060 7028 8697 8860 10253 10463 11251 11719 12633A 13368 14766 14905 15326 15452A 15607 15928 16093 16126 16163 16186 16189 16294 16300 16519
YBA049	Yemen	T1a	73 114A 263 315.1C 709 750 1438 1888 2706 4216 4769 4917 5123 6060 7028 8697 8860 10253 10463 11251 11719 12633A 13368 14766 14905 15326 15452A 15607 15928 16126 16163 16185 16186 16189 16294 16300 16519
YBA050	Yemen	J2a2b	73 150 195 204 263 295 315.1C 489 523iAC 750 1438 1676 2706 3337 4216 4769 5054 5291 6671 7028 7476 7741 8860 9323 10398 10499 11002 11251 11377 11719 12570 12612 12858 13708 14766 14947 15257 15326 15452A 15679 16069 16126 16145 16241
YBA061	Yemen	T2c1b2	73 146 152 263 309.1C 315.1C 523dAC 709 750 1438 1888 2380 2706 3221 4216 4769 4917 6261 6521 7028 8697 8860 10289 10463 10822 11251 11719 11812 13368 14233 14766 14905 15326 15452A 15607 15928 16126 16288 16292 16294 16296 16311
YBA063	Yemen	J1d1a1	73 152 263 295 309.2C 315.1C 462 489 750 1007 1438 2706 3010 4216 4769 7028 7789 7963 8860 10398 11251 11719 12612 13392 13708 13967 14766 15326 15452A 16069 16126 16193 16256 16300 16309
YBA068	Yemen	J1d1a1	73 152 263 295 315.1C 370 462 489 513 750 1007 1438 2706 3010 4216 4769 7028 7789 7963 8152 8860 9380 10398 11251 11719 12612 13392 13708 14766 15326 15452A 16069 16126 16193 16300 16309

YBA072	Yemen	J1d1a1	73 152 263 295 315.1C 370 462 489 513 750 1007 1438 2706 3010 4216 4769 7028 7789 7963 8152 8860 9380 10398 11251 11719 12612 13392 13708 14766 15326 15452A 16069 16126 16193 16300 16309
DL70	Dubai	J1d2c2	73 152 263 295 315.1C 462 489 750 1438 2706 3010 4011 4216 4769 5262 7028 7521 7789 7963 8860 10398 11251 11719 12612 13708 14766 15326 15452A 16069 16126 16193 16519
DL90	Dubai	J1b1b	73 263 271 295 315.1C 462 489 750 1438 2706 3010 4216 4769 5460 7028 8269 8494 8860 10398 11251 11719 12612 13656 13708 13879 14766 15326 15452A 15941 16069 16126 16145 16222 16261 16519
DL68	Dubai	J2a2b	73 150 195 200 263 295 315.1C 489 750 1438 2483 2706 4216 4257 4769 6671 6749 7028 7476 8860 10398 10499 11002 11251 11377 11719 12570 12612 13708 14058 14766 15217 15257 15326 15452A 15679 16069 16126 16241 16278
DL72	Dubai	J1d2c2	73 152 263 295 315.1C 462 489 750 1438 2281C 2706 3010 4011 4216 4769 5262 7028 7521 7789 7963 8860 10398 11251 11719 12612 13708 14766 15326 15452A 16069 16126 16193 16519
DL73	Dubai	J1d2	73 152 263 295 309.1C 315.1C 462 489 750 1438 2706 3010 3523 4216 4769 7028 7789 7963 8473 8860 10398 11251 11719 12612 13708 14766 15326 15452A 16069 16114 16126 16193 16519
DL76	Dubai	J1b	73 151 152 263 295 315.1C 462 489 750 1438 2706 3010 4216 4354 4769 6962 7028 7080 7364 8110 8269 8860 10398 10873 11251 11719 12612 12757 13708 13933 14020 14766 15326 15452A 16069 16145 16189 16261
DL77	Dubai	J1b	73 151 152 263 295 309.1C 462 489 523d 750 1438 1555 2706 3010 4216 4354 4769 6962 7028 7364 8269 8860 8994 10398 10873 11251 11719 12612 13708 13933 14020 14353 14766 15326 15452A 15663 16069 16124 16145 16224 16261 16519
DL79	Dubai	J1b	73 151 152 263 295 309.1C 315.1C 462 489 750 1438 2706 3010 4216 4354 4769 6962 7028 7364 8269 8860 10398 10873 11251 11719 12612 13488 13708 13933 14020 14766 15326 15452A 16069 16145 16261
DL82	Dubai	J1b2	73 183 263 295 315.1C 462 489 709 750 1438 1733 2706 3010 4216 4769 6340 7028 8269 8860 10398 11204 11251 11719 12406 12612 13708 14766 15326 15452A 16069 16093 16126 16145 16222 16261 16300
DL80	Dubai	J1b	73 151 152 263 295 309.1C 315.1C 462 489 750 1438 2706 3010 4216 4354 4769 6962 7028 7364 8269 8860 10398 10873 11251 11719 12172 12612 13708 13933 14020 14766 15326 15452A 15773 16069 16145 16261
DL95	Dubai	J1b1a1	73 146 242 263 295 315.1C 462 489 523d 750 1438 2158 2706 3010 3840 4216 4769 5460 7028 8269 8557 8835 8860 10322 10398 11251 11719 11935 12007 12612 13708 13879 14766 15326 15452A 16069 16145 16172 16261
DL200	Dubai	T1a6	73 263 315.1C 523d 709 750 1438 1888 2706 3867 4216 4769 4917 7028 8697 8860 10376 10463 11251 11719 12633A 13368 14766 14905 15326 15452A 15607 15928 16126 16163 16186 16189 16294 16519
DL67	Dubai	J2a2b	73 150 195 200 263 295 315.1C 489 750 1438 2483 2706 4216 4257 4769 6671 6749 7028 7476 8860 10398 10499 11002 11251 11377 11719 12570 12612 13708 14058 14766 15217 15257 15326 15452A 15679 15924 16069 16126 16241 16278
DL93	Dubai	J1b2	73 152 263 295 309.1C 315.1C 462 489 750 1438 1733 2706 3010 4216 4769 7028 8269 8860 10398 11251 11719 12612 13708 14766 15326 15452A 16069 16126 16145 16222 16261 16519
DL201	Dubai	T1a1a1	73 152 195 207 263 309.1C 315.1C 709 750 1438 1888 2706 4065 4216 4769 4917 7028 8697 8860 9899 10463 11251 11719 12633A 13368 14766 14905 15326 15452A 15607 15928 16126 16163 16186 16189 16294 16519
DL204	Dubai	T2c1b2	73 146 263 309.1C 315.1C 523d 709 750 1438 1888 2706 4216 4769 4917 6261 7028 8697 8860 10289 10463 10822 11251 11719 11812 13368 14233 14766 14905 15326 15452A 15607 15928 16126 16189 16269 16292 16294 16296 16380 16519

DL205	Dubai	T2b	73 263 309.1C 315.1C 709 750 930 1438 1888 2706 3394 4216 4769 4890 4917 5147 6299 7028 8697 8860 10463 11251 11719 11812 13368 14233 14766 14905 15326 15452A 15607 15928 16126 16239 16256 16289 16294 16296 16304 16519
DL75	Dubai	J1d	73 263 271 295 309.1C 315.1C 462 489 523d 462 489 750 1438 2706 3010 3474 3483 4216 4769 7028 7789 7963 8860 10398 10410A 11251 11719 12612 13708 14766 15326 15452A 16069 16114 16126 16193
DL198	Dubai	T1a3	73 151 152 263 315.1C 709 750 1192 1438 1888 2706 4216 4769 4917 6152 7028 8697 8860 10463 10867 11251 11719 12633A 13368 14766 14905 15326 15412 15452A 15607 15928 16126 16163 16186 16189 16294

Table S 2 Published mitochondrial complete sequences used in all phylogenetic tree studied with the corresponding origin and subhaplogroup affiliation

Accession Number/ID	Geographic region	Country/ ethnicity	Haplogroup	Ref.	Accession Number/ID	Geographic region	Country/ ethnicity	Haplogroup	Ref.
EF556183	Anatolia	Turkish Jew	X2	1	FJ147321	Europe	Russia	X2c1	18
GU290214	Anatolia	Kazakhstan	N1a1a1a	2	GU122989	Europe	Russia	W3a1	19
FJ234984	Anatolia	Turkey	I1b	Family Tree DNA	GU123002	Europe	Russia	W1	19
HM454265	Anatolia	Turkey	I1a	Family Tree DNA	GU123029	Europe	Russia	W1b1	19
HM852869	Anatolia	Turkey	I5a	3	EF486517	Europe	Russia	N1a	14
EF556165	North Africa	Libyan Jew	X2e1	1	EF486518	Europe	Russia	N1a1a1	14
EF556175	North Africa	Libyan Jew	X2b1	1	GU290215	Europe	Russia	N1a1a1	2
EU935450	North Africa	Egypt	X2j	4	GU290216	Europe	Russia	N1a1a3	2
EU935453	North Africa	Egypt	X2j	4	FJ493516	Europe	Russia	N1b1b	18
EU935454	North Africa	Egypt	X2j	4	GU122992	Europe	Russia	N1b1	19
EU935456	North Africa	Egypt	X1a	4	GU122984	Europe	Russia	I2	19
EU935459	North Africa	Egypt	X1a	4	GU123027	Europe	Russia	I1a1	19
AF381986	North Africa	Morocco	X2b2	5	GU123019	Europe	Russia	N1c	19
FJ460521	North Africa	Tunisia	X1c	6	EU787451	Europe	Russia	N2a	21
FJ460557	North Africa	Tunisia	X2	6	EF153785	Europe	Russia	N1e	14
FJ460561	North Africa	Tunisia	N1b1	6	GU361771	Europe	Ukrainian Jew	N1b2	Family Tree DNA
FJ460562	North Africa	Tunisia	I1a1	6	HQ423397	Europe	Belarus	W6	Family Tree DNA
ReidlaOm1538	Arabian Peninsula	Oman	X2b2	7	FJ441666	Europe	France	X2b	Family Tree DNA
HQ456226	South Caucasus	Armenia	X4	Family Tree DNA	HM134923	Europe	France	X2b4	Family Tree DNA
HQ529295	South Caucasus	Armenia	X2	Family Tree DNA	GU290212	Europe	France	N1a1a2	2
HM852756	South Caucasus	Armenia	X2f	3	EU746658	Europe	France	I3	Family Tree DNA
HM852758	South Caucasus	Armenia	X2e1	3	HQ420832	Europe	France	I3	Family Tree DNA
EU515252	South Caucasus	Armenia	W6	Family Tree DNA	EU677423	Europe	Belgium	X2c1	Family Tree DNA
HQ844617	South Caucasus	Armenia	W	Family Tree DNA	AY245555	Europe	The Netherlands	I	22

JF286634	South Caucasus	Armenia	W6	Family Tree DNA	GQ231312	Europe	Hungary	X2d	Family Tree DNA
HQ286324	South Caucasus	Armenia	N1b1	Family Tree DNA	FJ461348	Europe	Hungary	W3	Family Tree DNA
HQ315687	South Caucasus	Armenia	N1b1b	Family Tree DNA	AY195773	Europe	Unknown	X2c1	23
HQ435319	South Caucasus	Armenia	N1b1b	Family Tree DNA	AY195768	Europe	Unknown	W3a1	23
JF265069	South Caucasus	Armenia	N1b1	Family Tree DNA	AY195779	Europe	Unknown	W1a	23
HM852768	South Caucasus	Armenia	N1b1b	3	AY195756	Europe	Unknown	N1b1c	23
JF904935	South Caucasus	Armenia	N2	Family Tree DNA	AY195769	Europe	Unknown	I1b	23
HM852814	South Caucasus	Azerbaijan	X2e1	3	HM852817	Near East	Iran	I5	3
HM852799	South Caucasus	Azerbaijan	X2	3	HM852831	Near East	Iran	I	3
HM852902	South Caucasus	Georgia	X2	3	HM852839	Near East	Iran	I1a	3
HM852896	South Caucasus	Georgia	X2d	3	EU600318	Near East	Druze, Israel	X1a	24
HM852901	South Caucasus	Georgia	N1b1b	3	EU600319	Near East	Druze, Israel	X1a	24
HM852893	South Caucasus	Georgia	N1c	3	EU600320	Near East	Druze, Israel	X3a	24
ReidlaGm66	South Caucasus	Georgia	X2e	7	EU600321	Near East	Druze, Israel	X2b4a	24
ReidlaGo41	South Caucasus	Georgia	X2f	7	EU600322	Near East	Druze, Israel	X3a	24
HM625691	Europe	Austria	X2d	8	EU600323	Near East	Druze, Israel	X2f	24
FJ384431	Europe	Austria	W1f	9	EU600324	Near East	Druze, Israel	X2e2	24
FJ384432	Europe	Austria	W1f	9	EU600325	Near East	Druze, Israel	X2h	24
HM625692	Europe	Austria	W1e	8	EU600326	Near East	Druze, Israel	X2h	24
HM347343	Europe	England, UK	X2e2	Family Tree DNA	EU600327	Near East	Druze, Israel	X2h	24
EF652811	Europe	England, UK	W5a1	Family Tree DNA	EU600328	Near East	Druze, Israel	X2	24
EU558696	Europe	England, UK	W1	Family Tree DNA	ReidlaDr09	Near East	Druze, Israel	X3a	7
FJ821289	Europe	England, UK	W3a1a	Family Tree DNA	EU742148	Near East	Palestinian	N1b1c	25
HM034304	Europe	England, UK	W4	Family Tree DNA	EU742149	Near East	Palestinian	N1b1d	25
HM057816	Europe	England, UK	W5a1	Family Tree DNA	EU742150	Near East	Palestinian	N1b1	25
EU869314	Europe	England, UK	I3	Family Tree DNA	EU742151	Near East	Palestinian	N1b1a	25
FJ911909	Europe	England, UK	I2	Family Tree DNA	EU742153	Near East	Palestinian	N1b1b	25
GU646872	Europe	Ireland	W5a1	Family Tree DNA	EU597573	Levant	Bedouin, Israel	I5a1	27
HQ724528	Europe	Ireland	I2a1	Family Tree DNA	AF381999	Levant	Jordan	N1b1	5
EU570217	Europe	Ireland	I2	Family Tree DNA	EF556154	Mesopotamia	Iraqi Jew	W1d	1
HQ714959	Europe	Scotland, UK	I2a	Family Tree DNA	DQ301794		Ashkenazi Jew	N1b2	26
DQ523631	Europe	Sardinia	X2c	15	DQ301811		Ashkenazi Jew	N1b2	26
DQ523637	Europe	Sardinia	X2c	15	EU742154		Ashkenazi Jew	N1b2	25
DQ523642	Europe	Sardinia	X2b3	15	EU742155		Ashkenazi Jew	N1b2	25

FJ966243	Europe	Italy	X2	Family Tree DNA	EU742156		Ashkenazi Jew	N1b2	25
EF660942	Europe	Italy	X2	10	EU742157		Ashkenazi Jew	N1b2	25
HM765467	Europe	Italy	X1c	11	EU742158		Ashkenazi Jew	N1b2	25
EF660966	Europe	Italy	W4	10	EU742159		Ashkenazi Jew	N1b2	25
EF660944	Europe	Italy	N1a1a1	10	EU742160		Ashkenazi Jew	N1b2	25
HM236190	Europe	Italy	N1b1d	Family Tree DNA	EU742161		Ashkenazi Jew	N1b2	25
EF660993	Europe	Italy	N1b1d	10	EU742162		Ashkenazi Jew	N1b2	25
EF661011	Europe	Italy	N1b1b	10	EU742163		Ashkenazi Jew	N1b2	25
HM765456	Europe	Italy	N1b1c	11	GU320211		Ashkenazi Jew	N1b2	Family Tree DNA
AY963586	Europe	Italy	I3	12	EF420985		Unknown	X2b	Family Tree DNA
EF660917	Europe	Italy	I5a1	10	EU367995		Unknown	X2b	Family Tree DNA
EF660987	Europe	Italy	I4	10	EU086510		Unknown	W	Family Tree DNA
EF660937	Europe	Italy	N1c	10	EU135972		Unknown	W5a1	Family Tree DNA
EF177418	Europe	Portugal	X2b4	13	EU257638		Unknown	W1c	Family Tree DNA
EF177437	Europe	Portugal	X3	13	EU400619		Unknown	W4	Family Tree DNA
EF177414	Europe	Portugal	I1a1	13	EU744586		Unknown	W5	Family Tree DNA
AF382003	Europe	Spain	W1	5	FJ449703		Unknown	W6	Family Tree DNA
AF382007	Europe	Spain	I5a1	5	FJ472633		Unknown	W1	Family Tree DNA
EF486519	Europe	Czech Republic	N1a1a1	14	FJ473381		Unknown	W6	Family Tree DNA
FJ008043	Europe	Germany	X2	Family Tree DNA	GQ149695		Unknown	W6	Family Tree DNA
HM370114	Europe	Germany	X2c1	Family Tree DNA	HM352797		Unknown	W	Family Tree DNA
FJ348156	Europe	Germany	X2c1a	16	JF829690		Unknown	W3a1	Family Tree DNA
FJ348158	Europe	Germany	X2c1a	16	EU091245		Unknown	I4	Family Tree DNA
FJ348159	Europe	Germany	X2c1a	16	EU564849		Unknown	I1	Family Tree DNA
FJ348160	Europe	Germany	X2c1a	16	EU694173		Unknown	I1a	Family Tree DNA
HM125971	Europe	Germany	W5a1	Family Tree DNA	HQ221886		Unknown	N1c	Family Tree DNA
GU290207	Europe	Germany	N1a1b	2	JF930639		Unknown	N2a	Family Tree DNA
GU290209	Europe	Germany	N1a1a1a	2	LevinCHR		Unknown	X2d	28
HM047061	Europe	Germany	N1a1a2	Family Tree DNA	FJ457949		USA	X2b	Family Tree DNA
FJ348184	Europe	Germany	N1a1a2	16	GQ200588		USA	X2	Family Tree DNA
GU817015	Europe	Germany	W5a	Family Tree DNA	GU593980		USA	X2c1	Family Tree DNA
HM448049	Europe	Poland	X2d	Family Tree DNA	GU945542		USA	X2b4a	Family Tree DNA
EF609013	Europe	Poland	W1	Family Tree DNA	HM453712		USA	X2b	Family Tree DNA
FJ472839	Europe	Poland	W3a1a	Family Tree DNA	GU045487		USA	W4a	Family Tree DNA
EF153811	Europe	Poland	N1c	14	GU147938		USA	W3a	Family Tree DNA

ReidlaEs39	Europe	Estonia	X2c1a	7	GU726895		USA	W5a	Family Tree DNA
AY339510	Europe	Finland	X2c1	17	GU828018		USA	W5a1	Family Tree DNA
AY339511	Europe	Finland	X2c1a	17	JF275845		USA	W6	Family Tree DNA
AY339512	Europe	Finland	X2c1a	17	JF902025		USA	W6	Family Tree DNA
AY339513	Europe	Finland	X2b	17	GU290210		USA	N1a1a2	2
AY339460	Europe	Finland	W4	17	GU290211		USA	N1a1a2	2
AY339461	Europe	Finland	W1	17	GU290213		USA	N1a1a3	2
AY339462	Europe	Finland	W1b	17	FJ968796		USA	I1b	Family Tree DNA
AY339463	Europe	Finland	W1b	17	HM804481		USA	I4	Family Tree DNA
AY339464	Europe	Finland	W1b	17	HQ326985		USA	I2a	Family Tree DNA
AY339465	Europe	Finland	W1b	17	HQ658465		USA	I5	Family Tree DNA
AY339466	Europe	Finland	W1b1	17	HQ695930		USA	I2a	Family Tree DNA
AY339467	Europe	Finland	W1b1	17	AY195787		Native American	X2a2	23
AY339468	Europe	Finland	W1b1	17	EU095242		Native American	X2a1b1a	30
AY339469	Europe	Finland	W1b1	17	EU095243		Native American	X2a1b	30
AY339470	Europe	Finland	W1b1	17	EU095244		Native American	X2a1c	30
AY339471	Europe	Finland	W1b1	17	EU095245		Native American	X2a1b1a	30
AY339472	Europe	Finland	W1b1	17	EU095246		Native American	X2a1b1a	30
AY339473	Europe	Finland	W1b1	17	EU095247		Native American	X2a1	30
AY339474	Europe	Finland	W1b1	17	EU095248		Native American	X2a1a1	30
AY339475	Europe	Finland	W1a	17	EU095249		Native American	X2a1a1	30
AY339476	Europe	Finland	W1a	17	EU095250		Native American	X2a1a1	30
AY339477	Europe	Finland	W1a	17	EU095251		Native American	X2a1a1	30
AY339478	Europe	Finland	W1a	17	EU439939		Native American	X2a1	31
AY339479	Europe	Finland	W1a	17	FJ168757		Native American	X2a2	32
AY339480	Europe	Finland	W1a	17	FJ168758		Native American	X2a1a	32
AY339481	Europe	Finland	W1a	17	FJ168759		Native American	X2a1a	32
AY339482	Europe	Finland	W1a	17	FJ168760		Native American	X2a1b	32
AY339483	Europe	Finland	W1a	17	FJ168762		Native American	X2a1b1	32
AY339484	Europe	Finland	W1a	17	FJ168763		Native American	X2a1b1a	32
AY339485	Europe	Finland	W1a	17	FJ168764		Native American	X2a1c	32
AY339486	Europe	Finland	W1a	17	FJ825753		Native American	X2a2	Family Tree DNA
AY339487	Europe	Finland	W1a	17	FJ887963		Native American	X2a2	Family Tree DNA
AY339488	Europe	Finland	W1a	17	FJ168756		Canada	X2g	32
AY339489	Europe	Finland	W1a	17	FJ168761		Canada	X2a1b	32

AY339490	Europe	Finland	W1a	17	FJ168765		Canada	X2a1	32
AY339491	Europe	Finland	W1a	17	FJ168766		Canada	X2a1	32
AY339492	Europe	Finland	W1a	17	HM214761		Canada	W3a	Family Tree DNA
AY339493	Europe	Finland	W1a	17	GU294854		Canada	I2	Family Tree DNA
AY339494	Europe	Finland	W1a	17	AY714018		India	W4a	29
AY339495	Europe	Finland	W1a	17	AY714039		India	W1c	29
AY339496	Europe	Finland	W1a	17	AY714043		India	W3a1	29
FJ543390	Europe	Finland	W1e	Family Tree DNA	GU002155		India	W3a1	Family Tree DNA
AY339497	Europe	Finland	I2a1	17	HM560726		India	W1c	Family Tree DNA
AY339498	Europe	Finland	I2b	17	HM589047		India	W3	Family Tree DNA
AY339499	Europe	Finland	I2b	17	GU290206		India	N1a1a1	2
AY339500	Europe	Finland	I2b	17	GU290208		India	N1a1a1	2
AY339501	Europe	Finland	I2b	17	AY714041		India	I1b	29
AY339502	Europe	Finland	I1a1a	17	AY714008		India	N1d	29
AY339503	Europe	Finland	I1a1a	17	EF153772		Siberia	X2e2	14
AY339504	Europe	Finland	I1a1a	17	EF153776		Siberia	X2e2	14
AY339505	Europe	Finland	I1a1a	17	EF153793		Siberia	X2e2	14
AY339506	Europe	Finland	I1a1a	17	EF153832		Siberia	X2e2	14
AY339507	Europe	Finland	I1a1a	17	EF153778		Siberia	N1a1a1a	14
AY339508	Europe	Finland	I1a1a	17	EF153786		Siberia	I4	14
AY339509	Europe	Finland	I1a1a	17	GU590993		Australia	I3	Family Tree DNA
FJ147306	Europe	Russia	X2e2	18	DQ372887		New Zealand	W3a1	33
JQ797762	Europe	Greece	J1b1a2a	34	JF837819	Europe	Finland	T1a1a1	36
HM852779	South Caucasus	Armenia	J1b1a2a	3	JN089342	Europe	England	T1a1a1h	36
JQ797763	Europe	Italy	J1b1a2b	34	JQ702680		Unknown	T1a1a1h	35
JQ797764	Europe	Italy	J1b1a2b	34	JF905570		USA	T1a1a1k	36
EF660916	Europe	Italy	J1b1a3	10	JF926125		Unknown	T1a1a1k	36
JQ703595	Europe	England	J1b1a3	35	JQ702959		Unknown	T1a1a1k	35
JF286633	South Caucasus	Armenia	J1b1a3	36	JQ705707		Unknown	T1a1a1k	35
AY714035	Europe	India	J1b1a1	29	JQ702728		Unknown	T1a1a1k	35
JQ705247		Unknown	J1b1a1	35	JF830642	Europe	England	T1a1a1	36
JQ702259		Unknown	J1b1a1	35	JF836809		Unknown	T1a1a1	36
JQ701916		Unknown	J1b1a1	35	JQ705663	Europe	England	T1a1a1	35
JQ705093		Unknown	J1b1a1	35	JQ702035		Unknown	T1a1a1	35
JQ702434	Europe	Spain	J1b1a1	35	JQ701940	Europe	Slovakia	T1a1a1	35

AY495235		USA	J1b1a1	37	JF830257		Canada	T1a1a1	36
JQ704791		Unknown	J1b1a1	35	JQ702274	Europe	England	T1a1a1	35
JQ704942	Europe	France	J1b1a1	35	JQ703708	Europe	Ireland	T1a1a1	35
JQ704609		Unknown	J1b1a1	35	JQ702962		Unknown	T1a1a1	35
FJ178380	Europe	Italy	J1b1a1	38	JQ704967	Europe	Netherlands	T1a1a1	35
JQ701902		Unknown	J1b1a1	35	JQ704594	Europe	Ireland	T1a1a1	35
JQ797760	Europe	Italy	J1b1a1d	34	JQ705269		Unknown	T1a1a1	35
JQ705588		Unknown	J1b1a1d	35	JQ705335	Europe	France	T1a1a1	35
AY495238		USA	J1b1a1d	37	JQ705497		Unknown	T1a1a1	35
JQ701852		Unknown	J1b1a1	35	JQ705955		Unknown	T1a1a1	35
JQ797761	Europe	France	J1b1a1e	34	JF930640		Unknown	T1a1a1	36
JQ705792		Unknown	J1b1a1e	35	EF645646		Unknown	T1a1a2	36
JQ704777		Unknown	J1b1a1e	35	JF832384		Unknown	T1a1a2	36
EU915478	Europe	Italy	J1b1a1e	38	JQ705352		Unknown	T1a1a2	35
JQ703585	Europe	Germany	J1b1a1e	35	AY495296		USA	T1a1a2	37
JQ704626	Europe	United Kingdom	J1b1a1e	35	JQ798007	Europe	Italy	T1a1b	34
HM856621		USA	J1b1a1e	36	JQ798008	Europe	Italy	T1a1b	34
JQ703182	Europe	Czech Republic	J1b1a1e	35	JQ798009	South Caucasus	Azerbaijan	T1a1b	34
FJ213450		USA	J1b1a1a	36	JQ798023	South Asia	India	T1a+16362	34
JQ704118	Europe	Ireland	J1b1a1a	35	JQ798010	Europe	Canary Islands	T1a1c1	34
JQ703822		Unknown	J1b1a1a	35	EF177441	Europe	Portugal	T1a1c1	13
JQ702976		Unknown	J1b1a1a	35	JQ798011	Europe	Crete	T1a1c1	34
HM600785	Europe	England	J1b1a1a	36	EF660961	Europe	Italy	T1a1c1	10
HQ286325		USA	J1b1a1a	36	FJ348220	Europe	Italy	T1a1c1	16
JQ703574		Unknown	J1b1a1a	35	JQ702460	Europe	France	T1a1c3	35
JQ703249		Unknown	J1b1a1a	35	JN083377	South Caucasus	Armenia	T1a1c	36
JQ702695		Unknown	J1b1a1a	35	JQ798016	Near East	Iraq	T1a1c	34
JQ703588	Europe	Ireland	J1b1a1a	35	JQ798012	Europe	Italy	T1a1c	34
JQ701847	Europe	Scotland	J1b1a1a	35	JQ798013	Europe	Italy	T1a1c2	34
JQ705909		Unknown	J1b1a1	35	HM852775	South Caucasus	Armenia	T1a1c2	3
JQ703802		Unknown	J1b1a1	35	JQ798014	Europe	Italy	T1a1c	34
JQ703569		Unknown	J1b1a1	35	JQ798015	Arabian Peninsula	Kuwait	T1a1c	34

AY495234		USA	J1b1a1	37	JQ798017	Europe	Italy	T1a1d	34
JQ705923	Europe	Germany	J1b1a1	35	JQ798018	Anatolia	Turkey	T1a1d	34
AY495231		USA	J1b1a1c	37	JF831941		Unknown	T1a1	36
JQ705964		Unknown	J1b1a1c	35	JQ798019	Europe	Italy	T1a1e1	34
HQ543056		USA	J1b1a1c	36	JQ798020	Europe	Crete	T1a1e1	34
AY495233		USA	J1b1a1b	37	EU979542		USA	T1a1g	36
JQ702866		Unknown	J1b1a1b	35	JQ702040		Unknown	T1a1g	35
AY495237		USA	J1b1a1b	37	JQ702952	Europe	Scotland	T1a1g	35
AY495232		USA	J1b1a1b	37	JQ798021	Arabian Peninsula	Kuwait	T1a1f	34
JQ701907	Europe	Russia	J1b1a1b	35	EU935439	North Africa	Egypt	T1a1f	54
AY495236		USA	J1b1a1b	37	EU935455	North Africa	Egypt	T1a1f	4
JQ704097		Unknown	J1b1a1b	35	JQ798022	Near East	Israel	T1a1	34
JQ797765	Arabian Peninsula	Kuwait	J1b1b1	34	EF177409	Europe	Portugal	T1a	13
EF397558	Europe	Russia	J1b1b1	14	JQ705058		Unknown	T1a	35
FJ624455	Europe	Italy	J1b1b1	36	JQ798024	Europe	Italy	T1a	34
JQ797766	Near East	Iran	J1b1b1a	34	JQ703693	Europe	Lithuania	T1a	35
EF556155	Near East	Iran Jew	J1b1b1a	1	JQ705713		Unknown	T1a	35
EF397562	Europe	Russia	J1b1b1a	14	JQ798025	Europe	Italy	T1a15	34
JQ797768	Near East	Iran	J1b1b1	34	JQ798027	Near East	Israel	T1a2	34
HM852784	South Caucasus	Armenia	J1b1b1	3	JQ798026	Near East	Iraq	T1a2	34
JQ797769	Near East	Iraq	J1b1b1b	34	EU935447	North Africa	Egypt	T1a2a	4
JF939049	South Caucasus	Armenia	J1b1b1b	36	EU935441	North Africa	Egypt	T1a2a	4
JF929909	South Caucasus	Armenia	J1b1b1	36	EU935435	North Africa	Egypt	T1a2a	4
JQ797767	Near East	Iraq	J1b1b1	34	EU935452	North Africa	Egypt	T1a2a	4
JQ703545	Near East	Lebanon	J1b1b1	35	EU935444	North Africa	Egypt	T1a2a	4
AY714033	South Asia	India	J1b1b	29	JQ798028	Near East	Iraq	T1a14	34
JQ797770	Near East	Iran	J1b1b	34	JQ798029	Europe	Crete	T1a14	34
JQ703656		Unknown	J1b1b2	35	JQ798030	South Asia	India	T1a3	34
JQ797772	Europe	France	J1b2	34	JQ798031	North Africa	Algeria	T1a3	34
DQ282492		USA	J1b2	39	JQ703724	Europe	Germany	T1a3	35
DQ282490		USA	J1b2	39	JQ798032	Europe	Greece	T1a3	34
DQ282489		USA	J1b2	39	JF937680	Europe	England	T1a3	36
DQ282488		USA	J1b2	39	JQ798033	Europe	Italy	T1a4a	34
GU390312		USA	J1b2	36	JQ798034	Near East	Iran	T1a4a	34

JQ797771	Arabian Peninsula	Kuwait	J1b2a	34	JQ798035	South Caucasus	Azerbaijan	T1a4b	34
JQ702001		Unknown	J1b2a	35	JQ798036	Arabian Peninsula	Kuwait	T1a4b	34
JQ702684		Unknown	J1b2	35	JQ702139	Europe	United Kingdom	T1a6	35
JQ703211		Unknown	J1b2	35	JN104727		Unknown	T1a6	36
AF381987	North Africa	Morocco	J1b2	5	EU369395	Europe	Spain	T1a6	36
EF556169	Near East	Iraq	J1b2b	1	JQ798037	South Caucasus	Georgia	T1b1	34
JN648827	South Caucasus	Armenia	J1b2b	36	JQ798038		Jordan	T1b1	34
EF660985	Europe	Italy	J1b2	10	FJ878777	Europe	United Kingdom	T1b5	36
JQ797773	Arabian Peninsula	Kuwait	J1b2	34	JQ798039	Europe	Italy	T1b2	34
JQ797774	Near East	Iraq	J1b2	34	JQ798040	Europe	Italy	T1b2	34
JQ797783	Anatolia	Turkey	J1b7	34	JQ798041	Europe	Italy	T1b2	34
JQ797784	Near East	Iraq	J1b7	34	JQ798042	Near East	Iran	T1b3	34
JQ797775	Europe	Greece	J1b3b	34	JQ798043	Anatolia	Turkey	T1b3	34
HM852838	Near East	Iran	J1b3b	3	JQ705372		Unknown	T1b3	35
EF583177	Europe	Slovakia	J1b3a	40	JQ798044	Europe	Karachay-Cherkessia Rep. (Russia)	T1b3	34
EF583175	Europe	Slovakia	J1b3a	40	JQ798045	Europe	Kabardino-Balkaria Rep. (Russia)	T1b3	34
HM594676	South Caucasus	Armenia	J1b3a	36	JQ798046	Arabian Peninsula	Saudi Arabia	T1b	34
JQ797786	Arabian Peninsula	Dubai	J1b	34	JQ798047	North Africa	Egypt	T1b	34
JQ064573	North Africa	Algeria	J1b8	36	JQ798048	Europe	Italy	T1b	34
HQ914447	South Caucasus	Armenia	J1b8	36	JQ798049	Near East	Iraq	T1b4	34
JQ797782	Near East	Iraq	J1b6b	34	JQ798050	Europe	Greece	T1b4	34
JQ797780	Arabian Peninsula	Kuwait	J1b6	34	JQ798051	South Asia	India	T1b	34
JQ797781	Europe	Italy	J1b6a	34	AY495298		USA	T2a1a	37
DQ282491		USA	J1b6a	39	AY495301		USA	T2a1a	37
JQ702538		Unknown	J1b6a	35	JN383991	Europe	Norway	T2a1a	36

JQ797777	North Ossetia	Alania Rep. (Russia)	J1b9a	34	JQ705057	Europe	England	T2a1a	35
JQ797778	Near East	Iraq	J1b9a	34	JQ703890	Europe	France	T2a1a	35
JQ797785	Europe	Cyprus	J1b	34	JQ702594	Europe	Scotland	T2a1a	35
JQ797776	Europe	Greece	J1b4	34	JQ702379		Unknown	T2a1a	35
JN561091	South Caucasus	Armenia	J1b4a1	36	JQ798052	Near East	Iraq	T2a1a	34
HM852827	Near East	Iran	J1b4a1	3	JQ798053	Europe	Italy	T2a1a	34
HM992836	South Caucasus	Armenia	J1b4a2	36	AY714022	South Asia	India	T2a1a	29
HQ637485	South Caucasus	Armenia	J1b4a2	36	AY495302		USA	T2a1a5	37
HM852835	Near East	Iran	J1b5	3	JQ705499		Unknown	T2a1a5	35
JQ797779	Near East	Iraq	J1b5a	34	JQ702880		Unknown	T2a1a5	35
HM852765	South Caucasus	Armenia	J1b5a	3	AY495304		USA	T2a1a	37
JQ797787	Europe	Ukraine	J1c1	34	JQ045864	Europe	Denmark	T2a1a7	36
AY495216		USA	J1c1a	37	JQ701880	Europe	Denmark	T2a1a7	35
AY495210		USA	J1c1a	37	AY495303		USA	T2a1a3	37
JQ705874		Unknown	J1c1a	35	AY495300		USA	T2a1a3	37
JQ705114		Unknown	J1c1a	35	JF833041		Unknown	T2a1a3	36
FJ447985	Europe	Ireland	J1c1a	36	JQ702969	Europe	Scotland	T2a1a	35
JQ702620		Unknown	J1c1a	35	JQ703088	Europe	England	T2a1a	35
FJ348216	Europe	Italy	J1c1	16	JQ706008		Unknown	T2a1a	35
JQ702440	Europe	Belarus	J1c1d	35	JQ705133	Europe	Ireland	T2a1a	35
JQ705051		Unknown	J1c1d	35	JQ702611		Unknown	T2a1a	35
JQ797788	Europe	Austria	J1c1b	34	FJ656215	Europe	Germany	T2a1a	54
JQ702224	Europe	England	J1c1b	35	FJ348181	Europe	Italy	T2a1a1	16
AY495202		USA	J1c1b	37	FJ348191	Europe	Italy	T2a1a1	16
JQ705470		Unknown	J1c1b	35	FJ348180	Europe	Italy	T2a1a1	16
JQ703783		Unknown	J1c1b	35	FJ348186	Europe	Italy	T2a1a1	16
JQ705562	Europe	Ireland	J1c1b	35	GU553285		USA	T2a1a2	36
HQ166708		USA	J1c1b2a	36	JQ701997		Unknown	T2a1a2	35
JQ702020	Europe	Ireland	J1c1b2a	35	JQ704617	Europe	United Kingdom	T2a1a2	35
JQ704908	Europe	United Kingdom	J1c1b2a	35	JN120787		Unknown	T2a1a2	36
JQ702981		Unknown	J1c1b2a	35	HQ704899		Australia	T2a1a2	36
JQ705141		Unknown	J1c1b2	35	JN630630		Unknown	T2a1a2	36
JQ703594	Europe	United Kingdom	J1c1b2	35	JN012468		Unknown	T2a1a2	36
JQ702081	Europe	Scotland	J1c1b2	35	JQ702036		Unknown	T2a1a6	35
JQ797789	Europe	Ukraine	J1c1b1a1	34	JQ704797	Europe	Czech Republic	T2a1a6	35

JQ797790	Europe	Romania	J1c1b1a1	34	JF958082		Unknown	T2a1a	36
JQ703825		Unknown	J1c1b1a	35	JQ701972		Unknown	T2a1a	35
JQ703599	Europe	Spain	J1c1b1a2	35	JQ798054	Europe	Italy	T2a1b	34
JQ705164	Europe	Scotland	J1c1b1a2	35	JQ702344	Europe	Greece	T2a1b	35
JQ797791	Europe	Italy	J1c1b1a	34	JQ798056	Europe	Adygea Rep. (Russia)	T2a1b2b	34
AY714034	South Asia	India	J1c1b1a	29	JQ798057	South Caucasus	Azerbaijan	T2a1b2b	34
FJ502349		Unknown	J1c1b1a	36	GU123001	Europe	Russia	T2a1b2a	20
JQ703029		Unknown	J1c1b1	35	JQ704821		Unknown	T2a1b2a	35
JQ797792	Europe	Italy	J1c1b1	34	JQ703730		Unknown	T2a1b1a	35
EF177420	Europe	Portugal	J1c1b1	13	JQ702356		Unknown	T2a1b1a	35
JQ703623	Europe	Ireland	J1c1b1	35	JQ703776		Unknown	T2a1b1a	35
EF452293		Unknown	J1c1b1	36	JQ046361	Europe	Ireland	T2a1b1a	36
JQ703464		Unknown	J1c1c	35	JQ702972	Europe	Italy	T2a1b1a	35
JQ704867		Unknown	J1c1c	35	JQ705641	Europe	Germany	T2a1b1a	35
JQ703653		Unknown	J1c1c	35	JQ798055	Europe	England	T2a1b1a1a	34
JQ705770	Europe	Ireland	J1c1f	35	JQ703692		Unknown	T2a1b1a1a	35
JQ702685		Unknown	J1c1	35	JQ703700	Europe	Germany	T2a1b1a1a	35
FJ348177	Europe	Italy	J1c1	16	JQ703868		Unknown	T2a1b1a1a	35
JQ797797		Brazil-France	J1c1	34	JQ703149	Europe	Germany	T2a1b1a1	35
JQ797793		USA	J1c1	34	JF930649		Unknown	T2a1b1a1	36
JQ797794	Europe	Italy	J1c1	34	JQ702937		Unknown	T2a1b1a1c	35
JQ703557		Unknown	J1c1	35	JN024624	Europe	England	T2a1b1a1b	36
JQ797795	Europe	Italy	J1c1	34	JN048933		Unknown	T2a1b1a1b	36
JQ797796	Europe	Italy	J1c1	34	JQ798058	Near East	Iraq	T2a1	34
EF177431	Europe	Portugal	J1c1	13	HM852781	South Caucasus	Armenia	T2a1	3
JQ797798	Europe	England	J1c1d	34	JF927949		Unknown	T2a1	36
AY195754		Unknown	J1c1d	23	JF944823		Unknown	T2a2	36
AY495198		USA	J1c1e	37	FJ238094	South Caucasus	Armenia	T2a2	36
AY495208		USA	J1c1e	37	HM852808	South Caucasus	Azerbaijan	T2a3	3
EU915479	Europe	Italy	J1c2c	38	AY195767	Europe	Europe	T2b1	23
JQ797822	Europe	Italy	J1c2c	34	AY495267		USA	T2b1	37

JQ702581		Unknown	J1c2	35	EF660922	Europe	Italy	T2b1	10
JQ705421	Europe	Finland	J1c2	35	EF177444	Europe	Portugal	T2b1	13
JN663354	South Caucasus	Armenia	J1c2	36	HM583750	Europe	France	T2b1	36
JQ703591		Unknown	J1c2	35	JQ703293		Unknown	T2b1	35
JQ704935		Unknown	J1c2	35	AY714016	South Asia	India	T2b2	29
JQ701851		Unknown	J1c2h	35	AY495305		USA	T2b2+16304!	37
JQ705593	Europe	United Kingdom	J1c2h	35	EU683970	Europe	Ireland	T2b2b	36
JQ705996		Unknown	J1c2h	35	JF939816		Unknown	T2b2b	36
JQ704033	Europe	Slovakia	J1c2	35	JF929911		Unknown	T2b2b	36
JQ797820	Anatolia	Turkey	J1c2b	34	EF660978	Europe	Italy	T2b2b	10
JQ797821	Europe	Italy	J1c2b	34	JN032298	Europe	Belgium	T2b2b1	36
EU155191		Unknown	J1c2a	36	JQ705119		Unknown	T2b2b1	35
JQ703794		Unknown	J1c2a	35	JN084060		Unknown	T2b2b1	36
JQ702445	Europe	Norway	J1c2a	35	JQ705208		Unknown	T2b2b1	35
JQ702219	Europe	Norway	J1c2a	35	JQ701832		Unknown	T2b2b1	35
JQ027716	Europe	Russia	J1c2a	36	JN035224		Unknown	T2b2+16304!	36
JQ704432	Europe	Germany	J1c2a6	35	HM852840	Near East	Iran	T2b2+16304!	3
JQ705715	Europe	Russia	J1c2a6	35	AY495299		USA	T2b2+16304!	37
JQ705129	Europe	Czechoslovakia	J1c2a6	35	GU183768	Europe	France	T2b2+16304!	36
JQ797813	Europe	Romania	J1c2a7	34	AY495275		USA	T2b3b	37
JQ702366		Unknown	J1c2a7a	35	JN024623		Unknown	T2b3b	36
JQ702724	Europe	Ireland	J1c2a7a	35	JF937112	Europe	England	T2b3b	36
JQ797814		Siberia	J1c2a7	34	JF965447		Unknown	T2b3b	36
FJ499471		USA	J1c2a8	36	JN043363		Unknown	T2b3b	36
JQ703785	Europe	England	J1c2a8	35	HM122274		Tunisia	T2b3	36
AY495223		USA	J1c2a9	37	JQ701927		Unknown	T2b3d	35
JQ704829	Europe	England	J1c2a9	35	JQ705037		Unknown	T2b3d	35
JQ797799	Europe	Ukraine	J1c2a10	34	HM852813	South Caucasus	Azerbaijan	T2b3+151	3
AY495225		USA	J1c2a10	37	JF837334	Europe	Ireland	T2b3e	36
AY495224		USA	J1c2a	37	JQ702951	Europe	Scotland	T2b3e	35
JQ703870	Europe	Crete	J1c2a	35	JQ705417	Europe	England	T2b3e	35
AY495219		USA	J1c2a	37	JQ705199		Unknown	T2b3e	35
FJ499472	Europe	Norway	J1c2a11	36	JQ705913		Unknown	T2b3+151	35
JQ704965		Unknown	J1c2a11	35	JQ703699	Europe	England	T2b3+151	35
DQ358973	Europe	Germany	J1c2a	41	JQ702932		Unknown	T2b3c	35

JQ705054	Europe	Germany	J1c2a	35	JQ703099	Europe	France	T2b3c	35
HM852774	South Caucasus	Armenia	J1c2a	3	JQ705786	Europe	Portugal	T2b3c	35
JQ705724		Unknown	J1c2a	35	JQ798063	Europe	Italy	T2b3a1	34
JQ702515		Unknown	J1c2a	35	JQ798060	North Africa	Morocco	T2b3a1	34
JQ797800	Europe	Italy	J1c2a5	34	JQ798062	Europe	Italy	T2b3a1	34
JQ703675	Europe	United Kingdom	J1c2a5	35	EF660963	Europe	Italy	T2b3a1	10
JQ702311	Europe	England	J1c2a5	35	JQ798059	Europe	Italy	T2b3a1	34
JQ797801	Europe	France	J1c2a5	34	JQ798061	Europe	Italy	T2b3a1	34
JQ797802	Europe	Canary Islands	J1c2a5	34	DQ523660	Europe	Sardinia	T2b3a	15
JQ797803	Europe	Bulgaria	J1c2a5a	34	DQ523651	Europe	Sardinia	T2b3a	15
GU123042	Europe	Russia	J1c2a5a	20	DQ523649	Europe	Sardinia	T2b3a	15
JQ797808		Siberia	J1c2a5a	34	JQ705834		Unknown	T2b	35
JQ797804	Europe	Greece	J1c2a5	34	HQ696459		USA	T2b9	36
JQ703746	Europe	Italy	J1c2a5	35	JQ703011	Europe	Germany	T2b9	35
JQ797805	Europe	Greece	J1c2a5	34	JQ702737		Unknown	T2b9	35
JQ797806	Europe	Greece	J1c2a5	34	EU747356	Europe	Germany	T2b+152	36
JQ797807	Europe	Greece	J1c2a5	34	JQ798086	Near East	Israel	T2b+152	34
JQ797809	Europe	Albania	J1c2a5	34	JQ701867		Unknown	T2b+152	35
JQ797810	Europe	Romania	J1c2a5	34	HQ840646		USA	T2b22	36
EF660952	Europe	Italy	J1c2a	10	JQ703263		Unknown	T2b22	35
JQ797811	Europe	Estonia	J1c2a12	34	JQ025225	Europe	England	T2b21a	36
FJ190383		USA	J1c2a12	36	JQ705895		Unknown	T2b21a	35
AY495226		USA	J1c2a	37	JF340114		Newfoundland	T2b21a	36
JQ797812	Europe	Lithuania	J1c2a	34	JQ705338		Unknown	T2b21b	35
JQ797815	Europe	Poland	J1c2a13	34	JQ702492		Unknown	T2b21b	35
JQ705224		Unknown	J1c2a13	35	HQ917079	Europe	Norway	T2b21	36
JQ797816	Europe	Greece	J1c2a13	34	AY195745		Unknown	T2b21	23
JQ797817	Europe	Greece	J1c2a	34	JQ705647		Unknown	T2b21	35
JQ703556		Unknown	J1c2a	35	JQ702188	Europe	Lithuania	T2b+152	35
JQ701878	Europe	Germany	J1c2a	35	JQ705242		Unknown	T2b4	35
JQ797818	Europe	Greece	J1c2a	34	JQ798064		Nepal	T2b4e	34
JQ797819	Europe	Romania	J1c2a3a	34	JQ798065		Nepal	T2b4e	34
FJ449571	Europe	England	J1c2a3a	36	HM852802	South Caucasus	Azerbaijan	T2b4e	3
JQ702499	Europe	British Isles	J1c2a3a	35	JQ704596		Unknown	T2b4f	35
JQ702018	Europe	Finland	J1c2a3	35	JQ703508		Unknown	T2b4f	35

JQ703718	Europe	England	J1c2a3	35	JQ701897		Unknown	T2b4f	35
JQ705961		Unknown	J1c2a3	35	JQ798066	Arabian Peninsula	Kuwait	T2b4h	34
GU949564	Europe	Russia	J1c2a3	36	EU007872	Europe	Kazakhstan Rep. (Russia)	T2b4h	48
EF177422	Europe	Portugal	J1c2a3	13	JQ704869		Unknown	T2b4	35
JQ703882		Unknown	J1c2a3	35	JQ702800	Europe	England	T2b4	35
EU573192		Unknown	J1c2a3b	36	JQ702464	Europe	Switzerland	T2b4	35
JQ702741	Europe	Germany	J1c2a3b	35	JQ701999		Unknown	T2b4	35
FJ348202	Europe	Italy	J1c2a3b	16	JQ798067	Arabian Peninsula	Kuwait	T2b4	34
HQ336424	Europe	Ukraine	J1c2a3b1	36	GU123018	Europe	Russia	T2b4a	20
GU592034	Europe	Austria	J1c2a3b1	42	EF177439	Europe	Portugal	T2b4a	13
GU592019	Europe	Austria	J1c2a3b1	42	JQ704911	Europe	England	T2b4a	35
AY495230		USA	J1c2a	37	JQ702564		Unknown	T2b4a	35
AY495228		USA	J1c2a	37	JQ705600		Unknown	T2b4a	35
AY495221		USA	J1c2a	37	JQ798068	Europe	Italy	T2b4a1	34
AY495218		USA	J1c2a	37	JQ704738		Unknown	T2b4a1	35
HM590710		USA	J1c2a	36	JQ702323		Unknown	T2b4a1	35
HM803933	Europe	Ukraine	J1c2a	36	JF836084		Unknown	T2b4a	36
HM856585	Europe	Finland	J1c2a	36	JF937089	Europe	England	T2b4a	36
JQ705442		Unknown	J1c2a	35	JQ703501	Europe	Ireland	T2b4	35
JQ703461		Unknown	J1c2a	35	JQ704597		Unknown	T2b4	35
JQ702715		Unknown	J1c2a	35	JQ702516	Europe	England	T2b4	35
JQ702110		Unknown	J1c2a	35	JQ798069		USA	T2b4	34
JQ703612	Europe	Scotland	J1c2a	35	JQ798070	Europe	Greece	T2b4	34
JQ703575	Europe	Norway	J1c2a14	35	JN037468		Unknown	T2b4	36
JQ703789	Europe	England	J1c2a14	35	JN232198		Unknown	T2b4	36
JQ703818	Europe	England	J1c2a	35	JQ704746	Europe	England	T2b4	35
HQ260985		USA	J1c2a15	36	JQ798071	Europe	Italy	T2b4g	34
JQ704946		Unknown	J1c2a15	35	EF660982	Europe	Italy	T2b4g	10
AY495220		USA	J1c2a1c	37	JQ702012		Unknown	T2b4	35
JQ705006		Unknown	J1c2a1c	35	EU492455		USA	T2b4	36
JQ702742		Unknown	J1c2a1c	35	JF904739		Unknown	T2b4	36
DQ787109		USA	J1c2a1b	36	JQ705939	Europe	England	T2b4b	35
JQ703955		Unknown	J1c2a1b	35	JF900491		Unknown	T2b4b	36

AY495227		USA	J1c2a1a	37	JQ705883		Unknown	T2b4b	35
JQ702390		Unknown	J1c2a1a	35	JQ704925		Unknown	T2b4b	35
AY495229		USA	J1c2a1a	37	JQ704688	Europe	Germany	T2b4b	35
AY495222		USA	J1c2a2	37	JQ703842	Europe	United Kingdom	T2b4b	35
JQ703778		Unknown	J1c2a2	35	JN899566		Unknown	T2b4	36
JQ704554		Unknown	J1c2a2	35	JQ702058	Europe	France	T2b4	35
JQ703567		Unknown	J1c2a2d	35	JQ703710	Europe	Germany	T2b4	35
JQ703672		Unknown	J1c2a2d	35	JQ702158		Unknown	T2b4d	35
JQ704384		Unknown	J1c2a2	35	JQ703734	Europe	Ukraine	T2b4d	35
JQ705986		Unknown	J1c2a2	35	JQ702604		Unknown	T2b4c	35
AY495203		USA	J1c2a2	37	JQ704742	Europe	Germany	T2b4c	35
HM776018	Europe	Russia	J1c2a2b	36	JQ702278		Unknown	T2b4	35
JQ703610	Europe	United Kingdom	J1c2a2b	35	EU494628		Unknown	T2b4	36
JQ704710		Unknown	J1c2a2a	35	JQ798072	Near East	Iran	T2b5	34
JQ703613		Unknown	J1c2a2a	35	JQ798073	Europe	Italy	T2b5	34
JQ701968	Europe	United Kingdom	J1c2a2c	35	AY495287		USA	T2b5	37
JQ702706		Unknown	J1c2a2c	35	JQ702978	Europe	Germany	T2b5	35
JQ797823	Europe	Greece	J1c3	34	JQ702884		Unknown	T2b5	35
AY495215		USA	J1c3	37	JQ702666		Unknown	T2b5	35
JQ705360		Unknown	J1c3	35	AY495286		USA	T2b5	37
JQ701920		Unknown	J1c3	35	GQ304744		Unknown	T2b5	53
AY495214		USA	J1c3i	37	JQ703716	Europe	Germany	T2b5	35
JQ703753		Unknown	J1c3i	35	AY495283		USA	T2b5+8504	37
JQ797824	Europe	Greece	J1c3a1	34	JF958132		Unknown	T2b5a	36
AY495211		USA	J1c3a1	37	JQ702982	Europe	Switzerland	T2b5a	35
HM560728		USA	J1c3a1	36	JF833039		Unknown	T2b5a	36
JQ702747		Unknown	J1c3a1	35	JQ705432	Europe	England	T2b5a	35
JQ797825	Europe	Greece	J1c3a1	34	JQ703023	Europe	Sweden	T2b5a	35
JQ704745	Europe	England	J1c3a1	35	JQ701828		Unknown	T2b5a	35
HQ839858		Unknown	J1c3a1	36	JQ703712	Europe	England	T2b5a	35
JQ702646		Unknown	J1c3a1	35	JQ702885		Unknown	T2b5a	35
AY495217		USA	J1c3a2	37	AY495284		USA	T2b6	37
AY495213		USA	J1c3a2	37	JF891418		Unknown	T2b6b	36
JQ704615	Europe	Scotland	J1c3a2	35	JQ704785		Unknown	T2b6b	35
JQ703792	Europe	Netherlands	J1c3a2	35	JQ705047		Unknown	T2b6b	35

JQ703469	Europe	United Kingdom	J1c3a2	35	JQ705972		Unknown	T2b6b	35
JQ797826	Europe	Adygea Rep. (Russia)	J1c3k	34	JQ705995		Unknown	T2b6b	35
JQ702912		Unknown	J1c3k	35	JQ705357	Europe	Germany	T2b6	35
FJ445407	Europe	Ireland	J1c3b1a	36	AY495282		USA	T2b6a	37
HM627754		USA	J1c3b1a	36	JQ705494	Europe	Czech Republic	T2b6a	35
JQ703586	Europe	Scotland	J1c3b1a	35	FJ573258		USA	T2b6a	36
JQ705472		Unknown	J1c3b1a	35	JQ703772	Europe	Switzerland	T2b6a	35
JQ703671	Europe	Netherlands	J1c3b1	35	JQ798074	Europe	Italy	T2b23	34
JQ705447		Unknown	J1c3b1	35	JQ798075	Europe	Italy	T2b23	34
HQ696458		USA	J1c3b1	36	JQ703458	Europe	Belgium	T2b23	35
JQ705085	Europe	United Kingdom	J1c3b	35	JQ798076	Near East	Iran	T2b23	34
JQ702525	Europe	Ireland	J1c3b	35	FJ348211	Europe	Italy	T2b23a	16
JQ705280		Unknown	J1c3b	35	FJ348208	Europe	Italy	T2b23a	16
JQ705636	Europe	Ireland	J1c3b	35	JQ704788		Unknown	T2b23a	35
HM026752	Europe	Ireland	J1c3b2	36	HM852879	South Caucasus	Georgia	T2b	3
JQ702630		Unknown	J1c3b2	35	JQ704992	Europe	England	T2b	35
JQ048704	Europe	Portugal	J1c3b	36	JQ703933		Unknown	T2b	35
JQ703786	Europe	Scotland	J1c3	35	JQ703031		Unknown	T2b	35
JQ702063		Unknown	J1c3	35	JQ702824	Europe	England	T2b	35
EF660962	Europe	Italy	J1c3	10	JQ798080	Europe	Italy	T2b29	34
EU597522		USA	J1c3	27	JQ798081	Europe	France	T2b29	34
FJ603099		Unknown	J1c3	36	JF830105		Unknown	T2b	36
HQ709108		USA	J1c3f	36	JQ798082	Near East	Iraq	T2b	34
JQ797828	Europe	Ukraine	J1c3f	34	AY495272		USA	T2b	37
JQ701854		Unknown	J1c3f	35	JQ798083	Europe	Italy	T2b19	34
JQ704665	Europe	Ireland	J1c3f	35	EU744542		USA	T2b19	36
JQ705635		Unknown	J1c3f	35	JQ705899		Unknown	T2b19	35
JQ797829	Europe	Italy	J1c3f	34	JQ704699		Unknown	T2b19	35
JQ797833	Europe	Sweden	J1c3f	34	JQ703009	Europe	Ireland	T2b19b	35
JQ797831	Europe	Kabardino-Balkaria Rep. (Russia)	J1c3f	34	JQ705680	Europe	Scotland	T2b19b	35
JQ797830	Europe	Ukraine	J1c3f	34	JQ705664	Europe	United Kingdom	T2b19	35
JQ797832	Europe	Kazakhstan Rep. (Russia)	J1c3f	34	JQ701939	Europe	Germany	T2b	35

JQ701995		Unknown	J1c3f	35	JN603188	Europe	Norway	T2b	36
JQ797827	Europe	Estonia	J1c3f	34	JQ703083		Unknown	T2b	35
JQ797836	Europe	Italy	J1c3e2	34	JN419194		Unknown	T2b	36
JQ705308	Europe	England	J1c3e2	35	JN004135		Unknown	T2b	36
JQ702686		Unknown	J1c3e2	35	JQ798084	Europe	Adygea Rep. (Russia)	T2b	34
JQ704840	Europe	Sweden	J1c3e2	35	JQ798085	Europe	Kabardino-Balkaria Rep. (Russia)	T2b11	34
AY495195		USA	J1c3e1	37	JQ705009	Europe	Denmark	T2b11	35
JQ705868		Unknown	J1c3e1	35	JQ703713	Europe	England	T2b11	35
HM852874	Anatolia	Turkey	J1c3e1	3	JF968593	Europe	Belarus	T2b11	36
JQ703895	Europe	Norway	J1c3e1	35	JN038393		Unknown	T2b13a	36
JQ701822		Unknown	J1c3e1	35	JQ702084		Unknown	T2b13a	35
JQ702097		Unknown	J1c3e1	35	JQ703762		Unknown	T2b13a	35
JQ702691		Unknown	J1c3	35	JQ702435		Unknown	T2b13	35
JQ703597		Unknown	J1c3	35	JQ702963	Europe	United Kingdom	T2b13	35
JQ797834	Europe	France	J1c3	34	JQ703042	Europe	Netherlands	T2b13b	35
JQ797835	Europe	Italy	J1c3	34	JQ705492		Unknown	T2b13b	35
JF703252		USA	J1c3g	36	JQ701982		Unknown	T2b13	35
JQ702788	Europe	Norway	J1c3g	35	JQ705475	Europe	Ireland	T2b	35
JQ704779	Europe	United Kingdom	J1c3g	35	JQ798087	Europe	Italy	T2b+16296!	34
JQ704803		Unknown	J1c3g	35	JQ653137		USA	T2b+16296!	36
JQ706013		Unknown	J1c3g	35	JQ798089	Europe	Crete	T2b31	34
JQ705212		Unknown	J1c3g	35	JQ703715		Unknown	T2b31	35
HQ709168	Europe	France	J1c3c1	36	JQ705484	Europe	Ireland	T2b+16296!	35
JQ702395		Unknown	J1c3c1	35	JQ798078	Europe	Italy	T2b+16296!	34
JQ703565	Europe	Switzerland	J1c3c	35	JQ702476		Unknown	T2b+16296!	35
JQ702727	Europe	England	J1c3c2	35	JQ704973	Europe	Denmark	T2b+16296!	35
JQ703602	Europe	England	J1c3c2	35	JQ702980	Europe	Germany	T2b+16296!	35
JQ797837	Europe	Italy	J1c3h	34	JQ702965	Europe	Germany	T2b	35
JQ797838	Europe	Italy	J1c3h	34	JQ798088	Arabian Peninsula	Dubai	T2b	34
DQ523659	Europe	Sardinia	J1c3h	15	JQ705328	Europe	Belgium	T2b	35
AY495201		USA	J1c3d	37	HM055613		USA	T2b	36
JQ705701		Unknown	J1c3d	35	JQ704780	Europe	United Kingdom	T2b	35

AY495207		USA	J1c3d	37	AY495274		USA	T2b	37
JQ701961		Unknown	J1c3	35	AY495285		USA	T2b	37
JQ704211		Unknown	J1c3	35	AY495279		USA	T2b	37
JQ705922		Unknown	J1c3	35	AY495277		USA	T2b	37
DQ358974	Europe	Germany	J1c3	43	AY495273		USA	T2b	37
JQ702017		Unknown	J1c3	35	AY495270		USA	T2b	37
JQ705804		Unknown	J1c3	35	FJ384436	Europe	Europe	T2b	9
JQ703047	Europe	Germany	J1c3	35	FJ384435	Europe	Europe	T2b	9
HQ154118		USA	J1c3	36	JN032303	Europe	Finland	T2b	36
JQ702826	Europe	Ireland	J1c3	35	JN106403		Unknown	T2b	36
JQ797839	Europe	Italy	J1c4	34	HM101563	Europe	Spain	T2b	36
JQ705388		Unknown	J1c4	35	JQ702461	Europe	Ireland	T2b	35
JQ705414		Unknown	J1c4	35	JQ702979	Europe	Russia	T2b	35
JQ797840		Siberia	J1c4	34	JQ701840		Unknown	T2b	35
AY495197		USA	J1c4	37	HQ231912		USA	T2b	36
AY495200		USA	J1c4	37	JQ703953		Unknown	T2b	35
AF346983		USA	J1c4	44	JN315800	Europe	England	T2b	36
GU123008	Europe	Russia	J1c4	20	JN558762	Europe	England	T2b	36
JQ797841	Europe	Ukraine	J1c4b	34	JQ702312		Unknown	T2b	35
AY195774		Unknown	J1c4b	23	EU926622		Unknown	T2b	36
JQ703916		Unknown	J1c4b	35	JQ705948		Unknown	T2b	35
AY665667	Europe	Germany	J1c4b	45	JQ702953		Unknown	T2b	35
GU808335	Europe	Russia	J1c4b	36	HQ638221	South Caucasus	Armenia	T2b	36
DQ341085	Europe	Italy	J1c4	46	EF660928	Europe	Italy	T2b	10
DQ341086	Europe	Italy	J1c4	46	AY495269		USA	T2b	37
EU915476	Europe	Italy	J1c4	38	JF975728		Unknown	T2b15	36
FJ538285		USA	J1c4	36	EU682394		USA	T2b15	36
HM016082		USA	J1c4c	36	JF979034		USA	T2b15	36
JQ705633		Unknown	J1c4c	35	HQ287876	Europe	England	T2b	47
JQ705439		Unknown	J1c4c	35	HQ287875	Europe	England	T2b	47
JQ797842		Siberia	J1c5	34	AY495280		USA	T2b17a	37
JQ703750	Europe	Poland	J1c5	35	JQ705106		Unknown	T2b17a	35
JQ702776	Europe	France	J1c5c	35	JQ704783	Europe	Ireland	T2b17a	35
JQ704051		Unknown	J1c5c	35	JF957699	Europe	Ireland	T2b17	36
JQ704816	Europe	Greece	J1c5c	35	FJ460542	North Africa	Tunisia	T2b	6

HQ287873	Europe	England	J1c5d	47	FJ460555	North Africa	Tunisia	T2b	6
AY495204		USA	J1c5d	37	AY495281		USA	T2b	37
HQ907957		USA	J1c5d	36	JQ704897	Europe	Ireland	T2b	35
JQ702201		Unknown	J1c5d	35	AY495276		USA	T2b	37
JQ703604		Unknown	J1c5b	35	AY495278		USA	T2b	37
JQ705397		Unknown	J1c5b	35	AY495268		USA	T2b	37
JQ703639	Europe	Germany	J1c5b	35	AY495271		USA	T2b27	37
JQ797843	Europe	Italy	J1c5	34	JQ705444		Unknown	T2b27	35
JQ797844	South Asia	India	J1c5e	34	GU123010	Europe	Russia	T2b	20
JQ703740		Unknown	J1c5e	35	GU170817	South Asia	India	T2b7	55
JQ797845	Europe	Ukraine	J1c5a	34	JQ798079	Europe	Italy	T2b7a3	34
JQ703803	Europe	Ireland	J1c5a	35	JQ704643	Europe	Italy	T2b7a2	35
HQ287874	Europe	England	J1c5a1	47	JQ704997	Europe	France	T2b7a2	35
JQ705488		Unknown	J1c5a1	35	JQ702544		Unknown	T2b7a1	35
JQ703558	Europe	Ireland	J1c5a1	35	JQ703062	Europe	Sweden	T2b7a1	35
JQ702996	Europe	Germany	J1c5a1	35	JQ703294	Europe	British Isles	T2b7a1	35
JQ701921		Unknown	J1c5a1	35	JQ703714		Unknown	T2b7a1	35
JQ704736		Unknown	J1c5a1	35	JQ705882	Europe	England	T2b25	35
JQ702386		Unknown	J1c5a1	35	JQ704697	Europe	Romania	T2b25	35
JQ705031		Unknown	J1c5	35	JN024625	Europe	Ireland	T2b30	36
JQ703572		Unknown	J1c5	35	HQ399469		USA	T2b30	36
JQ705581		Unknown	J1c5	35	JQ704319	Europe	England	T2b	35
EU007859		Yakut	J1c5	48	JQ701986		Unknown	T2b	35
AY495212		USA	J1c5	37	JQ702239	Europe	United Kingdom	T2b	35
AY495199		USA	J1c5	37	JQ702645		Unknown	T2b	35
HM159445		Ashkenazi Jew	J1c5	36	JN004272	Europe	Germany	T2b	36
JQ703448	Europe	England	J1c5	35	JN419195	Europe	France	T2b	36
AY495209		USA	J1c6	37	JF980711		Unknown	T2b	36
JQ703919	Europe	Italy	J1c6	35	JN106183	Europe	Germany	T2b	36
AY495196		USA	J1c6	37	JQ798077	Europe	Estonia	T2b16	34
EU073970		Unknown	J1c6	36	EU007870	Europe	Kazakhstan Rep. (Russia)	T2b16	48
JQ705489		Unknown	J1c7	35	GU123025	Europe	Russia	T2b16	20
JQ797847	Europe	Greece	J1c7	34	JN126047	Europe	Germany	T2b	36
JQ797848	Europe	Greece	J1c7	34	GU122998	Europe	Russia	T2b	20
JQ797849	Europe	Romania	J1c7	34	JQ705351	Europe	Scotland	T2b	35

JQ797850	Europe	Romania	J1c7	34	JQ704505	Europe	Germany	T2b	35
EF583178	Europe	Slovakia	J1c7	40	JQ703722		Unknown	T2b	35
EU547187	Europe	Romania	J1c7	36	JQ703680	Europe	Scotland	T2b	35
EF459669		Unknown	J1c7	36	JQ702894	Europe	Finland	T2b	35
JQ797851	Europe	Greece	J1c7	34	JQ703728		Unknown	T2b	35
JQ797852	Europe	Bosnia and Herzegovina	J1c7	34	JQ705446	Europe	Ireland	T2b	35
JQ797853	Europe	Romania	J1c7	34	JQ705758		Unknown	T2b	35
JQ797854	Europe	Greece	J1c7	34	JQ705854	Europe	Norway	T2b	35
EU284668		Unknown	J1c7b	36	JF929201		Unknown	T2b24	36
JQ703516		Unknown	J1c7b	35	JQ705038	Europe	United Kingdom	T2b24	35
JQ704581		Unknown	J1c7b	35	JF946696		Unknown	T2b24a	36
FJ348222	Europe	Italy	J1c7b	16	JQ702547		Unknown	T2b24a	35
JQ702088		Unknown	J1c7b	35	JQ798090	South Asia	India	T2c	34
JQ702208	Europe	England	J1c7b	35	EF556164		Unknown	T2c1a1	1
JQ797861	Europe	Italy	J1c7	34	EF556160		Unknown	T2c1a1	1
JQ702609		Unknown	J1c7c	35	JQ703030	Near East	Iraq	T2c1a1	35
JQ703584		Unknown	J1c7c	35	EF660941	Europe	Italy	T2c1a2	10
JQ797860	Europe	Italy	J1c7	34	HM765471		Unknown	T2c1a	11
EU151466	Europe	Spain	J1c7a1	49	JN202494		Unknown	T2c1a	36
JQ703861	Europe	England	J1c7a1	35	JQ703691		Unknown	T2c1a	35
JQ701821		Unknown	J1c7a1	35	JQ701934		Unknown	T2c1a	35
JQ703682		Unknown	J1c7a1	35	JQ798091	Europe	Italy	T2c1b	34
JQ797855	Europe	Belarus	J1c7a1	34	JQ703823	Europe	Belgium	T2c1b2	35
	Europe	Italy	J1c7a1a	34	JQ705571	Europe	Germany	T2c1b2	35
FJ348162		USA	J1c7a1a	16	JF951723		Unknown	T2c1b1a	36
FJ348163		USA	J1c7a1a	16	JN252308		Unknown	T2c1b1a	36
FJ348164		USA	J1c7a1a	16	JN580589		Unknown	T2c1b1a1	36
FJ348165		USA	J1c7a1a	16	JQ705668		Unknown	T2c1b1a1	35
JQ705332		Unknown	J1c7a1a	35	JQ798092	Europe	Spain	T2c1b1a1	34
JQ703562	Europe	Sweden	J1c7a1a	35	JQ704020		Unknown	T2c1b1a	35
FJ348161		USA	J1c7a1a	16	JQ798093	Arabian Peninsula	Kuwait	T2c1b1a	34
JQ702929	Europe	Germany	J1c7a1a	35	JF833037		Unknown	T2c1b1	36
JQ703932		Unknown	J1c7a1a	35	JF892552		Unknown	T2c1b1	36
JQ701859		Unknown	J1c7a1a	35	JQ798094	Europe	Spain	T2c1b1	34

JQ702243	Europe	Hungary	J1c7a1a	35	JQ702928		Unknown	T2c1b1	35
HM627319		Ashkenazi Jew	J1c7a1a	36	JQ798095	Europe	Italy	T2c1b1b1	34
JQ705271	Europe	England	J1c7a1a	35	DQ523629	Europe	Sardinia	T2c1b1b1	15
JQ705072		Unknown	J1c7a1a	35	DQ523633	Europe	Sardinia	T2c1b1b1	15
JF812166		Ashkenazi Jew	J1c7a1a	36	DQ523667	Europe	Sardinia	T2c1b1b1	15
JQ797856	Europe	Belarus	J1c7a1a	34	JQ798096	Near East	Israel	T2c1b1b	34
JQ797857	Europe	Belarus	J1c7a1a1	34	GU048747		Unknown	T2c1b3	36
JQ705094		Unknown	J1c7a1a1	35	JQ798097	Europe	Czech Republic	T2c1b3	34
JQ705322	Europe	Norway	J1c7a1a1	35	JQ798098	Near East	Iran	T2c1b	34
JQ797858	Europe	Sweden	J1c7a1a1	34	JQ798099	Europe	Italy	T2c1c1	34
JQ705394	Europe	Norway	J1c7a1a1	35	JQ798100	Near East	Israel	T2c1c1	34
JQ703583	Europe	Finland	J1c7a1a1	35	JQ798101	Europe	Italy	T2c1c	34
JQ704103		Unknown	J1c7a1a1	35	JQ798103	Arabian Peninsula	Dubai	T2c1c	34
JQ702857		Unknown	J1c7a1a1	35	JQ798102		Crete	T2c1c2	34
JQ702338		Unknown	J1c7a1a1	35	JQ702795	Europe	Finland	T2c1c2	35
JQ797859	Europe	Ukraine	J1c7a1a1	34	JQ798105		Siberia	T2d1b1	34
HM171294		USA	J1c7a1a	36	JQ798106		Siberia	T2d1b1	34
HQ876599	Europe	Spain	J1c7	36	JQ798107		Siberia	T2d1b1	34
JN966735	Europe	Italy	J1c7	36	HM765473		Unknown	T2d1b1	11
JQ702486		Unknown	J1c7	35	DQ437577		Mongolia	T2d1b	56
JQ702176	Europe	Austria	J1c7d	35	AY714037	South Asia	India	T2d1a	29
JQ705424	Europe	Germany	J1c7d	35	JQ703754	Europe	Scotland	T2d1a	35
JQ705811		Unknown	J1c7d	35	JQ798104	South Asia	India	T2d1a	34
JQ705683		Unknown	J1c7d	35	JQ798108	Near East	Iran	T2d2	34
JQ704569	Europe	Poland	J1c7d	35	JQ798109	Europe	Italy	T2d2	34
JQ702827		Unknown	J1c7d	35	HM852899	South Caucasus	Georgia	T2d2	3
JQ705891		Ashkenazi Jew	J1c7d	35	EF177410	Europe	Portugal	T2e	13
JQ797862	Arabian Peninsula	Kuwait	J1c7e	34	GU565218		Unknown	T2e	36
FJ348155		USA	J1c7e	16	JQ705673		Unknown	T2e	35
FJ348154		USA	J1c7e	16	JQ702954	Europe	Italy	T2e	35
FJ348153		USA	J1c7e	16	JQ701853		Unknown	T2e	35
JQ797863	Europe	Greece	J1c8	34	EF060363	Europe	Italy	T2e2a	57
JQ702228	Europe	United Kingdom	J1c8	35	JQ798114	Near East	Iraq	T2e2a	34

JQ702868		Unknown	J1c8	35	AY714029	South Asia	India	T2e2	29
JQ797864		Unknown	J1c8b	34	EU597536	Europe	Czechoslovakia	T2e	27
GU906781		Unknown	J1c8b	36	HM852862	Anatolia	Turkey	T2e	3
JQ797865	South Asia	India	J1c8	34	JQ798115	Europe	Italy	T2e	34
JQ797866	Europe	Sweden	J1c8	34	JQ798116	Europe	Italy	T2e	34
JQ704604	Europe	England	J1c8	35	JQ703777	Europe	Croatia	T2e	35
JQ702876	Europe	Ireland	J1c8	35	JQ798111	Europe	Greece	T2e3	34
JQ797867	Europe	Greece	J1c8	34	JQ798112	Europe	Greece	T2e3	34
JQ705860	Europe	England	J1c8	35	EU703624	Europe	Italy	T2e3	36
JQ797868	Europe	England	J1c8	34	JN030346		Unknown	T2e4	36
JQ704891		Unknown	J1c8	35	JN828512		USA	T2e4	36
JQ702068		Unknown	J1c8a	35	JN819272		Sephardic Jew	T2e4	36
JQ705760	Europe	Germany	J1c8a	35	AF381985	North Africa	Morocco	T2e4	5
JQ797869		Unknown	J1c8a	34	JQ705148		Unknown	T2e+16296!	35
JQ702554		Unknown	J1c8a	35	JQ798110	Europe	Italy	T2e+16296!	34
JQ702049		Unknown	J1c8a	35	EU700086		Unknown	T2e+16296!	36
JQ704576	Europe	Ireland	J1c8	35	EF556188	Europe	Bulgaria Jew	T2e1	1
JQ703402	Europe	Ireland	J1c8	35	JQ705465	Europe	Ireland	T2e1	35
JQ797870	Europe	Italy	J1c	34	FJ178379	Europe	Italy	T2e1	38
GU123016	Europe	Russia	J1c12	20	JQ702286		Unknown	T2e1	35
JQ797875	Near East	Iraq	J1c12a	34	JQ702731	Europe	Belgium	T2e1	35
JQ797876	Near East	Israel	J1c12a	34	EU258890		Unknown	T2e1	36
JQ701870		Unknown	J1c12	35	JQ798113	Europe	Italy	T2e1	34
JQ797879	Europe	Italy	J1c12	34	GU944474		USA	T2e1	36
JQ797877	Arabian Peninsula	Kuwait	J1c12b	34	JF893457		Unknown	T2e1	36
HM852793	South Caucasus	Azerbaijan	J1c12b1	3	JF903810		Unknown	T2e1	36
JQ797878	Near East	Iraq	J1c12b1	34	JF831146	Europe	England	T2e1	36
JQ797881	South Caucasus	Azerbaijan	J1c16	34	JQ702210		Unknown	T2e	35
JQ705083		Unknown	J1c16	35	JQ701985		Unknown	T2e5	35
JQ704306	Europe	England	J1c16	35	JQ702891		Unknown	T2e5	35
JQ797872	Europe	Italy	J1c10a	34	JF965448		Unknown	T2e	36
JQ797873	North Africa	Morocco	J1c10a	34	JQ703661	Europe	England	T2e	35
JQ704751	Europe	England	J1c10	35	JQ704818	Europe	England	T2e	35
EF660915	Europe	Italy	J1c10	10	JF921152		Unknown	T2e	36
JQ797874	Europe	Italy	J1c11a	34	JF831421		Unknown	T2e	36

HM765470	Europe	Italy	J1c11a	11	JF937679		Unknown	T2e	36
GU592047	Europe	Austria	J1c11a	42	JQ798117	Arabian Peninsula	Kuwait	T2e	34
GU592032	Europe	Austria	J1c11a	42	JQ798118	Europe	Italy	T2e6	34
JQ702967	Europe	Germany	J1c11	35	JQ798119	Europe	Italy	T2e6	34
JQ704937	Europe	Scotland	J1c	35	AF346982	South Caucasus	Georgia	T2e	44
JQ797880	Europe	Italy	J1c17	34	JQ703702	Europe	Germany	T2f+195	35
JQ706022		Unknown	J1c17	35	JQ798120	Europe	Ukraine	T2f1	34
JQ797871	Europe	Italy	J1c9	34	JQ798121	Europe	England	T2f1a1a	34
JQ703686		Unknown	J1c9	35	JF940522	Europe	Sweden	T2f1a1a	36
AY495205		USA	J1c9	37	JQ798122	Europe	Italy	T2f1a1a	34
JQ705977	Europe	United Kingdom	J1c9	35	JQ702642		Unknown	T2f1a1a	35
JQ797883	North Africa	Morocco	J1c	34	JQ705857		Unknown	T2f1a1a	35
JQ797846	South Caucasus	Armenia	J1c	34	JQ798123	Europe	Denmark	T2f1a1a	34
JQ797882	Europe	Greece	J1c	34	JQ703663		Unknown	T2f1a1a	35
JQ705449		Unknown	J1c	35	JQ705015	Europe	Denmark	T2f1a1a	35
JQ701992		Unknown	J1c	35	JQ705023		Unknown	T2f1a1a	35
AY495206		USA	J1c	37	JQ705644		Unknown	T2f1a1a	35
JQ797884	Europe	Greece	J1c19	34	JQ703806		Unknown	T2f1a1a	35
HM775495	South Caucasus	Armenia	J1c19	36	JQ705892		Unknown	T2f1a1a	35
JQ797885	Europe	Dagestan Rep. (Russia)	J1	34	JQ705879		Unknown	T2f1a1a	35
JQ797886	Europe	Italy	J1d1a	34	JQ705897	Europe	Bosnia and Herzegovina	T2f1a1a	35
JQ797887	Arabian Peninsula	Kuwait	J1d1a	34	JQ705927		Unknown	T2f1a1a	35
JQ797888	Arabian Peninsula	Kuwait	J1d1a1	34	JQ704807		Unknown	T2f1a1a	35
JQ797889	Near East	Iraq	J1d1a1	34	GU932663		Unknown	T2f1a1a	36
JQ705319		Unknown	J1d1a1	35	JF905568		Unknown	T2f1a1a	36
JQ797890	East Africa	Uganda	J1d1a1a	34	JF906520		Unknown	T2f1a1a	36
JQ797891	East Africa	Uganda	J1d1a1a	34	JF926842		Unknown	T2f1a1a	36
EF184636	East Africa	Tanzania	J1d1a1	50	JN203494		Unknown	T2f1a1a	36
JF292900	South Caucasus	Armenia	J1d1a	36	JN887353	Europe	Sweden	T2f1a1a	36
JQ797892	Europe	Cyprus	J1d1b1	34	JQ702350		Unknown	T2f1a1	35

HM852780	South Caucasus	Armenia	J1d1b1	3	JN003847		Unknown	T2f1a1	36
AF382001	Europe	Iberia	J1d1b	5	JQ702975		Unknown	T2f1a1	35
JQ797893	Near East	Iran	J1d2a1	34	GU123028	Europe	Russia	T2f1a	20
HQ325739	South Caucasus	Armenia	J1d2a1	36	HQ287877	Europe	Ireland	T2f1a	47
DQ341087		Brazil-Italy	J1d2a	46	JN084792		Unknown	T2f	36
DQ341088		Brazil-Italy	J1d2a	46	JQ705528		Unknown	T2f4	35
JQ797902	Near East	Iraq	J1d	34	JN034044		Canada	T2f4	36
JQ797900	South Asia	India	J1d2d	34	JQ798124	Europe	Italy	T2f2	34
HM453206	South Caucasus	Armenia	J1d2d	36	JQ798125	Europe	Greece	T2f2	34
JQ797894	North Ossetia	Alania Rep. (Russia)	J1d2b	34	HQ286590	South Caucasus	Armenia	T2f2	36
EU007880		Mansi	J1d2b1	48	JQ702294		Unknown	T2f5	35
JQ797895		Siberia	J1d2b1	34	JQ702957		Unknown	T2f5	35
GU122987	Europe	Russia	J1d2b	20	JQ619780	Europe	Finland	T2f	36
JQ797896	South Caucasus	Georgia	J1d2c1	34	JQ703997	Europe	Italy	T2f	35
JQ797897	Europe	Adygea Rep. (Russia)	J1d2c1	34	JQ703697		Unknown	T2f	35
JQ797898	North Ossetia	Alania Rep. (Russia)	J1d2c1	34	JF960209		Unknown	T2f3	36
JQ797899	Europe	Italy	J1d2c	34	JQ705221		Unknown	T2f3	35
JQ704809		Unknown	J1d2c2	35	JQ705639		Unknown	T2f3	35
EU597552	South Asia	Pakistan	J1d2c2	27	JQ704426	Europe	United Kingdom	T2f3	35
JQ797901	Near East	Iran	J1d2e	34	JQ704129	Europe	Ireland	T2f	35
JQ703915	Anatolia	Turkey	J1d2e	35	JQ798129	Europe	Italy	T2i	34
HM852829	Near East	Iran	J1d2e	3	HM852881	South Caucasus	Georgia	T2i	3
JQ704041	Europe	Ireland	J2a1	35	JQ798130	Near East	Iraq	T2i1	34
JQ797903	Europe	Italy	J2a1a1	34	JQ704055		Unknown	T2i1	35
HQ699438	Europe	Sweden	J2a1a1	36	JQ798131	Near East	Iran	T2i	34
JQ705358		Unknown	J2a1a1	35	JQ702108	Europe	Germany	T2g2	35
JQ702216		Unknown	J2a1a1	35	JQ702574		Unknown	T2g2	35
JQ702091		Unknown	J2a1a1c	35	JQ702710		Unknown	T2g2	35
JQ705512		Unknown	J2a1a1c	35	JQ798126	Europe	Italy	T2g1a	34
JQ797904	Europe	France	J2a1a1c	34	JN086657	Europe	Lithuanian Jew	T2g1a1	36
JQ703552	Europe	British Isles	J2a1a1c	35	JQ798127	Europe	Italy	T2g1a1a	34
JQ797911	Europe	Italy	J2a1a1	34	EF556186	Near East	Iran Jew	T2g1a1a1	1

JQ797905	Europe	Italy	J2a1a1d	34	JQ705079	Near East	Iraq	T2g1a1a1	35
JQ797906	Europe	Italy	J2a1a1d	34	JQ705097		Unknown	T2g1a1a1	35
EU931680		USA	J2a1a1d	36	JQ703123	Europe	Hungary	T2g1a1a1	35
JQ797912	Europe	Italy	J2a1a1d	34	EU007862		Yakut	T2g1a1a	48
JQ702950	Europe	British Isles	J2a1a1	35	EU935442	North Africa	Egypt	T2g1a1a	4
JQ797907	Europe	Italy	J2a1a1a2	34	EU935446	North Africa	Egypt	T2g1a1a	4
JQ703638	Europe	England	J2a1a1a2	35	EU935448	North Africa	Egypt	T2g1a1a	4
JQ702622		Unknown	J2a1a1a2	35	JQ798132	South Asia	India	T2	34
JQ702772		Unknown	J2a1a1a2	35	JQ798133	South Caucasus	Azerbaijan	T2	34
JQ703330		Unknown	J2a1a1a2	35	JN037469		USA	T2h	36
JQ703524		Unknown	J2a1a1a2	35	JQ798128	Europe	Italy	T2h	34
JQ705214	Europe	England	J2a1a1a2	35	JQ702170	Europe	Slovakia	T2h	35
JQ703570	Europe	Czechoslovakia	J2a1a1a2	35	JN202724	Europe	Sicily	T2h	36
JQ703566		Unknown	J2a1a1a2	35	JQ705018		Unknown	T2h	35
JQ703673	Europe	Sweden	J2a1a1a2	35	HM852818	Near East	Iran	T2h	3
JQ704906		Unknown	J2a1a1a2	35	HM852796	South Caucasus	Azerbaijan	T2h	3
JQ705625	Europe	Netherlands	J2a1a1a2	35	JQ705300	Europe	Poland	T2	35
JQ705631	Europe	Germany	J2a1a1a2	35	JF707633		USA	T2	36
JQ705632		Unknown	J2a1a1a2	35	EF660960	Europe	Italy	T2+16296!	10
JQ705535		Unknown	J2a1a1a2	35	JF937088		Unknown	T2+16296!	36
HQ436101		Unknown	J2a1a1a2	36	JQ704762	Europe	England	T2+16296!	35
FJ348157		USA	J2a1a1a2	16	JQ702704		Unknown	T2+16296!	35
HQ104930		USA	J2a1a1a2	36	GQ304741	Europe	Italy	T2+16296!	53
FJ194438	Europe	Sweden	J2a1a1a2	36	JQ704995	Europe	United Kingdom	T2+16296!	35
EU721734		USA	J2a1a1a2	36	JQ798139	Europe	England	T2+16296!	34
JF343122		USA	J2a1a1a2	36	JQ798136	Europe	Greece	T2m	34
HQ424014		USA	J2a1a1a2	36	JQ798134	Europe	Italy	T2+16296!	34
JQ702027		Unknown	J2a1a1a2	35	HQ912960	Europe	Romania	T2l	36
GU903270		Unknown	J2a1a1a	36	JQ798138	Europe	Adygea Rep. (Russia)	T2l	34
HM485691		USA	J2a1a1a	36	JQ798137	Europe	Adygea Rep. (Russia)	T2	34
JQ705575		Unknown	J2a1a1a	35	EF177442	Europe	Portugal	T2j	13
JQ704947	Europe	England	J2a1a1a	35	HQ877824	Europe	Austria	T2j	36

JQ704600		Unknown	J2a1a1a	35	JQ704786	Europe	Italy	T2j	35
JQ703801	Europe	Czech Republic	J2a1a1a	35	HM852810	South Caucasus	Azerbaijan	T2k	3
JQ702561		Unknown	J2a1a1a	35	JQ705690	Europe	Ireland	T2k	35
JQ703568	Europe	Germany	J2a1a1a3	35	JQ798135	Europe	Crete	T2	34
JQ705323		Unknown	J2a1a1a3	35	HM852766	South Caucasus	Armenia	T2	3
GU123013	Europe	Russia	J2a1a1a	20	ADY17	Europe	Adygea Rep. (Russia)	T2i	58
JQ797908	Europe	Lithuania	J2a1a1e	34	ADY18	Europe	Adygea Rep. (Russia)	J1c3k	58
JQ797909	Near East	Israel	J2a1a1e	34	ADY20	Europe	Adygea Rep. (Russia)	T2	58
JQ797910	Europe	Greece	J2a1a1e	34	ADY29	Europe	Adygea Rep. (Russia)	T2a3	58
JQ705774	Europe	Poland	J2a1a1e	35	ADY38	Europe	Adygea Rep. (Russia)	T1a1a1	58
JQ702605		Unknown	J2a1a1e	35	ADY41	Europe	Adygea Rep. (Russia)	J1d2c1	58
JQ701807		Unknown	J2a1a1	35	ADY46	Europe	Adygea Rep. (Russia)	T2b	58
JQ797913	Europe	Italy	J2a1a1	34	ADY5	Europe	Adygea Rep. (Russia)	T2a1b2b	58
FJ605154		USA	J2a1a1b	36	MZB17	North Africa	Algeria	J2a2d2	58
JQ705042		Unknown	J2a1a1b	35	MZB24	North Africa	Algeria	J2a2d1	58
JQ705390		Unknown	J2a1a1b	35	MZB31	North Africa	Algeria	T1a3	58
DQ523640	Europe	Sardinia	J2a1a1	15	AZB11	South Caucasus	Azerbaijan	T2a1b1a1c	58
DQ341089	Europe	Italy	J2a1a2	46	AZB15	South Caucasus	Azerbaijan	t1a1a1	58
JQ797914	Europe	Greece	J2a1a2a	34	AZB8	South Caucasus	Azerbaijan	T2b	58
JQ702364	Europe	Italy	J2a1a2a	35	AZB9	South Caucasus	Azerbaijan	T2f4	58
JQ797915	Arabian Peninsula	Yemen	J2a2a1a	34	B5	Europe	Bulgaria	J1c2a6	58

JQ797916	Arabian Peninsula	Yemen	J2a2a1a	34	B13	Europe	Bulgaria	T1a1a1	58
JQ797917	Arabian Peninsula	Kuwait	J2a2a1a1a	34	B2	Europe	Bulgaria	J1c2a5	58
JQ797918	Near East	Iraq	J2a2a1a1a	34	B8	Europe	Bulgaria	J2c2a5a	58
JQ797919	Near East	Iraq	J2a2a1a1a	34	BUL7096	Europe	Bulgaria	J1c8a	58
JQ797920	Near East	Israel	J2a2a1a1	34	BUL7102	Europe	Bulgaria	T1a15	58
JQ797921	Arabian Peninsula	Kuwait	J2a2a1a	34	BUL7107	Europe	Bulgaria	T2+16296!	58
JQ797922	Europe	England	J2a2a	34	BUL7116	Europe	Bulgaria	J1c4	58
JQ797923	Europe	Italy	J2a2a2	34	BUL-HJB3	Europe	Bulgaria	J1c7	58
GU065327		Unknown	J2a2a2	51	SE1013	Europe	Canary Islands	J1b1a1	58
EF660967	Europe	Italy	J2a2a	10	SE145	Europe	Canary Islands	J2a2b1b	58
JQ797924	Europe	Italy	J2a2	34	SE155	Europe	Canary Islands	T2c1b1	58
JQ797925	Europe	Italy	J2a2c	34	CHE9	Europe	Chechen Rep. (Russia)	T2h	58
JQ797926	Europe	Italy	J2a2c1a	34	CHV44	Europe	Chuvashia Rep. (Russia)	T2b4	58
JQ797927	Europe	Italy	J2a2c1a	34	CZE377	Europe	Czech Republic	J1b1a3	58
JQ797928	Arabian Peninsula	Yemen	J2a2c1b	34	CZE235	Europe	Czech Republic	J1c2a3b1	58
JQ797929	North Africa	Morocco	J2a2b1a	34	CZE236	Europe	Czech Republic	J1c19	58
FJ460543	North Africa	Tunisia	J2a2b1a	6	CZE238	Europe	Czech Republic	T2b4	58
JQ797930	North Africa	Morocco	J2a2b1c	34	CZE245	Europe	Czech Republic	J1c3k	58
JQ797931	Europe	Greece	J2a2b2	34	CZE254	Europe	Czech Republic	T1a1a1i	58
JQ703605	Europe	United Kingdom	J2a2b2	35	CZE257	Europe	Czech Republic	J1c7a1a	58
JQ797932	Europe	Sardinia	J2a2b	34	CZE268	Europe	Czech Republic	T1a1a1g	58
JQ797933	Arabian Peninsula	Yemen	J2a2b	34	CZE316	Europe	Czech Republic	T2c1b3	58
JQ797934	Europe	Canary Islands	J2a2d1	34	CZE373	Europe	Czech Republic	J2a1a1	58
FJ460559	North Africa	Tunisia	J2a2d1	6	DRZ13	Near East	Israel-Druze	J2a2a1a1	58
JQ797935	North Africa	Algeria	J2a2d2	34	DRZ16	Near East	Israel-Druze	T1a1e	58
GQ249257		USA	J2b1a2	36	DRZ42	Near East	Israel-Druze	J2a1a1e	58
JQ797936		Unknown	J2b1a2a	34	DRZ6	Near East	Israel-Druze	J1c12a	58
JF938916	Europe	Portugal	J2b1a2a	36	DUB25	Arabian Peninsula	Dubai	J1b2	58

JQ797937	Europe	Italy	J2b1a2	34	DUB22	Arabian Peninsula	Dubai	J2b1	58
JQ703587		Unknown	J2b1a2	35	DUB34	Arabian Peninsula	Dubai	J1d2c2	58
JQ797938	Europe	Italy	J2b1a2	34	DUB39	Arabian Peninsula	Dubai	T2	58
JQ705677		Unknown	J2b1a2	35	DUB7	Arabian Peninsula	Dubai	J1b6b	58
JQ705925		Unknown	J2b1a	35	DUB83	Arabian Peninsula	Dubai	J1b2	58
JQ702593	Europe	England	J2b1a	35	DUB84	Arabian Peninsula	Dubai	T1a4b	58
EU673448		Unknown	J2b1a	36	CE70	Europe	England	J1b1a1e	58
JQ797939	Europe	Italy	J2b1a	34	NE665	Europe	England	J1b1a1a	58
JQ702199		Unknown	J2b1a	35	NE707	Europe	England	J1b1a1a	58
JQ797940	Europe	Italy	J2b1a3	34	NW119	Europe	England	J1b1a1	58
JQ702702		Unknown	J2b1a3	35	NW644	Europe	England	J1b1a1	58
JQ702858	Europe	Italy	J2b1a3	35	NW721	Europe	England	J1b1a1	58
JQ797941	Europe	Greece	J2b1a3	34	SC207	Europe	England	J1b1a1	58
AY195778		Unknown	J2b1a3	23	SE354	Europe	England	J1b1a1a	58
JQ797942	North Africa	Morocco	J2b1a	34	SE733	Europe	England	J1b1a1	58
JQ797943	Europe	Russia	J2b1a6	34	SE870	Europe	England	J1b1a1a	58
JQ704523	Europe	Poland	J2b1a6	35	WA64	Europe	England	J1b1a1a	58
JQ797944	Europe	Moldova	J2b1a6	34	FRA-CE23	Europe	France	T1a1a1e	58
DQ523671	Europe	Sardinia	J2b1a7	15	FRA-NF258	Europe	France	J2a1a1a	58
EU597520	Europe	Sardinia	J2b1a7	27	FRA-NF263	Europe	France	T2b2b1	58
JQ797945	Europe	Italy	J2b1a5	34	FRA-NF298	Europe	France	J1c1	58
JQ797946	Europe	Italy	J2b1a5	34	FRA-NF331	Europe	France	T2b	58
DQ523653	Europe	Sardinia	J2b1a5	15	FRA-NF345	Europe	France	J1c3f	58
JF915700	Europe	Ireland	J2b1a	36	FRA-NF351	Europe	France	T2f1a1a	58
JQ702346		Unknown	J2b1a4	35	FRA-NF66	Europe	France	T2f1a1a	58
JQ705466		Unknown	J2b1a4	35	HEL-T062	Europe	Greece	T2a1b	58
JQ703538		Unknown	J2b1a4	35	HEL-T066	Europe	Greece	T1b3	58
JQ703515		Unknown	J2b1a4	35	IRL42	Europe	Ireland	T2b	58
JQ702196		Unknown	J2b1a	35	IRL7	Europe	Ireland	J1c1b	58
JQ702758		Unknown	J2b1a	35	IRL868	Europe	Ireland	J2c2a2a	58

JQ702477	Europe	England	J2b1a	35	KAB12	Europe	Kabardino-Balkaria Rep. (Russia)	J1b1a1	58
DQ341090	Europe	Italy	J2b1a	46	KAB31	Europe	Kabardino-Balkaria Rep. (Russia)	J1d2c1	58
JQ797947	Europe	Italy	J2b1a	34	KUR71		Kurd	T1b3	58
JQ705623	Europe	England	J2b1a	35	J14	North Africa	Morocco	T2c1b1a	58
JQ702442	Europe	United Kingdom	J2b1a	35	J41	North Africa	Morocco	J1c3i	58
JQ702553		Unknown	J2b1a	35	J5	North Africa	Morocco	T1a6	58
JQ703517		Unknown	J2b1a	35	J50	North Africa	Morocco	J1d1a1	58
JQ705021	Europe	Italy	J2b1a	35	J54	North Africa	Morocco	J1c9	58
FJ445408	Europe	Ireland	J2b1a1a	36	J58	North Africa	Morocco	J2b1	58
FJ380056		Unknown	J2b1a1a	36	J60	North Africa	Morocco	J2b1a	58
JQ702863		Unknown	J2b1a1a	35	J76	North Africa	Morocco	J2b1a	58
JQ705356		Unknown	J2b1a1	35	NOS30	North Ossetia	Alania Rep. (Russia)	T1a1a1	58
JQ703580	Europe	France	J2b1a1	35	NOS51	North Ossetia	Alania Rep. (Russia)	J2b1	58
JQ705064	Europe	Finland	J2b1	35	NOS9	North Ossetia	Alania Rep. (Russia)	T2b24	58
JQ702424		Unknown	J2b1	35	PAL-AO1036	Near East	Palestine	J1c12	58
JQ797949	Near east	Syria	J2b1	34	PAL-AO1056	Near East	Palestine	J1b6b	58
EU862198	Europe	Ireland	J2b1b	36	PAL-AO530	Near East	Palestine	J1b1b2	58
JQ703784	Europe	Netherlands	J2b1b	35	PAL-AO549	Near East	Palestine	T2g1a1a	58
JQ702042		Unknown	J2b1b	35	PAL-AO593	Near East	Palestine	T1b	58
JQ797950	Europe	Greece	J2b1c	34	PAL-AO647	Near East	Palestine	T2a1a	58
JQ797951	Europe	Greece	J2b1c	34	PAL-AO672	Near East	Palestine	T2g1	58
JQ797952	Europe	Greece	J2b1c	34	PAL-AO783	Near East	Palestine	J1c1	58
EU807741		Siberia	J2b1c1	52	PAL-AO841	Near East	Palestine	J1b1a3	58
JQ702563		Unknown	J2b1	35	PAL-AO947	Near East	Palestine	T2a1b1a	58
JQ797953	Europe	Italy	J2b1	34	X0000	Europe	Portugal	J2a2d2	58
JQ797954	Europe	Italy	J2b1	34	X1384	Europe	Portugal	J1c3f	58
JQ797955	Europe	Italy	J2b1	34	X2342	Europe	Portugal	T2a1a	58
JQ797956	Europe	Italy	J2b1	34	X2881	Europe	Portugal	T2b4f	58
JQ797957	Europe	Italy	J2b1d1	34	X3305	Europe	Portugal	T2c1a2	58

JQ797958	Europe	Russia	J2b1d1	34	X3368	Europe	Portugal	J1c3	58
JQ797959	Near East	Syria	J2b1d1	34	X3590	Europe	Portugal	J1c3	58
JQ797960	Europe	Italy	J2b1d	34	X369	Europe	Portugal	T2e	58
FJ445409		USA	J2b1d	36	X3691	Europe	Portugal	J1b3a	58
JQ797961	North Africa	Algeria	J2b1	34	X4059	Europe	Portugal	J1c1f	58
JQ797962	Europe	Italy	J2b1	34	X4062	Europe	Portugal	T2b	58
JQ797963	Europe	Slovakia	J2b1	34	X4064	Europe	Portugal	T2b3+151	58
JQ702459		Unknown	J2b1	35	X4084	Europe	Portugal	J1c7a	58
JQ702488		Unknown	J2b1	35	X4118	Europe	Portugal	T2b3+151	58
JQ701981		Unknown	J2b1e	35	X4229	Europe	Portugal	T2a1b1a	58
JQ797964	Europe	Italy	J2b1e1	34	X4381	Europe	Portugal	J1c8	58
JQ797965	Near East	Lebanon	J2b1e1	34	X4644	Europe	Portugal	J1c1	58
JQ797966	North Ossetia	Alania Rep. (Russia)	J2b1	34	X4646	Europe	Portugal	T2b3+151	58
JQ797967	Europe	Kabardino-Balkaria Rep. (Russia)	J2b1	34	X5399	Europe	Portugal	T1a1a1	58
JQ797968	Near East	Syria	J2b1f	34	X5459	Europe	Portugal	J2b1c1	58
HQ727682	South Caucasus	Armenia	J2b1f	36	X5516	Europe	Portugal	T2e4	58
EF556152	South Caucasus	Azerbaijan Jew	J2b1f	1	X5577	Europe	Portugal	J1c1b	58
JQ797969	Anatolia	Turkey	J2b1f	34	X5591	Europe	Portugal	J2b1c1	58
JQ797970	Anatolia	Turkey	J2b1f	34	X5595	Europe	Portugal	T2e1	58
JQ797971	North Africa	Egypt	J2b1	34	X5679	Europe	Portugal	T1a1a1	58
JQ797948	Europe	Italy	J2b1	34	X5688	Europe	Portugal	T1a1c3	58
JQ797972	Europe	Romania	J2b2	34	X5703	Europe	Portugal	T1a1c3	58
JQ797973	Near East	Iran	J2b2	34	ROM-MA66	Europe	Romania	J1c7	58
JQ797974	Europe	Albania	J2b2	34	ROM-MA68	Europe	Romania	T1a1a1	58
FJ213765		USA	J2b2	36	ROM-MA7	Europe	Romania	T1a1a1	58
HQ889848		USA	J2b2	36	ROM-MA70	Europe	Romania	J1c5a	58
JQ797975	South Caucasus	Azerbaijan	T1	34	ROM-MA8	Europe	Romania	T1a15	58
JQ797976	Near East	Iraq	T1+16189	34	ROM-MA81	Europe	Romania	T2f2	58
JQ797977	North Africa	Morocco	T1a1a1	34	ROM-MA90	Europe	Romania	J1c2a7	58
JQ797978	Europe	Estonia	T1a1a1	34	ROM-VR203	Europe	Romania	T1a1a1b	58
JQ797979	Europe	Sweden	T1a1a1	34	ROM-VR210	Europe	Romania	J1c7	58

JQ797980	Europe	Kabardino-Balkaria Rep. (Russia)	T1a1a1	34	ROM-VR224	Europe	Romania	T2b	58
JQ797981	Anatolia	Turkey	T1a1a1	34	ROM-VR230	Europe	Romania	J1c3k	58
AY495293		USA	T1a1a1	37	ROM-VR236	Europe	Romania	T1a15	58
AY495289		USA	T1a1a1	37	ROM-VR238	Europe	Romania	T2a1b1a1	58
AF382006	Europe	Iberia	T1a1a1	5	RUS7	Europe	Russia	T2b	58
HQ167734	Europe	Ukraine	T1a1a1	36	SAR-SS110	Europe	Sardinia	T2b	58
GU122980	Europe	Russia	T1a1a1	20	SAR-SS133	Europe	Sardinia	T2b	58
GU365880		USA	T1a1a1	36	SAR-SS56	Europe	Sardinia	J1c3	58
JN880467	Europe	Ireland	T1a1a1	36	SAR-SS64	Europe	Sardinia	J2b1a5	58
JQ704249	Europe	Belarus	T1a1a1	35	SAR-SS77	Europe	Sardinia	T2b25	58
JQ704735		Unknown	T1a1a1	35	SAR-SS88	Europe	Sardinia	T2i	58
JQ705463		Unknown	T1a1a1	35	SIC-TO509	Europe	Sicily	T2h	58
JQ705353		Unknown	T1a1a1	35	SIC-TO515	Europe	Sicily	T1a1a1	58
JQ702716	Europe	Ireland	T1a1a1	35	SIC-TO582	Europe	Sicily	J1c	58
JQ702556	Europe	England	T1a1a1	35	SIC-TO608	Europe	Sicily	T2h	58
JQ702340		Unknown	T1a1a1	35	SIC-TR129	Europe	Sicily	T2m	58
JQ797982	North Africa	Morocco	T1a1a1f	34	SIC-TR21	Europe	Sicily	T2m	58
JQ797983	Europe	Greece	T1a1a1f	34	SIC-TR26	Europe	Sicily	J2b1	58
JQ702643		Unknown	T1a1a1f	35	SIC-TR4	Europe	Sicily	T1a1a1	58
JQ797984	North Africa	Morocco	T1a1a1	34	SIC-TR51	Europe	Sicily	J1c2a	58
JQ797985	Europe	Italy	T1a1a1	34	SIC-TR56	Europe	Sicily	Jb9	58
JQ702173		Unknown	T1a1a1	35	SUD78	North Africa	Sudan	T2b7a3	58
JQ797986	Europe	Estonia	T1a1a1	34	SYR-SJS17	Near east	Syria	T2b	58
GQ304742		Unknown	T1a1a1	53	SYR-SJS20	Near east	Syria	J1d2	58
JQ705207		Unknown	T1a1a1	35	SYR-SJS8	Near east	Syria	T2i1	58
JQ701804		Unknown	T1a1a1	35	SYR-SJSb11	Near east	Syria	T2c1a	58
JQ797987	Europe	Italy	T1a1a1	34	SYR-SJSb2	Near east	Syria	T1b4	58
JQ797988	Europe	Kabardino-Balkaria Rep. (Russia)	T1a1a1	34	SYR-SJSb33	Near east	Syria	T2c1a	58
JQ797989	North Ossetia	Alania Rep. (Russia)	T1a1a1	34	SYR-SJSb41	Near east	Syria	T2f	58

JQ797990	Europe	Karachay-Cherkessia Rep. (Russia)	T1a1a1	34	AL21	North Africa	Tunisia	J2b1a	58
EU597578	Europe	French Basque	T1a1a1	27	AL23	North Africa	Tunisia	T2d2	58
HM184912		USA	T1a1a1b	36	AL34	North Africa	Tunisia	T2b	58
JQ797991	Anatolia	Turkey	T1a1a1b	34	AL35	North Africa	Tunisia	T1a7	58
JQ797992	Europe	Crete	T1a1a1b	34	AL43	North Africa	Tunisia	T1a	58
JQ702958	Europe	Ukraine	T1a1a1b	35	AL8	North Africa	Tunisia	T2m	58
DQ358975	Europe	Germany	T1a1a1b	43	ET42	North Africa	Tunisia	T1a	58
JQ797994	Europe	Italy	T1a1a1b1	34	K19	North Africa	Tunisia	T1a1c3	58
JQ797993	Near East	Iran	T1a1a1b2	34	K2	North Africa	Tunisia	J2a2d1	58
HM852798	South Caucasus	Azerbaijan	T1a1a1b3	3	QL18	North Africa	Tunisia	T2m	58
JQ797995	Near East	Iran	T1a1a1b4	34	R187	North Africa	Tunisia	T2b	58
AY714036	South Asia	India	T1a1a1b5	29	R582	North Africa	Tunisia	J1b1a1	58
JQ702784		Unknown	T1a1a1	35	SK15	North Africa	Tunisia	T1a6	58
JQ702853	Europe	Finland	T1a1a1	35	SK25	North Africa	Tunisia	J1c7	58
JQ797996	Anatolia	Turkey	T1a1a1	34	SL15	North Africa	Tunisia	T2b3	58
JQ703240		Unknown	T1a1a1	35	SL17	North Africa	Tunisia	T2a1b1a1b	58
JQ797997	Near East	Iraq	T1a1a1	34	SL20	North Africa	Tunisia	T1b	58
JQ703670		Unknown	T1a1a1	35	TR29	North Africa	Tunisia	J2a2b1b	58
JQ705778	Europe	Scotland	T1a1a1	35	TR35	North Africa	Tunisia	J1c7	58
JQ797998	Near East	Iran	T1a1a1	34	TR39	North Africa	Tunisia	T2c1b1	58
JQ797999	Near East	Iran	T1a1a1	34	TR40	North Africa	Tunisia	J2a2a1a1a	58
JQ798000	Near East	Iran	T1a1a1	34	TR47	North Africa	Tunisia	T1a6	58
JQ798001	South Asia	India	T1a1a1	34	TR48	North Africa	Tunisia	J1b2	58
JQ703739		Unknown	T1a1a1	35	TR49	North Africa	Tunisia	J2a2d1	58
JQ798002	South Asia	India	T1a1a1	34	TR7	North Africa	Tunisia	J1b2	58
JQ798003	Europe	Greece	T1a1a1l	34	Z12	North Africa	Tunisia	T2b	58
JQ798004	Europe	Greece	T1a1a1l	34	TUR-AT21	Anatolia	Turkey	J2b1f	58
JF979131		Unknown	T1a1a1l	36	TUR-AT26	Anatolia	Turkey	T2a1b1a1c	58
JQ798005		Siberia	T1a1a1d	34	TUR-AT27	Anatolia	Turkey	J1c3f	58
EU007876		Kchanti	T1a1a1d	48	TUR-AT31	Anatolia	Turkey	T2g	58
FJ348197	Europe	Italy	T1a1a1	16	TUR-AT33	Anatolia	Turkey	T2a1a	58
JQ798006	Europe	Sweden	T1a1a1e	34	TUR-AT51	Anatolia	Turkey	T2a3	58
FJ480957	Europe	Sweden	T1a1a1e	36	TUR-AT78	Anatolia	Turkey	T1a1a1	58
EF661000	Europe	Italy	T1a1a1	10	TUR-AT80	Anatolia	Turkey	T1a1c	58

AY714015	South Asia	India	T1a1a1	29	TUR-AT92	Anatolia	Turkey	T2c1c	58
AY495288		USA	T1a1a1	37	TUR-ATb10	Anatolia	Turkey	T2g1a1a	58
AY495297		USA	T1a1a1	37	TUR-ATb13	Anatolia	Turkey	T2b	58
AY495290		USA	T1a1a1	37	TUR-ATb18	Anatolia	Turkey	T2a3	58
EF177406	Europe	Portugal	T1a1a1	13	TUR-ATb28	Anatolia	Turkey	T2h	58
JQ702825		Unknown	T1a1a1	35	TUR-ATb29	Anatolia	Turkey	T1b3	58
AY495292		USA	T1a1a1	37	TUR-ATb3	Anatolia	Turkey	J1c2a5	58
JQ705266		Unknown	T1a1a1	35	TUR-ATb31	Anatolia	Turkey	T1a15	58
AY495291		USA	T1a1a1j	37	TUR-ATb36	Anatolia	Turkey	J1b1a1	58
JQ702925	Europe	Greece	T1a1a1j	35	TUR-ATc1	Anatolia	Turkey	T2c1a	58
JQ703446		Unknown	T1a1a1j	35	TUR-ATc11	Anatolia	Turkey	T1b3	58
EU367994		Unknown	T1a1a1	36	TUR-ATc2	Anatolia	Turkey	T2b4	58
AY495294		USA	T1a1a1a1	37	TUR-ATc8	Anatolia	Turkey	J1c2a	58
EU882063		USA	T1a1a1a1	36	YJ-R1	Arabian Peninsula	Yemen	J2a2a1a	58
JQ703032	Europe	Switzerland	T1a1a1a1	35	YJ-R12	Arabian Peninsula	Yemen	J1b2a	58
JQ705982		Unknown	T1a1a1a	35	YJ-R16	Arabian Peninsula	Yemen	J2a2c1	58
JQ703395	Europe	Scotland	T1a1a1	35	YJ-R2	Arabian Peninsula	Yemen	J2a2a1a	58
AY495295		USA	T1a1a1	37	YJ-R22	Arabian Peninsula	Yemen	J2a2a1a	58
GU944473		USA	T1a1a1	36	YJ-R25	Arabian Peninsula	Yemen	J1b3b	58
JQ705441	Europe	Czech Republic	T1a1a1i	35	JF905566	Europe	Ireland	T1a1a1c	36
JQ705542		Unknown	T1a1a1i	35	JQ705630		Unknown	T1a1a1c	35
JN021256		Unknown	T1a1a1i	36	JQ705856	Europe	Ireland	T1a1a1c	35
JN127414		Unknown	T1a1a1i	36	JQ704801	Europe	Germany	T1a1a1c	35
JF929200	Europe	England	T1a1a1	36	JQ703894		Unknown	T1a1a1c	35
JQ702988	Europe	British Isles	T1a1a1	35	JQ702923		Unknown	T1a1a1c	35
JQ702343		Unknown	T1a1a1	35	JF833038		Unknown	T1a1a1	36
JF905569		Unknown	T1a1a1	36	JQ702986	Europe	Scotland	T1a1a1	35
JF833040	Europe	British Isles	T1a1a1g	36	JQ704732	Europe	Switzerland	T1a1a1g	35
EU092802	Arabian Peninsula	Yemen	L6a	59	DQ341064	North Africa	Ethiopia	L4a1a	60

EU092803	Arabian Peninsula	Yemen	L6a	59	EU092938	North Africa	Ethiopia	L4b2a2a	59
EU092924	Arabian Peninsula	Yemen	L6a	59	EU092942	North Africa	Ethiopia	L4b2a1	59
EU092748	Arabian Peninsula	Saudi Arabia	L4a1a	59	EU092951	North Africa	Ethiopia	L4b2a2b	59
EU092799	Arabian Peninsula	Yemen	L4a2	59	DQ341065	North Africa	Ethiopia	L4b2a1	60
EU092800	Arabian Peninsula	Yemen	L4a2	59	EU092838	South Africa	South Africa; San	L4b2a2	59
EU092808	Arabian Peninsula	Yemen	L4b1	59	JQ044811	West Africa	Burkina Faso	L4b1	61
EU092750	Arabian Peninsula	Saudi Arabia	L4b2a2	59	JQ044834	West Africa	Burkina Faso	L4b1	61
EU092780	Arabian Peninsula	Kuwait	L4b2a2b	59	JQ044848	West Africa	Burkina Faso	L4b1	61
EU092743	Near East	Syria	L4b2a2a	59	JQ045081	West Africa	Burkina Faso	L4b1	61
EU092773	North Africa	Egypt	L6a	59	EU092673		Israel; Ethiopian Jew	L6b	59
DQ341063	North Africa	Ethiopia	L6b	60	EU092686		Israel; Yemenite Jew	L6b	59
EU092935	North Africa	Ethiopia	L4a2	59	EU092678		Israel; Yemenite Jew	L4a1a	59
EU092949	North Africa	Ethiopia	L4a2	59	HM771233		Pygmy	L4b2b	62
FJ460531	North Africa	Tunisia	L4a1	6	JQ702504		Africa	L4b2b	Family Tree DNA
EU092662		Israel; Ethiopian Jew	L4b2a2a	59					

References:

1. Behar DM, et al. (2008) Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 3, e2062.

2. Palanichamy MG, et al. (2010) Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. BMC Evol Biol 10, 304.

3. Schönberg A, et al. (2011) High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. Eur J Hum Genet 19, 988-94.

4. Kujanová M, et al. (2009) Near eastern Neolithic genetic input in a small oasis of the Egyptian Western Desert. Am J Phys Anthropol 140, 336-46.

5. Maca-Meyer N, et al. (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2, 13.

6. Costa MD, et al. (2009) Data from complete mtDNA sequencing of Tunisian centenarians: testing haplogroup association and the "golden mean" to longevity. Mech Ageing Dev 130, 222-6.

7. Reidla M, et al. (2003) Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet 73, 1178-90.

8. Kloss-Brandstätter A, et al. (2010) Somatic mutations throughout the entire mitochondrial genome are associated with elevated PSA levels in prostate cancer patients. Am J Hum Genet 87, 802-12.

9. Fendt L, et al. (2008) Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. BMC Genomics 10, 139.

10. Gasparre G, et al. (2007) Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors. Proc Natl Acad Sci U S A 104, 9001-6.

11. Zaragoza MV, et al. (2010) Mitochondrial DNA variant discovery and evaluation in human Cardiomyopathies through next-generation sequencing. PLoS One 5, e12295.

12. Bandelt H-J, et al. (2005) Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies. Biochem Biophys Res Commun 333, 122-30.

13. Pereira L, et al. (2007) No evidence for an mtDNA role in sperm motility: data from complete sequencing of asthenozoospermic males. Mol Biol Evol 24, 868-74.

14. Derenko M, et al. (2007) Phylogeographic analysis of mitochondrial DNA in northern Asian populations. Am J Hum Genet 81, 1025-41.

15. Fraumene C, et al. (2006) High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates. Mol Biol Evol 23, 2101-11.

16. Pichler I, et al. (2010) Drawing the history of the Hutterite population on a genetic landscape: inference from Y-chromosome and mtDNA genotypes. Eur J Hum Genet 18, 463-70.

17. Finnilä S, et al. (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68, 1475-84.

18. Mazunin et al. (unpublished)

19. Malyarchuk B, et al. (2010a) The peopling of Europe from the mitochondrial haplogroup U5 perspective. PLoS One 5, e10285.

20. Malyarchuk B, et al. (2010b) Mitogenomic diversity in Tatars from the Volga-Ural region of Russia. Mol Biol Evol 27, 2220-6.

21. Derbeneva OA, et al. (2002) Mitochondrial DNA variation in Kets and Nganasans and the early peoples of Northern Eurasia. Genetika 38, 1554-60.

22. Janssen GM, et al. (2006) Novel mitochondrial DNA length variants and genetic instability in a family with diabetes and deafness. Exp Clin Endocrinol Diabetes 114, 168-74.

23. Mishmar D, et al. (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100, 171-6.

24. Shlush LI, et al. (2008) The Druze: a population genetic refugium of the Near East. PLoS One 3, e2105.

25. Feder J, et al. (2008) Differences in mtDNA haplogroup distribution among 3 Jewish populations alter susceptibility to T2DM complications. BMC Genomics 9, 198.

26. Behar DM, et al. (2006) The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. Am J Hum Genet 78, 487-97.
27. Hartmann A, et al. (2008) Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. Hum Mutat 30, 115-22.
28. Levin BC, et al. (1999) A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection. Genomics 55, 135-46.
29. Palanichamy MG, et al. (2004) Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75, 966-78.
30. Fagundes NJ, et al. (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. Am J Hum Genet 82, 583-92.
31. Achilli A, et al. (2008) The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. PLoS One 3, e1764.
32. Perego UA, et al. (2008) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. Curr Biol 19, 1-8.
33. Pierson MJ, et al. (2006) Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. Mol Biol Evol 23, 1966-75.
34. Pala M, et al. (2012) Mitochondrial DNA signals of Late Glacial recolonization of Europe from Near Eastern refugia. Am J Hum Genet 90, 915-924.
35. Behar DM, et al. (2012) The Basque paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. Am J Hum Genet 90, 486-493.
36. Greenspan, T. FTDNA Direct submission. GenBank.
37. Coble MD, et al. (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int J Legal Med 118, 137-146.
38. Pello R, et al. (2008) Mitochondrial DNA background modulates the assembly kinetics of OXPHOS complexes in a cellular model of mitochondrial disease. Hum Mol Genet 17, 4001-4011.
39. Just RS, et al. (2008) Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. Forensic Sci Int Genet 2, e45-48.
40. Malyarchuk BA, et al. (2008) Mitochondrial DNA variability in Slovaks, with application to the Roma origin. Ann Hum Genet 72, 228-240.
41. Detjen AK, et al. (2007) Analysis of mitochondrial DNA in discordant monozygotic twins with neurofibromatosis type 1. Twin Res Hum Genet 10, 486-495.
42. Fendt L, et al. (2011) Accumulation of mutations over the entire mitochondrial genome of breast cancer cells obtained by tissue microdissection. Breast Cancer Res Treat 128, 327-336.
43. Detjen AK. (direct submission).
44. Ingman M, et al. (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408, 708-713.
45. Zsurka. (Direct submission).
46. Carelli V, et al. (2006) Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees. Am J Hum Genet 78, 564-574.
47. Pope AM, et al. (2011) Mitogenomic and microsatellite variation in descendants of the founder population of Newfoundland: high genetic diversity in an historically isolated population. Genome 54, 110-119.
48. Ingman M, et al. (2007) A recent genetic link between Sami and the Volga-Ural region of Russia. Eur J Hum Genet 15, 115-120.
49. Amati-Bonneau. (Direct submission).
50. Gonder MK, et al. (2007) Whole-mtDNA genome sequence analysis of ancient African lineages. Mol Biol Evol 24, 757-768.
51. Desquiret V, et al. (Direct Submission).
52. Brown DT, et al. (2001) Random genetic drift determines the level of mutant mtDNA in human primary oocytes. Am J Hum Genet 68, 533-536.

53. Ghelli A, et al. (2009) The background of mitochondrial DNA haplogroup J increases the sensitivity of Leber's hereditary optic neuropathy cells to 2,5-hexanedione toxicity. PLoS One 4, e7922.
54. Rogaev EI, et al. (2009) Genomic identification in the historical case of the Nicholas II royal family. Proc Natl Acad Sci U S A 106, 5258-5263.
55. Rani DS, et al. (2010) Mitochondrial DNA haplogroup 'R' is associated with Noonan syndrome of south India. Mitochondrion 10, 166-173.
56. Li SB. (Direct Submission).
57. La Morgia C, et al. (2008) Rare mtDNA variants in Leber hereditary optic neuropathy families with recurrence of myoclonus. Neurology 70, 762-770.
58. Pereira JB (2013) Genetic characterisation of modern human dispersals in the Greater Mediterranean. PhD Thesis submitted in the University of Leeds.
59. Behar DM, et al. (2008) The dawn of human matrilineal diversity. Am J Hum Genet 82, 1130-40.
60. Torroni A, et al. (2006) Harvesting the fruit of the human mtDNA tree. Trends Genet 22, 339-45.
61. Barbieri C, et al. (2012) Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. Mol Biol Evol 29, 1213-23.
62. Batini C, et al. (2011) Insights into the demographic history of African Pygmies from complete mitochondrial genomes. Mol Biol Evol 28, 1099-110.

Table S 3 Frequency values used in the reconstruction of the interpolation maps for the haplogroups N1a, I, W, X and N1b.

Geographic region	N Total	Frequency Haplogroup N1a	Frequency Haplogroup I	Frequency Haplogroup W	Frequency Haplogroup X	Frequency Haplogroup N1b
Albania	102	0.0098	0.0098	0.0196	0.0196	0.0294
Armenia	191	0.0105	0.0000	0.0105	0.0105	0.0262
Austria	374	0.0053	0.0134	0.0134	0.0134	0.0000
Balkaria and Kabardia	53	0.0189	0.0000	0.0566	0.0566	0.0000
Bosnia	144	0.0000	0.0278	0.0139	0.0139	0.0069
Bulgaria	138	0.0000	0.0290	0.0072	0.0072	0.0072
Caucasus	103	0.0097	0.0000	0.0097	0.0097	0.0097
Croatia	204	0.0000	0.0196	0.0294	0.0294	0.0000
Cyprus	91	0.0000	0.0330	0.0330	0.0330	0.0110
Czech Republic	175	0.0057	0.0114	0.0114	0.0114	0.0000
Denmark	201	0.0100	0.0249	0.0050	0.0050	0.0000
Druzes	433	0.0000	0.0277	0.0092	0.0092	0.0000
Dubai	249	0.0040	0.0361	0.0281	0.0281	0.0120
Egypt	594	0.0067	0.0236	0.0067	0.0067	0.0236
Estonia	266	0.0113	0.0113	0.0301	0.0301	0.0000
Ethiopia	559	0.0029	0.0000	0.0029	0.0029	0.0000
Finland	661	0.0045	0.0348	0.0756	0.0756	0.0000
France	1285	0.0023	0.0195	0.0195	0.0195	0.0039
Germany	1628	0.0025	0.0104	0.0147	0.0147	0.0006
Greece	706	0.0042	0.0184	0.0184	0.0184	0.0071
Hungary	211	0.0000	0.0284	0.0237	0.0237	0.0095
Iceland	985	0.0000	0.0396	0.0051	0.0051	0.0000
Iran	738	0.0054	0.0217	0.0271	0.0271	0.0081
Iraq	167	0.0000	0.0180	0.0000	0.0060	0.0120
Ireland	266	0.0000	0.0226	0.0226	0.0226	0.0000
Israel	216	0.0000	0.0139	0.0093	0.0093	0.0556
Italy	1712	0.0006	0.0158	0.0158	0.0158	0.0123
Jordan	142	0.0000	0.0141	0.0070	0.0070	0.0141
Karachay-Cherkess and NW Caucasus	71	0.0000	0.0000	0.0282	0.0282	0.0000

Kenya	188	0.0053	0.0000	0.0000	0.0000	0.0000
Kurd (from Turkey and Georgia)	160	0.0000	0.0250	0.0500	0.0500	0.0000
Kuwait	381	0.0052	0.0157	0.0157	0.0157	0.0157
Latvia	413	0.0000	0.0460	0.0363	0.0363	0.0000
Libya	398	0.0025	0.0025	0.0000	0.0151	0.0050
Lithuania	343	0.0087	0.0321	0.0204	0.0204	0.0146
Macedonia	308	0.0000	0.0097	0.0325	0.0325	0.0162
Morocco	1014	0.0000	0.0010	0.0079	0.0079	0.0039
North Ossetia	231	0.0000	0.0202	0.0519	0.0519	0.0130
Norway	305	0.0066	0.0164	0.0164	0.0164	0.0066
Pakistan	189	0.0000	0.0159	0.0423	0.0423	0.0000
Poland	882	0.0023	0.0193	0.0351	0.0351	0.0045
Portugal	1612	0.0019	0.0217	0.0211	0.0211	0.0037
Romania	600	0.0017	0.0083	0.0500	0.0500	0.0150
Russia	1110	0.0009	0.0243	0.0198	0.0144	0.0009
Saudi Arabia	553	0.0217	0.0090	0.0108	0.0108	0.0217
Serbia	104	0.0096	0.0192	0.0577	0.0577	0.0096
Slovakia	710	0.0056	0.0239	0.0211	0.0211	0.0028
Slovenia	232	0.0000	0.0302	0.0259	0.0259	0.0000
Somalia	183	0.0068	0.0068	0.0000	0.0000	0.0068
Spain	2062	0.0000	0.0131	0.0112	0.0112	0.0015
Sudan	178	0.0000	0.0000	0.0056	0.0056	0.0000
Sweden	634	0.0032	0.0284	0.0110	0.0110	0.0000
Switzerland	153	0.0000	0.0131	0.0196	0.0196	0.0000
Syria	116	0.0000	0.0000	0.0345	0.0345	0.0172
Tunisia	551	0.0000	0.0091	0.0036	0.0036	0.0127
Turkey	448	0.0000	0.0156	0.0201	0.0201	0.0134
United Kingdom	3907	0.0010	0.0386	0.0100	0.0100	0.0013
Yemen	452	0.0288	0.0022	0.0066	0.0066	0.0000

Table S 4 Diversity values of ρ and Pi used for the interpolation maps of the haplogroups I,W,X, J, T and L4.

Geographic region	Haplogroup I		Haplogroup W		Haplogroup X		Haplogroup J		Haplogroup T		Haplogroup L4	
	ρ	Pi	ρ	Pi	ρ	Pi	ρ	Pi	ρ	Pi	ρ	Pi
Algeria	0.0000	0.0000	1.9000	3.4440	0.7000	1.3720	0.0000	0.0000	0.0000	0.0000	4.2000	6.3700
Armenia	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.6500	4.0050	3.4884	4.6420	0.0000	0.0000
Cameroon	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	4.9750	7.3920
Dubai	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	3.3333	3.8940	3.7500	6.4000	0.0000	0.0000
Egypt	1.0000	1.7250	0.0000	0.0000	2.0667	2.4900	2.9189	3.5980	3.7692	5.2500	0.0000	0.0000
Ethiopia	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.4000	3.9110	3.3333	5.2310	3.6694	6.0380
France	1.2400	1.8130	0.4800	0.9200	1.2143	2.3520	1.4752	2.7920	2.2069	3.6000	0.0000	0.0000
Germany	1.3607	2.0680	1.2000	2.2180	1.4615	2.5510	1.6989	3.1350	2.2761	3.6410	0.0000	0.0000
Greece	2.3548	3.9700	1.4250	2.6600	1.2444	2.4040	1.3053	2.3970	2.7000	4.2450	0.0000	0.0000
Hungary	1.7619	2.7660	1.7955	3.1190	1.3913	2.4030	1.0783	2.0660	2.7217	4.0940	0.0000	0.0000
Iceland	1.0000	1.3050	0.0000	0.0000	0.1333	0.2670	2.6567	3.8080	2.0259	3.0970	0.0000	0.0000
Iraq	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.8462	4.4260	2.8571	4.5000	0.0000	0.0000
Iran	1.3750	2.4500	1.3500	2.4500	1.0625	2.0250	2.8879	4.6300	3.3393	4.8760	0.0000	0.0000
Israel	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.5000	4.6670
Italy	1.9630	3.3940	1.6667	3.0090	0.9111	1.8340	1.6929	3.1550	2.5661	4.4570	0.0000	0.0000
Kuwait	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.8197	4.0480	3.8077	6.1820	0.0000	0.0000
Libya	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.3704	3.7780	2.7333	4.5710	0.0000	0.0000
Morocco	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.7143	3.2320	3.3243	5.0200	0.0000	0.0000
Niger	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.2500	4.4000
North Ossetia	0.0000	0.0000	1.7500	2.6380	0.6429	1.3100	2.6786	3.9870	1.8188	3.3180	0.0000	0.0000
Pakistan	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.2222	3.6110	1.8333	3.6430	0.0000	0.0000
Russia	1.9277	1.9740	0.8235	1.4850	1.2941	2.1320	1.5805	2.9220	2.2476	3.4810	0.0000	0.0000
Saudi Arabia	2.1429	3.3620	1.7619	2.7370	0.9697	1.7840	2.9386	3.9150	3.3333	4.9840	4.1111	5.8890
Sierra Leone	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	5.3330	8.1670
Spain	1.5968	2.5990	1.2105	2.1740	0.8070	1.4850	1.7815	3.3390	2.5260	4.1480	0.0000	0.0000

Sweden	0.7826	1.2570	0.4167	0.8330	1.0000	1.5560	1.5342	2.8930	2.2381	3.6410	0.0000	0.0000
Switzerland	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.4681	2.8420	2.4085	3.7930	0.0000	0.0000
Syria	1.5294	2.6620	1.0909	1.9640	1.9672	2.3270	2.1786	3.8600	3.0959	4.6160	0.0000	0.0000
Turkey	2.0000	3.3820	1.4706	2.6760	1.3750	2.6500	2.5088	4.1350	2.5932	4.2570	0.0000	0.0000
United Kingdom	1.2420	2.0670	0.8444	1.5890	0.8621	1.6630	1.7326	3.1410	2.0259	3.7860	0.0000	0.0000
Yemen	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.1884	3.3580	2.6800	4.6800	0.0000	0.0000
Zambia	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	3.4000	5.3450

Table S 5 Founder lineages identified when using a *f1* criterion from Near East to Europe.

f1			From Near East to Europe		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**
N1a1a1	F1	147A	2	0.5000	0.5000
N1a1a1a	F2	147A 320	27	1.2593	0.5316
N1a1a1a	F3	147A 154 320	9	1.4444	0.8389
N1a1a1a	F4	147A 320 325	1	0.0000	0.0000
N1a	F5	295	1	0.0000	0.0000
N1a	F6	Root	7	0.4286	0.2474
N1b	F7	126 223	2	1.0000	0.7071
N1b	F8	291	3	1.6667	0.8819
N1b	F9	176A	9	1.3333	0.7536
N1b	F10	400	1	0.0000	0.0000
N1b	F11	390	5	1.2000	0.7483
N1b	F12	209	8	1.0000	0.7706
N1b	F13	311	5	0.6000	0.3464
N1b	F14	126	1	0.0000	0.0000
N1b	F15	309	1	0.0000	0.0000
N1b	F16	Root	57	0.4737	0.1203
N1cde	F17	201 93 265	2	0.0000	0.0000
N1cde	F18	111 201 265	2	0.0000	0.0000
N1cde	F19	301 356	1	0.0000	0.0000
N1cde	F20	201 265	8	1.0000	0.4677
N1cde	F21	265	1	0.0000	0.0000
N1cde	F22	Root	2	0.0000	0.0000
I5	F23	148 311	4	0.5000	0.5000
I5a	F24	148 294	5	1.4000	0.7746
I5a	F25	148	25	1.0000	0.3274
I	F26	294	2	1.0000	0.7071
I	F27	184	1	0.0000	0.0000
I1a1d	F28	172 189 311	8	0.1250	0.1250
I1a	F29	172 311	152	0.3224	0.1369
I1a	F30	172 355	5	0.0000	0.0000
I1a	F31	172	2	0.0000	0.0000
I	F32	223 311	2	1.5000	0.8660
I	F33	311	36	0.3611	0.1690
I	F34	311 391	6	0.6667	0.4082
I	F35	391	21	0.1905	0.1166
I1	F36	129 311	3	0.0000	0.0000
I	F37	270	7	3.7143	1.7261
I	F38	169	8	1.5000	0.8660
I	F39	Root	253	0.6838	0.1894
W	F40	292	3	0.3333	0.3333
W6	F41	172 192 325	1	0.0000	0.0000
W	F42	223 294	1	0.0000	0.0000
W6	F43	211 325	1	0.0000	0.0000
W6	F44	192 325	53	0.3208	0.1084
W	F45	172	9	1.1111	1.0062
W1	F46	260	6	2.0000	0.8498
W4a	F47	286	17	0.9412	0.6281
W	F48	344	1	0.0000	0.0000
W1	F49	298	1	0.0000	0.0000
W	F50	192	1	0.0000	0.0000
W	F51	295	47	0.6383	0.2571
W6	F52	325	18	1.6111	0.7307
W	F53	root	310	0.6935	0.1716
N2a	F54	111	2	0.5000	0.5000
N2a	F55	root	3	0.3333	0.3333
X1	F56	104	1	0.0000	0.0000
X2j	F57	179	5	2.6000	1.0770

X2h	F58	223	20	0.2000	0.1581
X4	F59	274	5	0.4000	0.2828
X	F60	362	1	0.0000	0.0000
X	F61	344	5	1.0000	0.6633
X2	F62	248	13	0.1538	0.1088
X1	F63	104 278	1	0.0000	0.0000
X2n	F64	266 274	4	1.0000	0.7906
X	F65	111 344	1	0.0000	0.0000
X	F66	256	1	0.0000	0.0000
X2	F67	108	1	0.0000	0.0000
X2	F68	265	7	0.8571	0.6061
X	F69	261	1	0.0000	0.0000
X2n	F70	266	3	0.6667	0.4714
X	F71	Root	320	0.9000	0.2384

Table S 6 Founder lineages identified when using a *f2* criterion from Near East to Europe.

f2			From Near East to Europe		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**
N1a1a1	F1	147A	2	0.5000	0.5000
N1a1a1a	F2	147A 320	28	1.3571	0.5175
N1a1a1a	F3	147A 154 320	9	1.4444	0.8389
N1a	F4	Root	8	0.7500	0.3062
N1b	F5	126	3	1.3333	0.8165
N1b	F6	Root	89	1.0562	0.2138
N1cde	F7	301	1	0.0000	0.0000
N1cde	F8	201 265	12	1.1667	0.4249
N1cde	F9	265	1	0.0000	0.0000
N1cde	F10	Root	2	0.0000	0.0000
I5a	F11	148	34	1.3824	0.3517
I1a	F12	172 311	160	0.3625	0.1395
I1a	F13	172	7	1.4286	1.0102
I	F14	311	41	0.5122	0.1776
I	F15	311 391	6	0.6667	0.4082
I	F16	Root	292	0.8801	0.1893
W	F17	292	3	0.3333	0.3333
W6	F18	192 325	54	0.3333	0.1080
W4a	F19	286	17	0.9412	0.6281
W1	F20	298	1	0.0000	0.0000
W6	F21	325	19	1.5789	0.6943
W	F22	root	375	0.8880	0.1963
N2a	F23	111	2	0.5000	0.5000
N2a	F24	root	3	0.3333	0.3333
X2h	F25	223	20	0.2000	0.1581
X	F26	344	6	1.1667	0.6009
X2	F27	248	13	0.1538	0.1088
X1	F28	104 278	1	0.0000	0.0000
X	F29	Root	349	1.0086	0.2230

Table S 7 Frequency values used in the reconstruction of the interpolation maps for the haplogroup J,T and the subhaplogroups J1d1a and J2a2b

Geographic region	N Total	Frequency Haplogroup J	Frequency Haplogroup J1d1a	Frequency Haplogroup J2a2b	Frequency Haplogroup T
Abkhazia	27	0.0370	0.0000	0.0000	0.0370
Albania	102	0.098	0.0000	0.0000	0.1274
Algeria	47	0.1277	0.0000	0.0000	0.0425
Angola	519	0.0019	0.0000	0.0000	0.0000

Armenia	191	0.0838	0.0000	0.0000	0.1152
Austria	374	0.0909	0.0000	0.0000	0.1230
Azerbaijan	48	0.0417	0.0000	0.0000	0.1875
Balkaria	14	0.0000	0.0000	0.0000	0.1429
Bosnia	144	0.0694	0.0000	0.0000	0.0486
Bulgaria	138	0.0725	0.0000	0.0000	0.1015
Cabo verde	292	0.0034	0.0000	0.0000	0.0034
Caucasus	137	0.0291	0.0000	0.0000	0.1165
Croatia	245	0.0735	0.0000	0.0000	0.1143
Cyprus	91	0.0549	0.0000	0.0000	0.0769
Czech Republic	175	0.0914	0.0000	0.0000	0.1200
Denmark	201	0.1244	0.0050	0.0000	0.0846
Druze	433	0.0485	0.0000	0.0000	0.0600
Dubai	249	0.1165	0.0000	0.012	0.0321
Egypt	594	0.0606	0.0034	0.0017	0.0976
Estonia	266	0.1053	0.0000	0.0000	0.1015
Ethiopia	559	0.0089	0.0054	0.0000	0.0197
Finland	661	0.0454	0.0000	0.0000	0.0378
France	1285	0.0755	0.0000	0.0000	0.0903
Georgia	45	0.0444	0.0000	0.0000	0.2444
Germany	1841	0.0918	0.0000	0.0000	0.1157
Greece	706	0.0949	0.0000	0.0057	0.0991
Hungary	211	0.0995	0.0000	0.0000	0.1422
Iceland	985	0.136	0.0000	0.0000	0.1178
Iran	738	0.145	0.0027	0.0027	0.0759
Iraq	167	0.1557	0.0120	0.0000	0.0838
Ireland	266	0.109	0.0000	0.0000	0.0789
Israel	216	0.0787	0.0093	0.0000	0.1065
Italy	1712	0.0765	0.0012	0.0006	0.1104
Jordan	142	0.0493	0.0000	0.0000	0.0704
Karachay-Cherkess	13	0.0769	0.0000	0.0000	0.0000
Kenya	188	0.0053	0.0000	0.0000	0.0000
Kurd	160	0.0938	0.0000	0.0000	0.1250
Kuwait	381	0.1549	0.0157	0.0026	0.0682
Latvia	413	0.0605	0.0000	0.0000	0.0920
Libya	418	0.0622	0.0239	0.0120	0.0383
Lithuania	343	0.0641	0.0000	0.0000	0.1020
Macedonia	308	0.0584	0.0000	0.0000	0.1169
Mali	262	0.0191	0.0000	0.0000	0.0000
Mauritania	94	0.0425	0.0000	0.0000	0.0000
Morocco	1103	0.0526	0.0018	0.0054	0.0571
Niger	165	0.0182	0.0000	0.0000	0.0000
Nigeria	1425	0.0035	0.0000	0.0000	0.0014
North Ossetia	231	0.0822	0.0000	0.0000	0.0649
Norway	305	0.0754	0.0000	0.0000	0.0885
Oman	196	0.0765	0.0051	0.0051	0.1020
Pakistan	189	0.0476	0.0000	0.0000	0.0265
Poland	882	0.0782	0.0000	0.0023	0.0952
Portugal	1612	0.0651	0.0000	0.0006	0.0961
Romania	600	0.105	0.0000	0.0000	0.0883
Russia	1110	0.0586	0.0000	0.0000	0.0703
Saudi Arabia	553	0.2061	0.0326	0.0036	0.0615
Senegal	280	0.0071	0.0000	0.0000	0.0000
Serbia	104	0.0865	0.0000	0.0000	0.1058
Slovakia	710	0.0958	0.0000	0.0000	0.0930
Slovenia	232	0.0733	0.0000	0.0000	0.0991
Somalia	183	0.0109	0.0000	0.0000	0.0109
South Africa	637	0.0000	0.0000	0.0000	0.0031
Spain	2062	0.0737	0.0000	0.0000	0.0834
Sudan	178	0.0056	0.0000	0.0000	0.0168

Sweden	634	0.0789	0.0000	0.0000	0.0899
Switzerland	153	0.1176	0.0000	0.0000	0.1634
Syria	116	0.0948	0.0000	0.0000	0.1207
Tanzania	96	0.0104	0.0104	0.0000	0.0000
Tunisia	551	0.0563	0.0000	0.0109	0.1034
Turkey	448	0.0938	0.0000	0.0000	0.0870
United Kingdom	3907	0.1211	0.0000	0.0003	0.0906
Ukraine	18	0.0556	0.0000	0.0000	0.1667
West Saharan	110	0.0091	0.0000	0.0000	0.0119
Yemen	552	0.1467	0.0199	0.0181	0.0580

Table S 8 Founder lineages identified when using a *f1* criterion from Near and Pakistan to Arabian Peninsula.

f1			**From Near East and Pakistan to Arabian Peninsula**		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**
HV1	F1	67 292 354	1	0.0000	0.0000
H	F2	316	2	1.0000	0.7071
HV1a1	F3	67 355	2	0.0000	0.0000
H6b	F4	300 362	8	0.8750	0.4841
HV	F5	69	3	0.0000	0.0000
H	F6	266	1	0.0000	0.0000
H	F7	92	2	0.0000	0.0000
HV	F8	210	1	0.0000	0.0000
H	F9	168	2	0.0000	0.0000
HV2	F10	217	5	1.4000	0.6633
HV	F11	114	1	0.0000	0.0000
HV	F12	145	1	0.0000	0.0000
H15a1b	F13	248	1	0.0000	0.0000
H15a1b	F14	319	2	0.5000	0.5000
HV	F15	243	2	0.5000	0.5000
HV	F16	172	1	0.0000	0.0000
HV	F17	355	2	1.0000	0.7071
HV	F18	220C 292	2	0.0000	0.0000
HV	F19	298	1	0.0000	0.0000
H2a1	F20	354	15	0.1333	0.0943
HV	F21	153	2	0.0000	0.0000
H	F22	218	6	0.5000	0.2887
HV1	F23	67	13	1.5385	0.6057
HV	F24	221	1	0.0000	0.0000
H5	F25	304	5	1.2000	0.4899
H6	F26	362	2	0.0000	0.0000
H	F27	261	2	0.0000	0.0000
H1	F28	278	2	0.5000	0.5000
H2a3	F29	274	2	1.0000	0.7071
HV	F30	240	1	0.0000	0.0000
HV	F31	192	4	0.0000	0.0000
H	F32	189	2	0.0000	0.0000
H	F33	93	2	0.0000	0.0000
H	F34	242	1	0.0000	0.0000
H	F35	239	4	0.0000	0.0000
HV	F36	Root	56	0.8036	0.1722
M1a3	F37	223 311	1	0.0000	0.0000
M1b1	F38	185	2	0.5000	0.5000
M1a3	F39	223	1	0.0000	0.0000
M1a1	F40	359	14	0.3571	0.1597
M1	F41	Root	8	1.1250	0.4146
N1b1	F42	145 176G 309 390	2	0.0000	0.0000
N1	F43	301	3	0.0000	0.0000
N1a1a1	F44	147G 172 248 355	12	0.8333	0.4082
N1a1a1	F45	147A 172 248 355	14	2.4286	0.8512

N1a3	F46	201 265	10	0.1000	0.1000
N1a3	F47	189 201 265	1	0.0000	0.0000
N1a3	F48	265	5	2.4000	1.3856
N1b1	F49	145 176G 390	13	0.8462	0.3846
N1b1	F50	93 145 176G 390	2	0.0000	0.0000
I5a	F51	129 148 391	4	1.5000	0.7071
I1	F52	129 311 391	1	0.0000	0.0000
I	F53	129 391	10	2.4000	1.0954
W	F54	292	7	0.7143	0.4286
W6a'b	F55	192 292 325	3	1.3333	0.6667
W6	F56	292 325	3	0.3333	0.3333
N2a	F57	153 319	2	0.0000	0.0000
R0a	F58	189	9	0.1111	0.1111
R0a1a	F59	355	85	0.6588	0.2445
R0a	F60	Root	78	0.8462	0.4205
R0a	F61	114	1	0.0000	0.0000
R2	F62	325 355 357	1	0.0000	0.0000
R2c	F63	320	1	0.0000	0.0000
R2	F64	234	1	0.0000	0.0000
R2	F65	Root	8	1.1250	0.5154
T2	F66	146 292 296!	4	0.2500	0.2500
T2	F67	288 292 296	2	0.0000	0.0000
T2b7a2	F68	239 296 304	1	0.0000	0.0000
T2	F69	93 296	2	0.5000	0.5000
T1b	F70	163 189 243	2	0.0000	0.0000
T2b	F71	296! 304	1	0.0000	0.0000
T2c1c	F72	146 292 296	2	0.5000	0.5000
T2c1	F73	292 296!	4	0.2500	0.2500
T1a	F74	163 186 189	18	0.2778	0.1242
T2e	F75	153 296	1	0.0000	0.0000
T2b7a1	F76	153 257 296!	5	1.2000	0.6325
T2b	F77	296 304	1	0.0000	0.0000
T2c1	F78	292 296	6	2.0000	0.9718
T2	F79	296!	2	0.0000	0.0000
T1	F80	163 189	3	0.6667	0.6667
T2	F81	296	9	0.8889	0.5879
J	F82	188 311	8	1.1250	1.0078
J2a1a1	F83	145 231 261	1	0.0000	0.0000
J1b1b2	F84	145 222 235 261	2	0.0000	0.0000
J1d1a	F85	193 300 309	22	0.0455	0.0455
J1d1	F86	193 287 300	1	0.0000	0.0000
J1b	F87	145 222 261!	1	0.0000	0.0000
J1b	F88	145 222 261 311	3	0.0000	0.0000
J1b	F89	145 222 256 261 278	2	0.0000	0.0000
J1b	F90	145 222 261 300	5	0.6000	0.6000
J1d5	F91	193 274	1	0.0000	0.0000
J	F92	231 319	6	0.0000	0.0000
J	F93	319	1	0.0000	0.0000
J1b	F94	126 145 261	5	0.4000	0.2828
J1d1	F95	193 300	3	0.3333	0.3333
J1b	F96	145 222 261	68	0.5294	0.2767
J	F97	145	3	0.6667	0.6667
J	F98	371	2	0.0000	0.0000
J	F99	231	2	0.5000	0.5000
J	F100	69	2	1.5000	0.8660
J2a2b	F101	241	9	0.7778	0.5556
J1d	F102	193	12	0.8333	0.6009
J1b	F103	145 261	32	0.5000	0.1466
J	F104	Root	22	0.7273	0.2727
U5a1	F105	192 256 399	2	0.5000	0.5000
U5a	F106	192 256	1	0.0000	0.0000
U5	F107	192	3	0.6667	0.4714
U5a	F108	256	1	0.0000	0.0000
U5	F109	Root	3	0.0000	0.0000
U8b1	F110	189 234 257 259 290	2	0.5000	0.5000

U6a2'3	F111	172 189 219 278	3	0.6667	0.4714
U2b	F112	51 239 288 353	1	0.0000	0.0000
U1a	F113	189 249 288 362	1	0.0000	0.0000
U2e	F114	51 129C 189 362	9	0.5556	0.5556
U2d1	F115	51 184 189! 234 294 342	1	0.0000	0.0000
U1	F116	184A 249 355	1	0.0000	0.0000
U6a	F117	172 219 278	1	0.0000	0.0000
U2b2	F118	51 353	11	4.0000	1.7953
U7	F119	207 309 318C	1	0.0000	0.0000
U7	F120	309 318T 362	1	0.0000	0.0000
U1a'c	F121	145 189 249	1	0.0000	0.0000
U6a'b'd	F122	172 219	2	0.0000	0.0000
U2	F123	51 247 254	1	0.0000	0.0000
U1a'c	F124	189 249	7	0.8571	0.4949
U9a	F125	51 129 259 278	5	0.2000	0.2000
U1b	F126	111 214A 249 327	1	0.0000	0.0000
K1a4c1	F127	224 246T 311	2	0.0000	0.0000
U7	F128	309 318T	2	0.5000	0.5000
U7	F129	309 318C	1	0.0000	0.0000
U8b1	F130	189 234	1	0.0000	0.0000
U8b1a1	F131	129 189 234	1	0.0000	0.0000
K	F132	192 224 311	5	0.0000	0.0000
U3a	F133	343 390	9	2.1111	0.7454
U3b3	F134	168 343	2	0.5000	0.5000
U7	F135	129 318T	2	1.0000	0.7071
U4a1	F136	134 356	6	0.5000	0.2887
U3b1a	F137	86 343	5	2.6000	1.0000
K1b1a	F138	224 311 319	2	0.0000	0.0000
K1	F139	93 224 311	13	0.0000	0.0000
U1	F140	249	1	0.0000	0.0000
U2	F141	51	5	3.2000	1.0583
U9a	F142	51 278	10	0.5000	0.5000
U7	F143	318T	3	0.6667	0.6667
U4	F144	356	5	2.6000	0.8718
U3	F145	343	5	0.8000	0.4000
U7a4	F146	126 209 309 318T 390	1	0.0000	0.0000
U7a4	F147	126 209 309 318T	1	0.0000	0.0000
K	F148	224 311	33	1.2424	0.3193
X2j	F149	179	1	0.0000	0.0000
X	F150	362	1	0.0000	0.0000
X	F151	344	8	0.3750	0.2165
X2	F152	248	2	1.0000	0.7071
X	F153	Root	11	0.1818	0.1286

Table S 9 Founder lineages identified when using a *f2* criterion from Near and Pakistan to Arabian Peninsula.

f2			**From Near East and Pakistan to Arabian Peninsula**		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**
HV1a1	F1	67 355	2	0.0000	0.0000
H6	F2	300 362	8	0.8750	0.4841
H	F3	92	2	0.0000	0.0000
HV2	F4	217	5	1.4000	0.6633
HV	F5	114	1	0.0000	0.0000
H15a1b	F6	248	1	0.0000	0.0000
HV	F7	355	2	1.0000	0.7071
HV	F8	298	1	0.0000	0.0000
H2a1	F9	354	15	0.1333	0.0943
H	F10	218	6	0.5000	0.2887
HV1	F11	67	14	1.6429	0.5759
H5	F12	304	5	1.2000	0.4899
H6	F13	362	2	0.0000	0.0000

H	F14	261	2	0.0000	0.0000
HV	F15	192	4	0.0000	0.0000
H	F16	189	2	0.0000	0.0000
H	F17	93	2	0.0000	0.0000
H	F18	239	4	0.0000	0.0000
HV	F19	Root	82	1.0122	0.1518
M1a1	F20	359	14	0.3571	0.1597
M1	F21	Root	12	1.3333	0.3909
N1a1a1	F22	147A 172 248 355	26	2.1538	0.6772
N1a3	F23	201 265	11	0.1818	0.1286
N1a3	F24	265	5	2.4000	1.3856
N1b1	F25	145 176G 390	17	0.8824	0.3379
I	F26	129 391	15	2.4667	0.8138
N1	F27	Root	3	0.0000	0.0000
W	F28	292	7	0.7143	0.4286
N2a	F29	153 319	2	0.0000	0.0000
W6a'b	F30	192 292 325	3	1.3333	0.6667
W6	F31	292 325	3	0.3333	0.3333
R0a	F32	189	9	0.1111	0.1111
R0a1a	F33	355	85	0.6588	0.2445
R0a	F34	114	1	0.0000	0.0000
R0a	F35	Root	78	0.8462	0.4205
R2	F36	Root	11	1.5455	0.4545
T2b	F37	296! 304	1	0.0000	0.0000
T2	F38	288 292 296	2	0.0000	0.0000
T2c1c	F39	146 292 296	6	1.0000	0.7071
T2c1	F40	292 296 296	4	0.2500	0.2500
T1a	F41	163 186 189	18	0.2778	0.1242
T2b	F42	296 304	2	1.0000	0.7071
T2c1	F43	292 296	6	2.0000	0.9718
T2	F44	296 296	2	0.0000	0.0000
T1	F45	163 189	5	0.8000	0.5657
T2	F46	296	17	2.0000	0.6707
J2a1a1	F47	145 231 261	1	0.0000	0.0000
J1d1a	F48	193 300 309	22	0.0455	0.0455
J1b	F49	145 222 256 261 278	2	0.0000	0.0000
J1b1b2	F50	145 222 235 261	2	0.0000	0.0000
J1d1	F51	193 300	4	0.7500	0.4330
J1b	F52	145 222 261	77	0.6364	0.2594
J	F53	145	3	0.6667	0.6667
J	F54	231	8	0.8750	0.7603
J	F55	69	2	1.5000	0.8660
J2a2b	F56	241	9	0.7778	0.5556
J1d	F57	193	13	0.8462	0.5600
J1b	F58	145 261	37	0.6216	0.1892
J	F59	Root	33	1.3333	0.4635
U5a1	F60	192 256 399	2	0.5000	0.5000
U5a	F61	192 256	1	0.0000	0.0000
U5	F62	192	3	0.6667	0.4714
U5a	F63	256	1	0.0000	0.0000
U5	F64	Root	3	0.0000	0.0000
U2d1	F65	51 184 189 234 294 342	1	0.0000	0.0000
U2e	F66	51 129C 189 362	9	0.5556	0.5556
U6a	F67	172 219 278	4	1.7500	0.9014
U5b2c2b	F68	189 249 288	1	0.0000	0.0000
U7a4	F69	126 209 309 318T	2	1.5000	0.8660
U1a'c	F70	189 249	8	1.2500	0.5000
U1b	F71	111 249 327	1	0.0000	0.0000
U7	F72	309 318T	3	0.6667	0.4714
U7	F73	309 318C 318T	2	1.0000	0.7071
U3a	F74	343 390	9	2.1111	0.7454
U3b3	F75	168 343	2	0.5000	0.5000
U8b	F76	189 234 257 259	2	2.5000	1.5000
U8b1	F77	189 234	1	0.0000	0.0000
K	F78	129 189 234	1	0.0000	0.0000

K1	F79	93 224 311	13	0.0000	0.0000
U1	F80	249	1	0.0000	0.0000
U1	F81	249 355	1	0.0000	0.0000
U2	F82	51	18	4.3333	1.3240
U7	F83	318T	5	1.2000	0.6325
U4	F84	356	5	2.6000	0.8718
U4a1	F85	134 356	6	0.5000	0.2887
U3	F86	343	10	2.2000	0.7348
K	F87	224 311	42	1.3095	0.3095
U	F88	Root	17	3.2941	1.3745
X2	F89	248	2	1.0000	0.7071
X	F90	344	8	0.3750	0.2165
X	F91	Root	13	0.3077	0.1538

Table S 10 Frequency values used in the reconstruction of the interpolation maps for the haplogroups L4 and L6.

Geographic region	N Total	Frequency Haplogroup L4	Frequency Haplogroup L6
Saudi Arabia	553	0.0036	0.0018
Angola	519	0.0039	0.0000
Burkina Faso	119	0.0084	0.0000
Cameroon	737	0.0095	0.0000
Chad	117	0.0171	0.0000
Dubai	249	0.0120	0.0000
Egypt	594	0.0084	0.0000
Ethiopia	636	0.0802	0.0189
Gabon	833	0.0048	0.0000
Ghana	238	0.0042	0.0000
Israel	216	0.0046	0.0000
Kenya	329	0.0851	0.0000
Kuwait	381	0.0079	0.0000
Libya	527	0.0057	0.0000
Morocco	1103	0.0027	0.0000
Niger	195	0.0102	0.0000
Nigeria	1425	0.0182	0.0000
Rwanda	42	0.0714	0.0000
Somali	186	0.0376	0.0161
South Africa	637	0.0078	0.0000
Sudan	178	0.0449	0.0000
Syria	116	0.0086	0.0000
Tanzania	171	0.1696	0.0000
Tunisia	551	0.0018	0.0018
Turkana	24	0.1250	0.0000
Yemen	552	0.0072	0.0235
Zambia	78	0.0256	0.0000
Zimbabwe	59	0.0169	0.0000

Table S 11 Founder lineages identified when using a *f1* criterion from Africa to Arabian Peninsula and Near East.

f1			From Africa to Arabian Peninsula and Near East		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**
L0a1'4	F1	129 287	1	0.0000	0.0000
	F2	93 293	1	0.0000	0.0000
	F3	278 293	1	0.0000	0.0000
	F4	114A	1	0.0000	0.0000
	F5	278	9	0.6667	0.2722

	F6	320	5	0.0000	0.0000
	F7	293	11	0.0000	0.0000
L0a*	F8	129 169 278	1	0.0000	0.0000
	F9	214 188A 234 320	1	0.0000	0.0000
	F10	209	1	0.0000	0.0000
	F11	129	2	1.0000	0.7071
	F12	93	3	0.0000	0.0000
	F13	root	33	0.2727	0.1249
L0b	F14	root	1	0.0000	0.0000
L0d	F15	290 300 399	1	0.0000	0.0000
L0f	F16	52 290 354 325	1	0.0000	0.0000
L0k	F17	223 291A	2	0.0000	0.0000
L1b	F18	187	1	0.0000	0.0000
	F19	239	1	0.0000	0.0000
	F20	293	1	0.0000	0.0000
	F21	root	4	0.2500	0.2500
L1c	F22	173 172 187 188A 256 293 368	3	0.6667	0.4714
L1c2	F23	71 145 213 234	4	0.0000	0.0000
	F24	78 187 278 320	1	0.0000	0.0000
	F25	78 278 320	2	0.0000	0.0000
	F26	93 320	1	0.0000	0.0000
L1c3a	F27	355 390	1	0.0000	0.0000
	F28	360	1	0.0000	0.0000
	F29	root	3	1.0000	0.5774
L2c	F30	234 249 294 295	1	0.0000	0.0000
	F31	311 264	1	0.0000	0.0000
	F32	264	2	0.0000	0.5774
L2a	F33	93 189 192 309 309	1	0.0000	0.0000
	F34	192 192 189 290 309	4	0.2500	0.2500
	F35	131 189 225 234 309	1	0.0000	0.0000
	F36	234 249 295 309 309	2	1.0000	0.7071
	F37	173 189 192 309	1	0.0000	0.0000
	F38	189 192 309 309	4	0.2500	0.2500
	F39	189 192 290 309	3	0.0000	0.0000
	F40	189 192 284 309	3	0.6667	0.6667
	F41	93 192 189 309	1	0.0000	0.0000
	F42	192 189 209 309	1	0.0000	0.0000
	F43	93 256 309	1	0.0000	0.0000
	F44	187 189 309	1	0.0000	0.0000
	F45	150 189 309	1	0.0000	0.0000
	F46	93 189 309	1	0.0000	0.0000
	F47	189 192 309	3	0.3333	0.3333
	F48	189 193 309	1	0.0000	0.0000
	F49	189 209 309	1	0.0000	0.0000
	F50	92 286 309	2	0.0000	0.0000
	F51	209 301 354	5	0.4000	0.2828
	F52	189 229 291 311	2	1.0000	0.7071
	F53	309 344	1	0.0000	0.0000
	F54	289 309	1	0.0000	0.0000
	F55	269 309	1	0.0000	0.0000
	F56	169 309	2	0.5000	0.5000
	F57	192 309	4	0.0000	0.0000
	F58	256 309	1	0.0000	0.0000
	F59	86 309	2	0.5000	0.5000
	F60	189 309	6	0.1667	0.1667
	F61	129 309	6	0.0000	0.0000
	F62	286 309	9	0.1111	0.1111
	F63	189	2	1.0000	0.7071
	F64	309	8	0.5000	0.3062
	F65	root	6	0.6667	0.4082
L2b	F66	145 179 189	1	0.0000	0.0000
	F67	169 362	2	0.5000	0.5000

	F68	355 362	2	0.0000	0.0000
	F69	145	2	0.0000	0.0000
L2d	F70	311	2	0.0000	0.0000
L2e	F71	355 390 399 400	1	0.0000	0.0000
	F72	root	1	0.0000	0.0000
L3a	F73	192	1	0.0000	0.0000
L3b	F74	311	1	0.0000	0.0000
	F75	93	9	0.0000	0.0000
	F76	root	12	0.2500	0.1443
	F77	189	1	0.0000	0.0000
L3d	F78	189 278 304 311	3	0.0000	0.0000
	F79	256 368	2	0.5000	0.5000
	F80	300 + 319	1	0.0000	0.0000
	F81	325	1	0.0000	0.0000
	F82	111	1	0.0000	0.0000
	F83	256	1	0.0000	0.0000
	F84	319	23	0.0870	0.0615
	F85	root	1	0.0000	0.0000
L3e1	F86	185 209 311	1	0.0000	0.0000
	F87	185 311	1	0.0000	0.0000
	F88	185 209	4	0.0000	0.0000
	F89	256 330	1	0.0000	0.0000
	F90	256	1	0.0000	0.0000
	F91	root	10	0.5000	0.3317
L3e2	F92	172 189	4	0.0000	0.0000
L3e3	F93	root	19	0.2105	0.1053
L3f	F94	93 292	2	0.0000	0.0000
	F95	292 295	1	0.0000	0.0000
	F96	189 292	1	0.0000	0.0000
	F97	311	1	0.0000	0.0000
	F98	292	13	0.6154	0.2665
	F99	111A	1	0.0000	0.0000
	F100	root	3	0.3333	0.3333
L3h	F101	179 215 256A 284	1	0.0000	0.0000
	F102	653	1	0.0000	0.0000
	F103	192 218 256A 362	1	0.0000	0.0000
	F104	179 256A 284	2	0.5000	0.5000
	F105	165 192 399	1	0.0000	0.0000
	F106	93 270	1	0.0000	0.0000
	F107	111 184 304 311	3	0.6667	0.4714
	F108	148 192	3	0.0000	0.0000
	F109	399	1	0.0000	0.0000
L3i	F110	184 260 311	2	0.0000	0.0000
	F111	260 311	8	0.2500	0.1768
	F112	153	2	0.0000	0.0000
	F113	260	1	0.0000	0.0000
L3x	F114	189 311 278 298	1	0.0000	0.0000
	F115	86 193 195 223	3	0.6667	0.6667
	F116	256 278 311	1	0.0000	0.0000
	F117	278 298 311	1	0.0000	0.0000
	F118	193 195 223	1	0.0000	0.0000
	F119	278	2	1.0000	0.7071
L4	F120	86 293T 355 399	1	0.0000	0.0000
	F121	172 293T 355 399	1	0.0000	0.0000
	F122	274 293T 355 399	1	0.0000	0.0000
	F123	207T 220 260 261	2	0.0000	0.0000
	F124	189 260 264	2	0.5000	0.5000
	F125	93 207T 260	1	0.0000	0.0000
	F126	207T 260 261	2	0.0000	0.0000
	F127	179 189 239 320	1	0.0000	0.0000
	F128	293T 355	1	0.0000	0.0000
	F129	207T 260	1	0.0000	0.0000

	F130	root	1	0.0000	0.0000
L5	F131	111 254 311 335 360	1	0.0000	0.0000
	F132	355 362	5	0.2000	0.2000
L6	F133	173 362	12	0.0000	0.0000
	F134	362	1	0.0000	0.0000
	F135	root	1	0.0000	0.0000

Table S 12 Founder lineages identified when using a *f2* criterion from Africa to Arabian Peninsula and Near East.

f2			**From Africa to Arabian Peninsula and Near East**		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**
L0a1'4	F1	129 287	1	0.0000	0.0000
	F2	278 293	1	0.0000	0.0000
	F3	114A	1	0.0000	0.0000
	F4	278	9	0.6667	0.2722
	F5	320	5	0.0000	0.0000
	F6	293	12	0.1667	0.1179
L0a*	F7	188A 214 234 320	1	0.0000	0.0000
	F8	169 278	1	0.0000	0.0000
	F9	129	2	1.0000	0.7071
	F10	93	3	0.0000	0.0000
	F11	Root	34	0.2941	0.1248
L0b	F12	Root	1	0.0000	0.0000
L0d	F13	290 300	1	0.0000	0.0000
L0f	F14	52 290 325	1	0.0000	0.0000
L0k	F15	root	2	0.0000	0.0000
L1b	F16	187	1	0.0000	0.0000
	F17	239	1	0.0000	0.0000
	F18	293	1	0.0000	0.0000
	F19	Root	4	0.2500	0.2500
L1c	F20	293 368	3	0.6667	0.4714
L1c2	F21	71 145 213 234	4	0.0000	0.0000
	F22	78 278 320	3	1.0000	0.7454
	F23	320	1	0.0000	0.0000
L1c3a	F24	360	1	0.0000	0.0000
	F25	355	1	0.0000	0.0000
	F26	Root	3	1.0000	0.5774
L2c	F27	234 249 294 295	1	0.0000	0.0000
	F28	264 311	1	0.0000	0.0000
	F29	264	2	0.0000	0.5774
L2a	F30	93 189 192 309 309	1	0.0000	0.0000
	F31	189 192 192 290 309	4	0.2500	0.2500
	F32	131 189 225 234 309	1	0.0000	0.0000
	F33	249 295 309 309 234	2	1.0000	0.7071
	F34	189 192 309 309	4	0.2500	0.2500
	F35	192 189 290 309	3	0.0000	0.0000
	F36	189 192 284 309	3	0.6667	0.6667
	F37	93 189 192 309	1	0.0000	0.0000
	F38	187 189 309	1	0.0000	0.0000
	F39	150 189 309	1	0.0000	0.0000
	F40	93 189 309	1	0.0000	0.0000
	F41	189 192 309	4	0.7500	0.4330
	F42	189 193 309	1	0.0000	0.0000
	F43	92 286 309	2	0.0000	0.0000
	F44	209 301 354	5	0.4000	0.2828
	F45	189 229 291 311	2	1.0000	0.7071
	F46	309 344	1	0.0000	0.0000
	F47	269 309	1	0.0000	0.0000
	F48	192 309	4	0.0000	0.0000

	F49	256 309	2	1.5000	0.8660
	F50	86 309	2	0.5000	0.5000
	F51	189 309	8	0.5000	0.3062
	F52	129 309	6	0.0000	0.0000
	F53	286 309	9	0.1111	0.1111
	F54	189	2	1.0000	0.7071
	F55	309	11	1.0000	0.3748
	F56	Root	6	0.6667	0.4082
L2b	F57	355 362	2	0.0000	0.0000
	F58	145	3	1.3333	0.8165
	F59	362	2	0.5000	0.5000
L2d	F60	Root	2	0.0000	0.0000
L2e	F61	355 400	1	0.0000	0.0000
	F62	Root	1	0.0000	0.0000
L3a	F63	192	1	0.0000	0.0000
L3b	F64	311	1	0.0000	0.0000
	F65	93	9	0.0000	0.0000
	F66	Root	12	0.2500	0.1443
	F67	189	1	0.0000	0.0000
L3d	F68	189 278 304 311	3	0.0000	0.0000
	F69	256 368	2	0.5000	0.5000
	F70	300 319	1	0.0000	0.0000
	F71	111	1	0.0000	0.0000
	F72	256	1	0.0000	0.0000
	F73	319	23	0.0870	0.0615
	F74	Root	2	1.0000	0.7071
L3e1	F75	185 209 311	1	0.0000	0.0000
	F76	185 311	1	0.0000	0.0000
	F77	185 209	4	0.0000	0.0000
	F78	256	2	1.0000	0.7071
	F79	Root	10	0.5000	0.3317
L3e2	F80	172 189	4	0.0000	0.0000
L3e3	F81	root	19	0.2105	0.1053
L3f	F82	292 295	1	0.0000	0.0000
	F83	189 292	1	0.0000	0.0000
	F84	311	1	0.0000	0.0000
	F85	292	15	0.6667	0.2667
	F86	111A	1	0.0000	0.0000
	F87	Root	3	0.3333	0.3333
L3h	F88	179 215 256A 284	1	0.0000	0.0000
	F89	192 218 256A 362	1	0.0000	0.0000
	F90	179 256A 284	2	0.5000	0.5000
	F91	165 192 399	1	0.0000	0.0000
	F92	93 270	1	0.0000	0.0000
	F93	111 184 304 311	3	0.6667	0.4714
	F94	148 192	4	0.5000	0.3536
	F95	399	1	0.0000	0.0000
L3i	F96	260 311	10	0.4000	0.2449
	F97	153	2	0.0000	0.0000
	F98	260	1	0.0000	0.0000
L3x	F99	256 278 311	1	0.0000	0.0000
	F100	193 195 223	4	1.2500	0.9014
	F101	278 311	2	0.5000	0.5000
	F102	278	2	1.0000	0.7071
L4	F103	172 293T 399 355	1	0.0000	0.0000
	F104	293T 274 399 355	1	0.0000	0.0000
	F105	207T 220 260 261	2	0.0000	0.0000
	F106	293T 355 399	1	0.0000	0.0000
	F107	93 207T 260	1	0.0000	0.0000
	F108	207T 260 261	2	0.0000	0.0000
	F109	179 189 239 320	1	0.0000	0.0000
	F110	293T 355	1	0.0000	0.0000

	F111	207T 260	1	0.0000	0.0000
	F112	260	2	2.0000	0.5000
	F113	Root	1	0.0000	0.0000
L5	F114	111 254 311 335 360	1	0.0000	0.0000
	F115	355 362	5	0.2000	0.2000
L6	F116	173 362	12	0.0000	0.0000
	F117	362	1	0.0000	0.0000
	F118	Root	1	0.0000	0.0000

Table S 13 Founder lineages identified when using a *f1* criterion from Arabian Peninsula and Near East to North Africa and to East Africa separately.

f1			From Arabian Peninsula and Near East To North Africa			From Arabian Peninsula and Near East To East Africa		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**	**n**	**rho**	**se**
HV1a3	F1	67 183 327A	1	0.0000	0.0000			
HV1d	F2	67 278 362				4	0.2500	0.2500
H6	F3	67 129 368	1	0.0000	0.0000			
H6b	F4	104 300 362	1	0.0000	0.0000			
H	F5	86 129 189	1	0.0000	0.0000			
H1b	F6	189 356	3	0.6667	0.4714			
H	F7	189 293	2	0.0000	0.0000			
H8	F8	288 362	3	0.0000	0.0000			
H1b1b	F9	355 362	2	0.0000	0.0000			
H8c2	F10	153 362	1	0.0000	0.0000			
HV1	F11	67 354	1	0.0000	0.0000			
HV1a3	F12	67 327A	8	0.0000	0.0000	1	0.0000	0.0000
HV1a1	F13	67 355	1	0.0000	0.0000			
H5	F14	189 304	2	0.0000	0.0000			
H6b	F15	300 362	4	0.5000	0.5000			
H13b1	F16	261 262	6	0.0000	0.0000			
HV	F17	159	1	0.0000	0.0000			
HV	F18	245	1	0.0000	0.0000			
H	F19	266	2	0.0000	0.0000			
H39	F20	299	1	0.0000	0.0000			
H4	F21	287	3	1.3333	0.6667			
H	F22	92	1	0.0000	0.0000			
H	F23	293	2	1.0000	0.7071			
HV2	F24	217	2	0.0000	0.0000			
HV	F25	288	2	0.0000	0.0000			
H7c	F26	265	2	1.0000	0.7071			
HV	F27	114	1	0.0000	0.0000			
HV	F28	145	6	0.0000	0.0000			
H15a1b	F29	248	1	0.0000	0.0000			
H15a1b	F30	184	3	0.0000	0.0000			
HV	F31	357	13	0.9231	0.8496			
HV	F32	193	2	0.0000	0.0000			
HV	F33	243	3	1.3333	0.6667			
HV	F34	172	6	1.0000	0.5774			
HV	F35	355	5	0.4000	0.2828	1	0.0000	0.0000
HV	F36	260	4	0.7500	0.7500			
H3p	F37	222	2	0.5000	0.5000			
H13a1a6	F38	207	7	0.4286	0.3194			
HV	F39	147	5	0.4000	0.2828			
HV	F40	298	145	0.6759	0.1936	2	1.5000	0.8660
H2a1	F41	354	5	0.4000	0.2828			
HV	F42	290	1	0.0000	0.0000			
HV	F43	187	7	0.1429	0.1429			
HV	F44	153	8	0.7500	0.7500			
HV	F45	111	10	0.0000	0.0000			

H	F46	218	11	0.7273	0.4066	1	0.0000	0.0000
HV1	F47	67	22	1.0000	0.3278	12	1.4167	0.7407
HV	F48	221	4	1.7500	0.8292			
H5	F49	304	28	0.7500	0.2448			
H6	F50	362	7	0.5714	0.2857			
H	F51	261	5	1.0000	0.5292			
H1	F52	278	2	0.5000	0.5000			
H2a3	F53	274	7	0.1429	0.1429			
H	F54	256	14	1.2143	0.4684			
HV	F55	240	4	0.0000	0.0000			
HV	F56	192	4	0.7500	0.4330			
H	F57	189	27	0.2963	0.1960	1	0.0000	0.0000
H	F58	93	9	0.2222	0.2222			
HV	F59	86	2	0.0000	0.0000			
H	F60	239	3	0.0000	0.0000			
HV	F61	Root	465	0.4000	0.0763	8	0.5000	0.3953
M1a3	F62	223 311	3	0.6667	0.4714			
M1b2	F63	399	3	0.6667	0.6667			
M1a5	F64	129	3	1.3333	0.9428	24	0.8750	0.4390
M1b1	F65	185	10	0.4000	0.2449			
M1a3	F66	223	8	0.2500	0.1768			
M1a1	F67	359	45	0.7111	0.4309	36	0.7222	0.2664
M1a1c'd	F68	93 359				10	0.6000	0.3162
M1	F69	357				4	0.2500	0.2500
M1	F70	240				3	0.0000	0.0000
M1	F71	Root	46	0.1304	0.0532	13	0.7692	0.3264
N1a	F72	147G 172 248 263 266 355	1	0.0000	0.0000			
N1b1	F73	126 145 176G 390	1	0.0000	0.0000			
N1b1	F74	145 176G 362 390	1	0.0000	0.0000			
N1b1	F75	145 176G 311 390	9	1.0000	0.5092			
N1a1	F76	147A 172 248 320 355	1	0.0000	0.0000			
N1a1	F77	147A 172 248 355	3	1.0000	1.0000	16	0.1875	0.1398
N1a	F78	147G 172 213 248 355				7	0.0000	0.0000
N1a3	F79	201 265	1	0.0000	0.0000			
N1b1	F80	145 176G 390	15	0.6000	0.2000			
N1b	F81	176G 390	1	0.0000	0.0000	5	0.0000	0.0000
I5a	F82	148 129 294 391	1	0.0000	0.0000			
I5a	F83	129 148 391	1	0.0000	0.0000			
I1c	F84	129 311 319 391	1	0.0000	0.0000			
I1a	F85	129 172 311 391	2	0.0000	0.0000			
I	F86	93 129 391	3	0.3333	0.3333			
I1	F87	129 311 391	7	0.8571	0.4949			
I	F88	129 391	6	0.3333	0.2357	3	0.0000	0.0000
W	F89	292!	1	0.0000	0.0000			
W	F90	292 295	1	0.0000	0.0000			
W	F91	292	12	1.5833	0.4640	4	2.7500	0.9682
N2a	F92	153 319				4	2.2500	1.1456
R0a	F93	189 232A	2	0.0000	0.0000			
R0a1a	F94	185 355	2	0.0000	0.0000			
R0a	F95	145	3	0.0000	0.0000	1	0.0000	0.0000
R0a	F96	189	1	0.0000	0.0000	2	0.0000	0.0000
R0a2i	F97	92	2	0.0000	0.0000			
R0a1a	F98	355	13	0.9231	0.4615	9	0.2222	0.2222
R0a	F99	266	1	0.0000	0.0000	3	1.3333	1.0541
R0a2c	F100	304	4	0.0000	0.0000	1	0.0000	0.0000
R0a	F101	Root	26	0.3462	0.1490	12	0.5000	0.3118
R0a	F102	93				1	0.0000	0.0000
R0a	F103	114				3	0.0000	0.0000
R0a2b	F104	305T				12	0.0000	0.0000

R0a1a1a	F105	172 184A				1	0.0000	0.0000
T1a	F106	163 172 186 189 298	3	1.3333	0.9428			
T2	F107	146 292 296!	1	0.0000	0.0000			
T2	F108	146 147 292 296	5	0.4000	0.4000			
T2	F109	288 292 296				2	0.0000	0.0000
T1a	F110	163 186 189 287	1	0.0000	0.0000			
T1a	F111	163 186 189 271	1	0.0000	0.0000			
T1a	F112	163 186 189 390	1	0.0000	0.0000			
T1b	F113	163 189 243				1	0.0000	0.0000
T2b	F114	209 296 304	1	0.0000	0.0000			
T2b	F115	292 296 304	1	0.0000	0.0000	1	0.0000	0.0000
T2b	F116	296! 304	1	0.0000	0.0000			
T2c1c	F117	146 292 296	4	0.7500	0.4330			
T2c1	F118	292 296!	9	0.5556	0.2940	3	0.0000	0.0000
T2c1c	F119	146 296!	2	0.0000	0.0000			
T1a	F120	163 186 189	69	1.3043	0.3672	5	1.2000	0.6325
T2e	F121	153 296				1	0.0000	0.0000
T2k	F122	291 296	1	0.0000	0.0000			
T2	F123	256 296	5	0.4000	0.2828			
T2f	F124	189 296	1	0.0000	0.0000			
T2b	F125	296 304	30	0.5667	0.1795			
T2a1b	F126	296 324	5	0.4000	0.2828			
T2c1	F127	292 296	12	0.8333	0.4410	1	0.0000	0.0000
T2	F128	296!	5	0.8000	0.4000			
T1	F129	163 189	3	2.0000	1.1547			
T2	F130	296	17	1.0000	0.4441	1	0.0000	0.0000
J	F131	51 188 311	4	0.2500	0.2500	1	0.0000	0.0000
J2a1a1	F132	145 231 261	7	0.0000	0.0000			
J1d1	F133	129 193 300	1	0.0000	0.0000			
J1d1a	F134	193 300 309	14	0.3571	0.1890	4	0.2500	0.2500
J1b	F135	145 222 261!	1	0.0000	0.0000			
J1b	F136	145 222 261 311	3	0.0000	0.0000			
J1b2a	F137	136 145 222 261	2	0.0000	0.0000			
J1b1a1	F138	145 172 222 261	1	0.0000	0.0000			
J	F139	231 319	1	0.0000	0.0000			
J	F140	69 145	2	0.0000	0.0000			
J2	F141	241 311	1	0.0000	0.0000			
J1d1	F142	193 300	2	0.0000	0.0000			
J1b	F143	145 261 355	1	0.0000	0.0000			
J1b	F144	129 145 261	1	0.0000	0.0000			
J1b	F145	145 222 261	7	0.7143	0.4286	1	0.0000	0.0000
J1b1a1	F146	145 172 261	1	0.0000	0.0000			
J	F147	271	1	0.0000	0.0000			
J	F148	256	2	0.0000	0.0000			
J	F149	231	2	1.5000	1.1180	1	0.0000	0.0000
J	F150	69	3	1.0000	0.5774			
J	F151	362	3	0.3333	0.3333			
J2a2b	F152	241	17	0.2353	0.1176			
J1d	F153	193	19	1.6842	0.5766			
J1c7	F154	261	2	0.0000	0.0000			
J1b	F155	145 261	3	0.0000	0.0000	1	0.0000	0.0000
J	F156	Root	54	0.5741	0.1493	2	0.5000	0.5000
U5b1b1b	F157	189 192 320	2	1.0000	0.7071			
U5a1b1	F158	291 256 399	4	0.7500	0.7500			
U5a1	F159	192 256 399	1	0.0000	0.0000			
U5b2a1	F160	189 270	3	0.0000	0.0000			
U5	F161	145 192	1	0.0000	0.0000			
U5b1b1	F162	189 192	12	0.8333	0.2635			
U5a1	F163	256 399	3	2.0000	0.8165			
U5a	F164	192 256	3	0.3333	0.3333			
U5b3	F165	304	2	0.5000	0.5000			
U5b2a	F166	189	12	0.5833	0.2205			

U5	F167	192	26	0.4615	0.2176	2	2.0000	1.0000
U5a	F168	256	8	1.1250	0.5449			
U5	F169	Root	3	2.0000	0.8165			
U6a1a	F170	172 189 219 278 295	8	0.0000	0.0000			
U6a1a	F171	147 172 189 219 278	2	0.0000	0.0000			
U1a	F172	129 189 249 288 362	2	0.5000	0.5000			
U2e2	F173	51 92 129C 189 362	1	0.0000	0.0000			
U6b'd	F174	172 219 311 362	1	0.0000	0.0000			
U6a1b	F175	219 172 235 278	8	1.5000	0.7071			
U6a2'3	F176	172 189 219 278	49	0.8163	0.3487	15	0.6000	0.3197
U2b2	F177	51 209 239 352 353	1	0.0000	0.0000			
U2e	F178	51 129C 189 362	4	0.0000	0.0000			
U2d	F179	51 189 234 294	1	0.0000	0.0000	2	1.0000	0.7071
U6b'd	F180	172 311 219	5	0.2000	0.2000			
U6a	F181	172 219 278	40	0.7500	0.2179			
U1a	F182	189 249 362	1	0.0000	0.0000			
U1a	F183	189 249 311	1	0.0000	0.0000			
U2b2	F184	51 352 353	1	0.0000	0.0000			
U2	F185	51 189 362	2	0.0000	0.0000			
U7	F186	309 318T 362	1	0.0000	0.0000			
U3a	F187	260 343 390	2	0.0000	0.0000			
U3b3	F188	168 343 355	2	0.0000	0.0000			
K1	F189	93 189 224 311	1	0.0000	0.0000			
U6a'b'd	F190	172 219	7	0.8571	0.5345			
U1a'c	F191	189 249	6	1.5000	0.7993			
U1b	F192	111 214A 249 327				1	0.0000	0.0000
U7	F193	318C 318T	1	0.0000	0.0000			
U7	F194	309 318T	3	1.6667	0.8819			
U4c2a	F195	261 356	3	0.3333	0.3333			
U4a2b	F196	223 356	1	0.0000	0.0000			
U3c	F197	193 343	2	2.5000	1.3229			
U3	F198	104 343	2	0.0000	0.0000			
U3a	F199	343 390	8	0.8750	0.4146			
U3b3	F200	168 343	1	0.0000	0.0000	1	0.0000	0.0000
U8b1b	F201	189 234 324	1	0.0000	0.0000			
K2c	F202	210 224 311	2	0.5000	0.5000			
K	F203	167 224 311	1	0.0000	0.0000			
K	F204	158 224 311	1	0.0000	0.0000			
K	F205	224 290 311	3	0.0000	0.0000			
K	F206	224 242 311	2	1.5000	0.8660			
K	F207	224 278 311	1	0.0000	0.0000	1	0.0000	0.0000
K1b1a	F208	224 311 319	2	1.5000	0.8660			
K1	F209	224 234 311	3	0.0000	0.0000			
K	F210	224 311 360	1	0.0000	0.0000			
K1	F211	224 311 362	2	0.0000	0.0000			
K1a27	F212	176 224 311	2	0.0000	0.0000			
K	F213	189 224 311	3	0.0000	0.0000	1	0.0000	0.0000
K1	F214	93 224 311	15	0.1333	0.0943	2	0.0000	0.0000
K	F215	224 304 311	2	0.5000	0.5000			
U6	F216	172	4	2.7500	1.3463			
U1	F217	249	1	0.0000	0.0000			
U2	F218	51	1	0.0000	0.0000			
U9a	F219	51 278	1	0.0000	0.0000	1	0.0000	0.0000
U7	F220	318T	2	2.0000	1.0000			
U4	F221	356	13	0.6154	0.3077			
U3	F222	343	21	1.0476	0.2935	1	0.0000	0.0000
K	F223	224 311	74	0.6216	0.1324	8	0.1250	0.1250
X1	F224	104 278	7	0.0000	0.0000			
X1	F225	104				1	0.0000	0.0000
X2j	F226	179	4	1.0000	1.0000			
X2h	F227	223	3	1.6667	1.0000			
X4	F228	274	1	0.0000	0.0000			

X2	F229	248	1	0.0000	0.0000			
X	F230	Root	35	0.6571	0.1959	3	0.6667	0.4714

Table S 14 Founder lineages identified when using a *f2* criterion from Arabian Peninsula and Near East to North Africa and to East Africa, separately.

f2			From Arabian Peninsula and Near East To North Africa			From Arabian Peninsula and Near East To East Africa		
Clade	**Founder**	**Sequence**	**n**	**rho**	**se**	**n**	**rho**	**se**
H1b	F1	189 356	3	0.6667	0.4714			
HV1a1	F2	67 355	1	0.0000	0.0000			
H6	F3	300 362	5	0.8000	0.4899			
H	F4	266	2	0.0000	0.0000			
H4	F5	287	3	1.3333	0.6667			
H	F6	92	1	0.0000	0.0000			
H	F7	293	4	1.2500	0.6614			
HV2	F8	217	2	0.0000	0.0000			
H7c	F9	265	2	1.0000	0.7071			
HV	F10	114	1	0.0000	0.0000			
HV	F11	145	6	0.0000	0.0000			
H15a1b	F12	248	1	0.0000	0.0000			
HV	F13	243	3	1.3333	0.6667			
HV	F14	172	6	1.0000	0.5774			
HV	F15	355	7	0.5714	0.3499	1	0.0000	0.0000
HV	F16	260	4	0.7500	0.7500			
H3p	F17	222	2	0.5000	0.5000			
HV	F18	298	145	0.6759	0.1936	2	1.5000	0.8660
H2a1	F19	354	5	0.4000	0.2828			
HV	F20	290	1	0.0000	0.0000			
HV	F21	187	7	0.1429	0.1429			
HV	F22	153	9	0.8889	0.6849			
H	F23	218	11	0.7273	0.4066	1	0.0000	0.0000
HV1	F24	67	33	1.3636	0.4318	17	1.7059	0.6308
H5	F25	304	30	0.7667	0.2380			
H6	F26	362	7	0.5714	0.2857			
H	F27	261	11	2.0909	0.9749			
H2a3	F28	274	7	0.1429	0.1429			
H	F29	256	14	1.2143	0.4684			
HV	F30	240	4	0.0000	0.0000			
HV	F31	192	4	0.7500	0.4330			
H	F32	189	27	0.2963	0.1960	1	0.0000	0.0000
H	F33	93	9	0.2222	0.2222			
HV	F34	86	3	1.0000	0.5774			
H	F35	239	3	0.0000	0.0000			
HV	F36	root	519	0.5318	0.0823	8	0.5000	0.3953
M1a3	F37	223 311	3	0.6667	0.4714			
M1a3	F38	223	8	0.2500	0.1768			
M1a1	F39	359	45	0.7111	0.4309	46	0.9130	0.3090
M1	F40	root	62	0.5161	0.1922	44	1.4318	0.6145
N1b1	F41	126 145 176G 390	1	0.0000	0.0000			
N1a1	F42	147A 172 248 320 355	1	0.0000	0.0000			
N1a1	F43	147A 172 248 355	4	1.2500	0.8292	23	0.4348	0.3195
N1a3	F44	201 265	1	0.0000	0.0000			
N1b1	F45	145 176G 390	25	1.1200	0.4233			
I5a	F46	129 148 391	2	1.0000	0.7071			
I1	F47	129 311 391	7	0.8571	0.4949			
I	F48	129 391	12	1.0833	0.3997	3	0.0000	0.0000
N1	F49	Root	1	0.0000	0.0000	5	0.0000	0.0000
W	F50	292	14	1.7857	0.4345	4	2.7500	0.9682

N2a	F51	153 319				4	2.2500	1.1456
R0a1a	F52	185 355	2	0.0000	0.0000			
R0a	F53	189	3	0.6667	0.6667	2	0.0000	0.0000
R0a1a	F54	355	13	0.9231	0.4615	9	0.2222	0.2222
R0a	F55	114				3	0.0000	0.0000
R0a2c	F56	304	4	0.0000	0.0000	1	0.0000	0.0000
R0a	F57	Root	32	0.5625	0.1926	30	1.0000	0.4497
T2	F58	146 292 296 296	1	0.0000	0.0000			
T2b	F59	296! 304	1	0.0000	0.0000			
T2	F60	288 292 296				2	0.0000	0.0000
T2c1c	F61	146 292 296	9	1.1111	0.6285			
T2c1	F62	292 296 296	9	0.5556	0.2940	3	0.0000	0.0000
T1a	F63	163 186 189	75	1.4533	0.3485	5	1.2000	0.6325
T2k	F64	291 296	1	0.0000	0.0000			
T2b	F65	296 304	32	0.6563	0.1795	1	0.0000	0.0000
T2c1	F66	292 296	12	0.8333	0.4410	1	0.0000	0.0000
T2	F67	146 296	2	0.0000	0.0000			
T2	F68	296 296	5	0.8000	0.4000			
T2e	F69	153 296				1	0.0000	0.0000
T1	F70	163 189	3	2.0000	1.1547	1	0.0000	0.0000
T2	F71	296	28	1.1429	0.3780	1	0.0000	0.0000
J2a1a1	F72	145 231 261	7	0.0000	0.0000			
J1d1a	F73	193 300 309	14	0.3571	0.1890	4	0.2500	0.2500
J1b	F74	145 222 261 261	1	0.0000	0.0000			
J1b2a	F75	136 145 222 261	2	0.0000	0.0000			
J1d1	F76	193 300	3	0.3333	0.3333			
J1b	F77	145 222 261	11	0.8182	0.3963	1	0.0000	0.0000
J1b1a1	F78	145 172 261	1	0.0000	0.0000			
J	F79	231	3	1.6667	0.8819	1	0.0000	0.0000
J	F80	69	5	1.0000	0.5292			
J2a2b	F81	241	18	0.2778	0.1242			
J1d	F82	193	19	1.6842	0.5766			
J1b	F83	145 261	5	0.6000	0.3464	1	0.0000	0.0000
J	F84	Root	66	0.8485	0.1780	3	1.3333	0.6667
U5a1	F85	192 256 399	1	0.0000	0.0000			
U5b2a1	F86	189 270	3	0.0000	0.0000			
U5a1	F87	256 399	7	1.8571	0.7954			
U5a	F88	192 256	3	0.3333	0.3333			
U5b2a	F89	189	14	1.0000	0.3030			
U5	F90	192	39	0.9231	0.3516	2	2.0000	1.0000
U5a	F91	256	8	1.1250	0.5449			
U5	F92	Root	5	2.0000	0.6928			
U2e2	F93	51 92 129C 189 362	1	0.0000	0.0000			
U2b2	F94	51 209 239 352 353	1	0.0000	0.0000			
U2e	F95	51 129C 189 362	4	0.0000	0.0000			
U6a	F96	172 219 278	107	1.4860	0.5284	15	1.6000	1.0499
U5b2c2b	F97	189 249 288	2	3.5000	1.8028			
U2b2	F98	51 352 353	1	0.0000	0.0000			
U2	F99	51 189 362	2	0.0000	0.0000			
U3b3	F100	168 343 355	2	0.0000	0.0000			
U6a'b'd	F101	172 219	13	1.0769	0.5547			
U1a'c	F102	189 249	8	1.3750	0.6250			
U1b	F103	111 214A 249 327				1	0.0000	0.0000
U2c'd	F104	51 234	1	0.0000	0.0000	2	3.0000	1.5811
U7	F105	318C 318T	1	0.0000	0.0000			
U7	F106	309 318T	4	1.5000	0.7071			
U3c	F107	193 343	2	2.5000	1.3229			
U3a	F108	343 390	10	0.9000	0.3873			
U3b3	F109	168 343	1	0.0000	0.0000	1	0.0000	0.0000
U8b1b	F110	189 234 324	1	0.0000	0.0000			

K	F111	167 224 311	1	0.0000	0.0000			
K1	F112	93 224 311	16	0.1875	0.1083	2	0.0000	0.0000
U1	F113	249	1	0.0000	0.0000			
U2	F114	51	1	0.0000	0.0000			
U9a	F115	51 278	1	0.0000	0.0000	1	0.0000	0.0000
U7	F116	318T	2	2.0000	1.0000			
U4	F117	356	17	0.7647	0.3057			
U3	F118	343	23	1.0435	0.2818	1	0.0000	0.0000
K	F119	224 311	98	0.9388	0.1414	10	0.5000	0.2236
U	F120	Root	4	2.7500	1.3463			
X1	F121	104 278	7	0.0000	0.0000			
X2	F122	248	1	0.0000	0.0000			
X	F123	Root	43	0.9302	0.2302	4	1.2500	0.5590