UNIVERSITY OF TARTU
FACULTY OF BIOLOGY AND GEOGRAPHY, INSTITUTE OF MOLECULAR AND CELL BIOLOGY, DEPARTMENT OF EVOLUTIONARY BIOLOGY

Piia Serk

HUMAN MITOCHONDRIAL DNA HAPLOGROUP J IN EUROPE AND NEAR EAST

M.Sc. Thesis

Supervisors: Ph.D. Ene Metspalu,
Prof. Richard Villems

Tartu
2004

Table of contents

Abbreviations

(Y)BP	(years) before present
CR	Control Region
CRS	Cambridge Reference Sequence
D-loop	displacement loop (= control region) of mtDNA
HVSI	the first hypervariable segment of the control region of mitochondrial genome
HVSII	the second hypervariable segment of the control region of mitochondrial genome
LHON	Leber hereditary optic neuropathy
MRCA	most recent common ancestor
mtDNA	mitochondrial DNA
np(s)	nucleotide position
RFLP	Restriction Fragment Length polymorphism

Definition of basic terms used in the thesis

haplotype	a sequence type that comprises all identical sequences
haplogroup	a group of haplotypes that share a common ancestor defined by an array of synapomorphic substitutions
lineage	any array of characters/mutations shared by more than one haplotype
star-like tree	a set of sequences is said to have a pattern of star-like phylogeny if most (ideally all of them) coalesce to one and the same haplotype (that has not necessarily been observed in the sample)
expansion time	coalescence
coalescence	coalescence time calculated to the founder that displays star-like phylogeny

Introduction

History, environmental variations, and cultural configurations dictated many of the outcomes, some of which played a fundamental role in the large-scale genesis of human cultural and biological patterning from the Neolithic/Formative times into the world of today. Variability of human mitochondrial DNA has provided valuable data about the genetic past of human maternal lineages**.** Analyses of the frequency, variation and distribution of mitochondrial DNA haplotypes have been used to evaluate current models concerning the process of colonization of the World. By now, there is already a substantial amount of mitochondrial evidence for an African exodus of humans, peopling of Eurasia, Australia and the Americas, as well as of the Pacific**.** Analysis of mitochondrial DNA haplotypes is, however, not only limited to human origin and evolution. It is also important for studies of human pathologies, many studies have shown that mitochondrial DNA polymorphisms can play an important role in modulating disease expression.

My research concentrates on mitochondrial DNA haplogroup J in Europe and in Near East. Haplogroup J offered interest in many aspects. Firstly, according to previous studies, the spread of haplogroup J was stated as a corner stone for the argument that haplogroup J was brought to Europe by the first farming and herding societies form Near East during Neolithic times. Secondly, polymorphisms specific to J haplogroup are known to be associated with a mitochondrial disease – Leber hereditary optic neuropathy. Elucidation of J haplogroup's genealogy would hopefully benefit exploring both of these hypotheses.

Literature overview

West–Eurasian mtDNA tree

One of the great virtues of mitochondrial DNA (mtDNA) is the opportunity it provides for detailed estimation of the human maternal genealogy. Usefulness of mtDNA molecule in studying humanity's demographic history arises from its large copy number, strictly maternal inheritance (Giles et al. 1980) and general homoplasmy. MtDNAs rarely, if ever mix and recombine (Elson et al. 2001; Ingman et al. 2000). Human mtDNA studies could be broadly categorized on the basis of how much of the molecule was assayed. Human mtDNA was fully sequenced already more than 20 years ago (Anderson et al. 1981), making it possible to design strategies to study its variation in human populations. Historically "low-restriction analysis" with 5 or 6 restriction enzymes, was used to reveal variation in mtDNA pool whereas later on "high-resolution analysis" included already 12 or 14. Sequencing studies have mostly been focused on the control region (CR), or alternatively called displacement loop (D-loop), of the molecule. CR includes first hypervariable segment (HVSI), the region between nucleotide positions (np) 16024–16383 numbered in according to the so-called Cambridge Reference Sequence (Anderson et al. 1981), and hypervariable segment two (HVSII), the region between 73–500 np. Generally, HVSII segment is considered to be less informative (Aris-Brosou and Excoffier 1996; Hasegawa et al. 1993; Ingman et al. 2000) and therefore so far has not been used widely in phylogenetic studies.

During the early 1990s a refined picture of mtDNA phylogeny started slowly to emerge with the application of high-resolution restriction fragment length polymorphism (RFLP) analysis of mtDNA from one continent at a time (Ballinger et al. 1992; Chen et al. 1995; Schurr et al. 1990; Torroni et al. 1996). These analyses carried out first of all in Douglas Wallace' laboratory in Emory, Atlanta, revealed that mtDNAs could be classified into a small number of monophyletic clades, or haplogroups. Defined by one or several unique for them ("diagnostic") restriction sites – i.e. loss or gain of such sites, relative to CRS. For example 175 mtDNA chromosomes from American individuals of European ancestry were analyzed by Antonio Torroni in 1994 (Torroni et al. 1994), revealing 117 haplotypes (lineages). Related haplotypes were gathered into groups (haplogroups), haplogroups were denoted by uppercase roman

letters. Four European-specific haplogroups were determined (H, I, J, K) and were defined as follows. Haplogroup J was characterised by the *Bst*NI 13704 restriction site (G to A transition at np 13708) and the *Hinf*I 16065 site (C to T transition at np 16069) losses. Haplogroup H has lost an *Alu*I restrition site at np 7028. Haplogroup I was defined by the *Dde*I np 1715 and the *Hae*II restriction site losses at np 4529 and the *Alu*I site gain at np 10028. In addition to that by a combined *Ava*II np 8249 site gain and the *Hae*III np 8250 site loss, and a combined *Bam*HI/*Mbo*I np 16389 site gain and the *Ava*II np 16390 site loss. Haplogroup K was delineated by a combined *Hae*II np 9052/*Hha*I np 9053 site loss. Further RFLP analysis of 49 Finnish, 37 Swedish and 48 Tuscan mtDNAs revealed the presence of additional European haplogroups that were named as T, U, V, X and W (Torroni et al. 1996). Haplogroup T was defined by the gain of the *Bam*HI and the *Alu*I restriction sites at nps 13366 and 15606. Haplogroup U was defined by the gain of the *Hin*fI site at np 12308. Haplogroup V was defined by the loss of the *Nla*III site at np 4577. Haplogroup W was characterised by the gain of the *Ava*II np 8249 site and the loss of the *Hae*III np 8994 site. Haplogroup X was characterised by the *Dde*I site loss at np 1715.

One of the earliest studies that used sequence variation in the control region to study the hypothesis that all mtDNA types in contemporary humans stem from a common African ancestor was performed in 1991 (Vigilant et al. 1991). This paper, from Alan Wilson's group, was an extension and further refinement of the famous Nature paper of the same group from 1987 (Cann et al. 1987) – the paper that was used to coin"African Eve" label. In other words, a paper, where a genetic argument in favour of the "Recent Out-of-Africa" theory was put forward. In sake of correctness, it is, however, fair to indicate that at least one much earlier mtDNA study, suggested essentially the same (Brown 1980).

There are at least two wide-spread theories about what happened after archaic pre-sapiens humans (*Homo erectus*, possibly, at least partially, also *Homo ergaster*) spread over out of Africa, presumably over southern Eurasia, a few millions of years ago. The first theory insists that, these anatomically archaic populations have gradually evolved into the modern *Homo sapiens sapiens* essentially locally, with a limited gene flow between the populations. This is the "multiregional theory" of human evolution and it is based, first of all, on morphological markers (Wolpoff et al. 1984). There are

several sub-variants of this theory, largely evolving towards allowing extensive long-distance gene flows ("trellis" models). Secondly, the "Out-of-Africa" theory of human evolution states that all modern human populations have descended from an anatomically modern group that arouse in Africa some 200 000 years before present (YBP). And later on, a sub-fraction of them left Africa and replaced all, presumably different archaic populations that may have still inhabited various geographic localities at that time (Stringer and Andrews 1988).

Two hypervariable segments of mtDNA were sequenced by Vigilant et al. (cit. above) from 189 individuals of diverse geographic origin in order to study the proposal that all mtDNA types in contemporary humans coalesce to a common ancestor present in an African population some 200 000 years ago. Coalescence is the process by which the individuals in the present time trace back in the genealogy to common ancestors (the numbers of which are given for each generation), eventually reaching the single most recent common ancestor (MRCA). 201 polymorphic sites were found. Each unique sequence was termed a mtDNA type (haplotype); these 201 polymorphic sites defined 135 types among the 189 individuals. An African origin for human mtDNA was supported by several findings. Firstly, mtDNA sequence variation was found to be considerably higher among the Africans than among the Asians or the Europeans. Secondly, topology of the resulting genealogical tree had the deepest brancing point between two clusters, where one such cluster encompassed mtDNA chromosomes, exclusively of African origin, whereas the other cluster included all Eurasian plus many African mtDNAs. Such a result can be considered as a strong hint in favour of the recent "Out-of-Africa" model.

The genealogical resolution (i.e. a detailed topology of the corresponding tree) is of an essential importance in the deciphering human mtDNA phylogeny. Nearly 30 % of the polymorphic sites are confined to the control region, which represents a mere 7 % of the genome (Ingman and Gyllensten 2001). High-resolution RFLP analysis screens about 20 % of a mitochondrial genome (Torroni et al. 1994). If the resolution is shallow, the underlying genealogy may be so poorly estimated that demographic and historical interpretations can come unstuck. Actually all studies of human evolution based on mtDNA sequencing or RFLP data are complicated by the variation in substitution rate between sites, and consequence as a of parallel and putative back

mutations (reversions) that cause difficulties in the estimation of genetic distances and make phylogenetic inferences difficult (Maddison 1991). Thus, one of the problems confusing phylogenetic reconstructions is raised by particularly fast mutation rates - multiple hits - on some of the sites. This results in a possibility of drawing, at least theoretically, millions of most parsimonious phylogenetic trees from a data set of approximately 100 mtDNA HVSI sequences (Cann et al. 1987; Templeton 1992; Vigilant et al. 1991).

This problem was at least partly resolved by summarising many likely trees in a graph, called phylogenetic network. In ideal a complete phylogenetic network is a graph that displays all alternative potential evolutionary paths. In other words it is a minimum spanning tree for a set of sequence types that connects all given types without creating any cycles such that the total length (the sum on distances between linked sequence types) would be minimal (Bandelt et al. 1995; Excoffier and Smouse 1994). It is achieved by algorithms that rely on sequential split decompositions of each informative character in the sequence matrix or on a sequential introduction of inner branches between components of tightly connected nodes (Bandelt et al. 1999).

The principles of cladistic nomenclature of mtDNA tree branches were established (Macaulay et al. 1999; Richards et al. 1998). Haplotype stands for a sequence type that comprises all identical mitochondrial nucleotide sequences. Haplogroup means a group of haplotypes that share a common ancestor defined by an array of synapomorphic substitutions. Therefore, mtDNA haplogroups are supposed to be monophyletic and, applying cladistic terminology, just *clades*. Lineage stands for any array of characters or mutations shared by more than one haplotype. Major haplogroups are denoted by uppercase roman letters. The haplogroups can be nested; for example, RFLP haplogroups C, D, E and G are contained in M (i.e. see (Torroni et al. 1994)). Successively nested haplogroups are named by alternating positive integers and lowercase roman letters; for example J1b1 $\subset$ J1b $\subset$ J1 $\subset$ J, where "$\subset$" means "is a subhaplogroup of ". A haplogroup that is composed of a set of named subhaplogroups is referred to by concatenating the subhaplogroup names; for example, the smallest haplogroup that includes J and T is called "JT". To designate a set of mtDNAs – in general, not a clade itself – that coalesce in an unresolved multifurcation but that are not

members of any of the clusters branching from that node, an asterisk (*) to the list of clusters was appended (Macaulay et al. 1999).

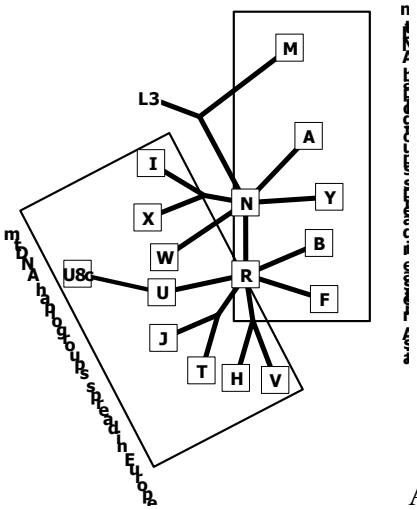


Figure 1. A tree relating mtDNA haplogroups spread in Eurasia. The root is specified by an African node L3.

At present, mtDNA network for European populations is established combining coding region and HVSI data. Arising from an African haplogroup L3, two mtDNA macrolineages (superhaplogroups) M (Chen et al. 1995) and N (Alves-Silva et al. 2000; Quintana-Murci et al. 1999) cover the variation of maternal lineages outside Africa. L3 superhaplogroup is defined by loss of *Hpa*I restriction site at np 3592. All Eurasian, Oceanian and Amerindian lineages also lack this restriction site and have arisen from L3 (Chen et al. 1995; Kivisild et al. 1999). In contrast to haplogroup M, which is spread in South and East Eurasia and Oceania while being virtually absent in Europe, though present in Africa as M1, the phylogenetic node N (including R) has spread its branches all over Eurasia. Most branches of it are, by and large, specific to either eastern [i.e. A, Y, B, R9 (Kivisild et al. 2002)] or western [i.e. N1, TJ, HV, U, R1 (Macaulay et al. 1999)] parts of the continent (Figure 1).

Fast mutation rate of mtDNA

MtDNA diverges at the rate of 2-4% per site per million years (Cann et al. 1987; Torroni et al. 1994), which is on the average 10 to 100 fold faster than the rate in the nuclear genome. The high substitution rate has been attributed to the lack of proofreading activity in mitochondrial DNA polymerase and because of a high concentration of oxydative radicals inside mitochondria**.** It is well known by now that mutation (substitution) rates in the human mitochondrial D-loop are highly heterogeneous. Initially, two segments of the control region were noted to be

particularly variable and were thus called as hypervariable regions (Vigilant et al. 1991). Transitions at sites like 16093, 16129, 16209, 16311 and 16362 in HVSI and 146, 150 and 152 in HVSII occur frequently in many different phylogenetic contexts and these sites can be considered as "mutational hotspots" (Hasegawa et al. 1993; Ingman et al. 2000; Stoneking 2000; Wakeley 1993). The variation of mutation rate is higher in HVSII, where one finds a few sites where substitutions occur very often (observed frequently in different lineages), while most of HVSII shows rather little sequence variation (Aris-Brosou and Excoffier 1996). That is why HVSII segment is usually considered to be less informative than HVSI and has gained little attention in mtDNA studies resolving humanity's past. Nevertheless, HVSII has provided molecular markers, useful to resolve haplogroup X' phylogeny (Reidla et al. 2003)**.** In addition to "mutational hotspots" there are some polymorphisms in HVSII that can not be used in phylogenetic analyses. Firstly, transition (G>A) at np 263 represents rare polymorphisms in the CRS mtDNA (Andrews et al. 1999). Secondly, transition (A>G) at np 73 has shared anchestry in haplogroups U, K, T, J, I, W, X , and Z (e.g. [Torroni, 1996; Macaulay et al. 1999), thus can not be regarded as a defining polymorphism. Thirdly, there is a tandem repetitive polycytosine tract, from np 302 to 309 in HVSII known to have a high rate of mutations that change the repeat length (Hauswirth et al. 1985). Insertion of at least one cytosine in this hypermutable C-tract has been reported (Crespillo et al. 2000; Sigurđardóttir et al. 2000).

Estimation of a coalescence time

Major events in demographic history of population are often dated by estimating the time since the beginning of their expansion. To infer time scale of such events, analysis of the variation of DNA sequences from the mitochondrial genome pool and of the non-recombining part of the Y chromosome are particularly useful because they are passed on from mother to daughter and father to son without recombination Consequently, these sequences can be traced back directly to the genealogical maternal or paternal MRCAs. Therefore, an entire population (here, more precisely, its mtDNA and Y-chromosomal pools) can be traced back to a (*pro:* respective) common ancestor(s), and the genealogy of the population (of the respective pools) is said to coalesce (Hartl et al. 1989). The time at which it occurs is the coalescence time.

Coalescent theory tells us what gene genealogies are expected to look like if populations have different demographic histories – that is, how genealogies are affected by changes in population size and structure. Populations may be reduced dramatically in size and subsequently recover (a "bottleneck effect"). In such case, the MRCA may be recent in a small population that remained constant in size. However, in a population that has expanded, many lineages will coalesce in a rather short time frame after the "bottleneck" and thus the genealogy will be star-like. In contrast, in a population of constant size, the two oldest lineages will on average need as much time to coalesce as all the other lineages and the genealogy of the lineages will show a deep "split". A standard approach is to derive a phylogenetic tree and then date branch lengths by reference to an assumed molecular clock, calibrated by means of "external" dates, usually derived from established fossil records or, in case of human societies, even from known archaeological/historical knowledge. Calibration is necessary to translate time estimates expressed as a number of mutations, to calendar years. For human mtDNA trees, palaeontological or archaeological dating for the divergence between human and chimpanzee, the settling of the New World, New Guinea or Australia, have been used for calibration (i.e. (Cann et al. 1987; Forster et al. 1996; Ingman et al. 2000; Torroni et al. 1994; Vigilant et al. 1991)). Taking the divergence estimate for the split between the ancestors of humans and chimpanzees in between 4 and 6 million years (Hasegawa et al. 1990), human mtDNA tree is estimated to coalesce about 150000 years ago (Cann et al. 1987; Ingman et al. 2000). The coalescence time has been estimated using mean pairwise difference between mtDNAs, or the average mutational distance from the root of a clade, designated by ρ. In the latter case, the coalescence time estimate of a tree depends upon the topology of that tree. If such a phylogeny followes a Poisson process, a tree would be perfectly star-like (Saillard et al. 2000). Star-likeliness is defined as the proportion of pairs of sequences in a tree that coalesce at the root node (Torroni et al. 1998). The more star-like a tree is, the more exact time estimate can be obtained (Saillard et al. 2000). That, however, does not solve problems arising from the calibration of molecular clock.

Topology of mtDNA haplogroup J

The first studies with high-resolution restriction mapping divided global mtDNA variation into a number of major ancient clades, called haplogroups (for a review see (Wallace 1995). Haplogroup J was first described in 1994 by Torroni and colleagues (Torroni et al. 1994), it was characterized by the 13704 *Bst*NI site and the 16065 *Hinf*I site losses.

The first study that revealed some inner features of haplogroup J structure was performed by Torroni and et al. in 1997. 37 Italian subjects affected by Leber hereditary optic neuropathy (LHON) were screened for most of the mutations that were known to be associated with LHON at that time. LHON is a maternally transmitted disease in which the primary clinical manifestation is acute or subacute bilateral loss of central vision, leading to blindness. Approximately 90% of LHON cases are associated with and likely directly caused by mtDNA mutations at nps 3460, 11778 or 14484. These are designated as "primary" mutations because they impart a high risk for LHON expression [Wallace, 1988 #957; Howell et al. 1991; Johns et al. 1992; 1993). Mutation at np 3460 is distributed randomly along the phylogenetic tree, without any preferential association with the nine haplogroups (H, I, J, K, T, U, V, W, and X) that characterize European populations, whereas the mutations at nps 11778 and 14484 show a strong preferential association with haplogroup J. Based on RFLP data of LHON affected individuals belonging to haplogroup J, a phylogenetic network was constructed. Haplogroup J was divided into two subclusters, defined the *Alu*I np 7474 (transition C>T at np 7476) and the *Acc*I np 15257 (transition G>A at np 15257) restriction site losses (Torroni et al. 1997).

More detailed phylogenetic networks for European mtDNA were constructed by use of sequence data from HVSI (Richards et al. 1998). Skeleton network for European mitochondrial phylogenetic structure was constructed (Figure 2), based on HVSI data, but also including informative coding region variation, established earlier (Torroni et al. 1996). The major founder subcluster, J*, harboring HVSI mutations 16069 (C>T) and 16126 (T>C) was defined. Several HVSI polymorphisms were found to be haplogroup J specific, i.e. an array of transitions comprising nps 16145 (G>A), 16193 (C>T), 16222 (C>T), 16231 (T>C) and, 16261 (C>T) (Richards et al. 1998).

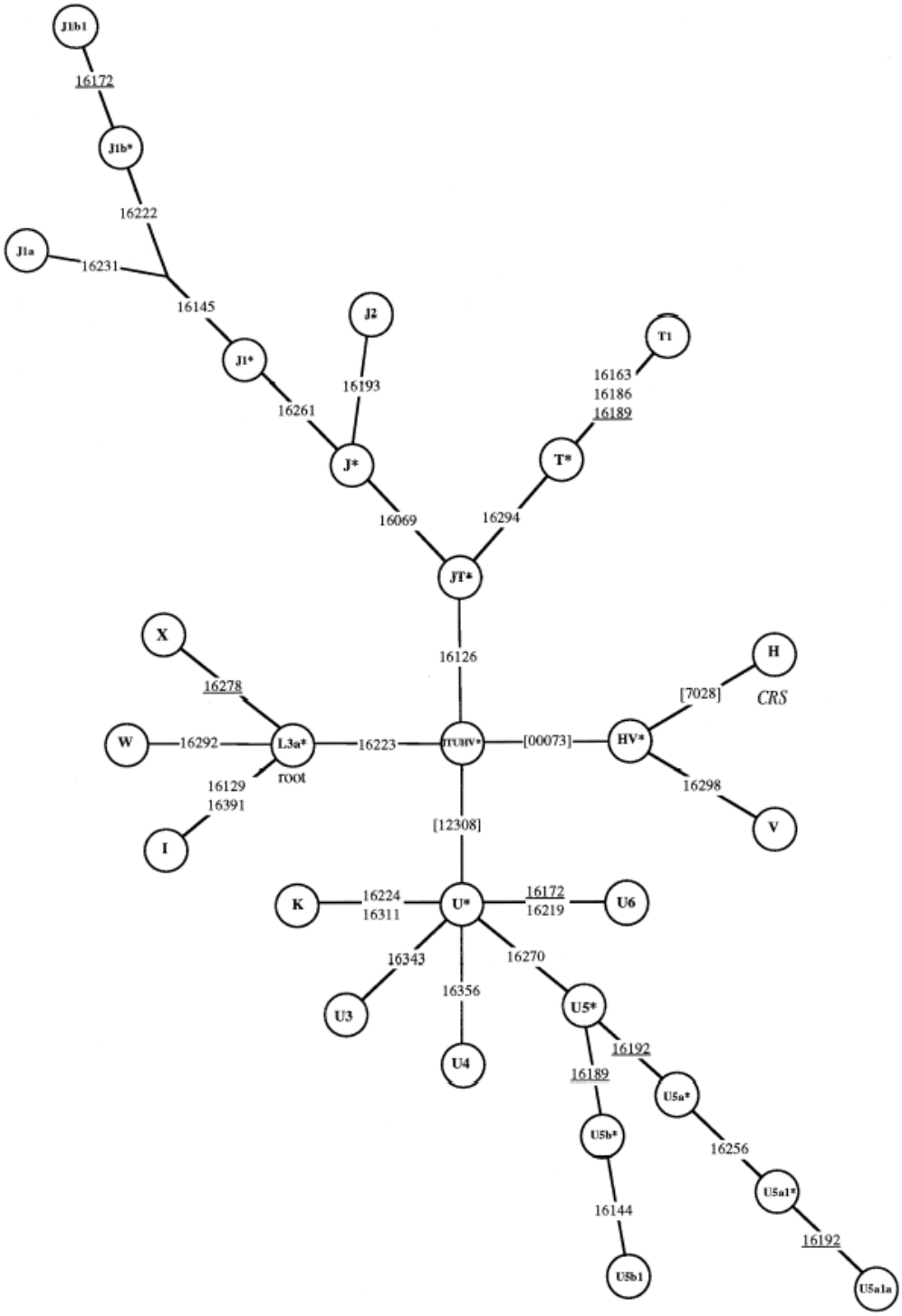


Figure 2. Schematic tree for European mtDNA variation. Clusters of sequences comprising named clades are outlined and labelled. The node marked CRS corresponds to the Cambridge Reference Sequence. The branches are labelled with HVSI polymorphisms, and otherwise by HVSII or coding-region polymorphisms in square brackets. Note that motif positions may occasionally revert within a cluster, particularly in the case of rapidly evolving positions in the control region [figure from (Richards et al. 1998)].

To avoid misunderstandings, I specify here that all mtDNA site numbers referred

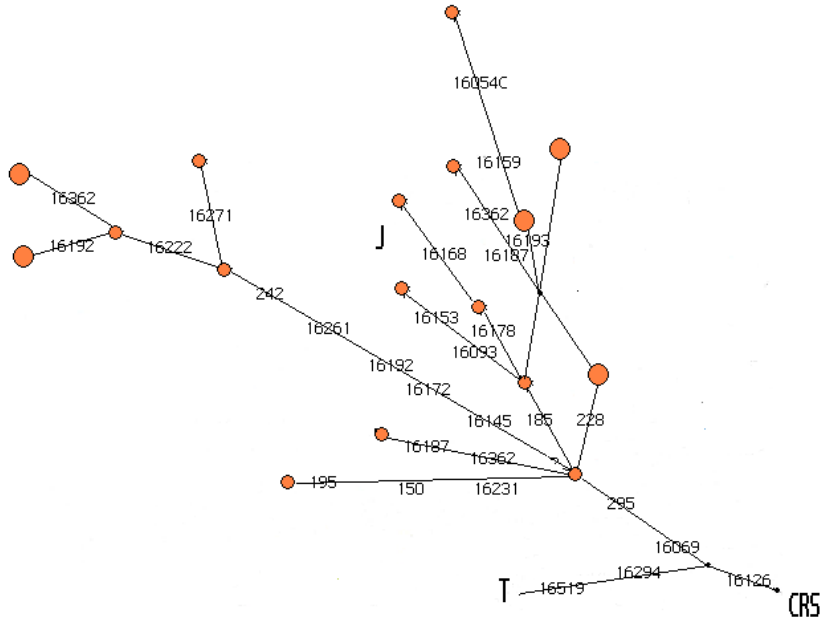
to in this study are in accordance with that for the first established complete mtDNA sequence (Anderson et al. 1981), as revised in (Andrews et al. 1999). HVSI sequence haplotypes (as well as HVSII and RFLP haplotypes) are reported in terms of transitions or transversions from the CRS; for example, 16069-16126 describes an HVSI sequence which differs from the CRS by transitions at the positions 16069 and 16126.

Several ambiguities in the phylogenetic branching of mtDNA were resolved when the HVSI data (Richards et al. 1998) was augmented with data from RFLP analysis of the coding region (Macaulay et al. 1999). It was shown that haplogroups J and T are two ancient subclusters of the same major mtDNA clade, and that the mutation in np 4216 (T>C) has occurred in this context only once - on the shared root of the JT clade. As for the 10394 *Dde*I site (transversion A>G at np 10398), it appeares, that it probably had occurred once deeply in the phylogeny and reverted subsequently. This site is distinguishing several major clusters (for example like J from T) and it is sufficiently stable within each cluster to be a potentially diagnostic in association with other markers (Macaulay et al. 1999).

Information about mtDNA HVSII of J haplogroup emerged in the following years by some studies. For example, 410 Icelandic individuals were analyzed according to HVSI and HVSII sequence and RFLP markers data (Helgason et al. 2000). A median-joining network was generated to infer phylogenetic relationships between Icelandic mtDNA lineages. MtDNA variants that clustered into haplogroup J are shown in Figure 3. It indicates that HVSII can be informative concerning the inner structure of haplogroup J. For example, HVSII polymorphisms at nps 185, 195, 228, 242 and 295 might be used in order to increase the resolution of halogroup J phylogenetic tree. Positions 150 and 152 are represented as having multiple mutation hits.

The conclusion that HVSII might be of importance in studying the topology of J haplogroup can also be drawn analyzing mtDNA polymorphisms of 13 Spaniards belonging to haplogroup J (Crespillo et al. 2000). In addition to the HVSII polymorphisms at nps 185, 195, 228, 242 and 295 specific to haplogroup J, HVSI polymorphism at np 16278 was reported to be abundant among the Spaniards belonging to haplogroup J.

Figure 3. Median-joining network of Icelandic mtDNA lineages for HVSI and HVSII regions. Circles are proportional to haplotype frequencies, in which the smallest circle represents single-copy haplotype. Lines between nodes (filled or empty) are



proportional in lenght to the number of substitutions. Transitions are indicated on the lines by np number; a single transversion present is indicated by np and base abbreviation for a derived state. A reticulation in the network, in which it has been impossible to resolve a recurrent mutation at one or more sites, is represented by a parallelogram-like square where the parallel lines offer alternative evolutionary trajectories in the phylogeny. CRS stands for Cambridge reference sequence [figure from (Helgason et al. 2000)].

A new and an "ultimate" phylogenetic tree-building era came when data about complete mtDNA sequences started to accumulate. For example, phylogenetic networks for mtDNA haplogroup JT were constructed based on 28 sequence data of the complete mtDNA coding region and HVSI motif (Finnilä and Majamaa 2001). According to this study, haplogroup J network found on sequence variation in the coding region, has been divided into subclusters that conformed to the topology proposed before (Torroni et al. 1997). Accordingly, np 7474 *Alu*I and np 15257 *Acc*I site losses were found to define subcluster J2, whereas subcluster J1, which has been defined by the lack of np 15257 *Acc*I site loss (Torroni et al. 1997), was found to be characterized by the G>A transition in the np 3010 that created a restriction site for *Bsh*12306I restriction enzyme at np 3008. A major part of this subcluster was further determined by the transition in the np 14798 (T>C) in the cytochrome b gene. Median-joining networks were constructed separately for complete coding region sequence data (Figure 4) and for the HVSI sequence data (Figure 5). Haplogroup JT network included parallel mutation at np 16172, 16182, 16183, 16186, 16189, 16304 and 16363 and one reticulation. In case of

haplogroup T, the HVSI network correlates well with the network based on the coding region sequence, with the exception of one sample (number 23), which was discrepant between the two networks. In case of haplogroup J, parallel mutations at np 16145, 16172, and 16261 in J1 and J2 had to be assumed in order to obtain concordant networks. There are a few discrepancies between J haplogroup networks based on HVSI segment and coding region variation. However, emphasizing seeming contradictions between the two types of phylogenetic reconstructions appears to be essentially artificial: there is only one true topology and parsimony criteria applied to the combined HVSI and coding region data set, allow to propose at least the most parsimonious solution, hopefully coinciding with the"true" geneology.

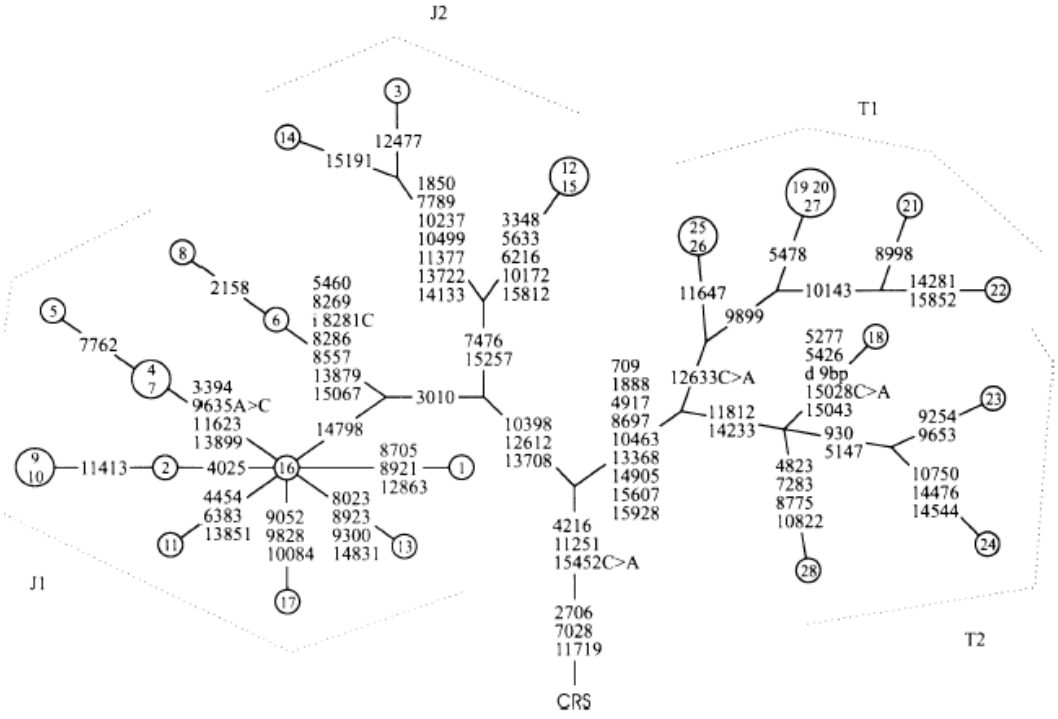


Figure 4. Phylogenetic network of haplogroup JT based on variation in the coding sequence. Numbers inside the nodes are samples. The polymorphic variants are shown on the lines connecting the nodes and are transitions unless otherwise marked. Broken lines around the branches show the locations of subclusters. CRS, Cambridge Reference Sequence, i8281C, insertion of variable number of cytosine; d 9bp, 9-bp deletion in the CO II-tRNA^{Lys} intergenic region [figure from (Finnilä and Majamaa 2001)].

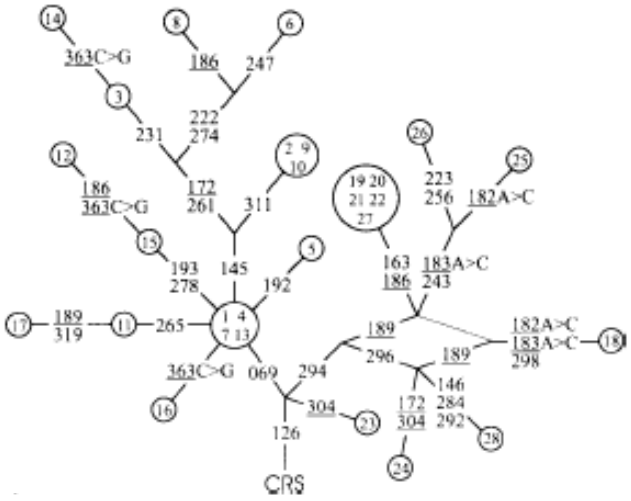


Figure 5. Phylogenetic network of haplogroup JT based on variation in the HVSI sequence. The numbers inside the nodes denote samples and numbers on the lines connecting the nodes denote polymorphic nucleotides, with the first two digits, 16, omitted. Underlined digits indicate parallel mutations. CRS, Cambridge Reference Sequence [figure from (Finnilä and Majamaa 2001)].

In the same year, article by Nicole Maca-Meyer et al. was published (Maca-Meyer et al. 2001). 42 complete human mtDNA lineages were sequenced. Phylogenetic network based on complete mtDNA genome sequences was presented (Figure 6). Due to small sample size, (the tree topology is based on two haplogroup J and two haplogroup T complete sequences), the resolution becomes poor and the tree topology bifurcating, and no confounding conclusions can be drawn. Several parallel mutations were detected, for example, 3010 G>A, has occurred three times in European populations (in the haplogroups J, U and H) and was also found in African haplogroup L2, suggesting either parallel mutations at these sites or very old mtDNA alleles that arose in African mtDNA sequences. Those subsequently evolved from a common ancestor to "newer" European mtDNA sequences in which they underwent "reversion" on multiple occasions. As for haplogroup J and subhaplogroup H1, the 3010 transition has arisen in the root of the haplogroup, in other cases, as for haplogroups U and L2, the 3010 polymorphism has arisen on the tips of the already formed branches of the phylogenetic tree, suggesting parallel mutations at this site.

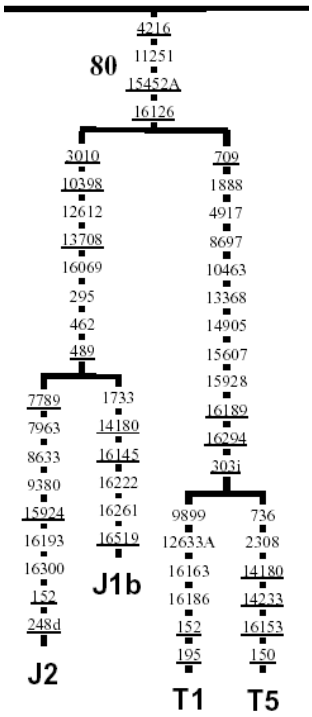


Figure 6. Phylogenetic network for haplogroups J and T based on complete mtDNA genome sequences. Numbers along the links refer to nucleotide positions, suffixes are transversions, underlining indicates recurrent mutations; the order of the mutations on a path not interrupted by any branching or distinguished nodes is arbitrary [figure from [(Maca-Meyer et al. 2001)].

A year later, in 2002, a study based on the nucleotide variation in the coding region of human mtDNA was performed by Corinna Herrnstadt and colleagues (Herrnstadt et al. 2002). 560 complete European, Asian, and African mtDNA coding region sequences from unrelated individuals were analyzed. Phylogenetic networks of African, Asian and European mtDNA sequences based on coding region variations were developed. Here, on Figure 7, haplogroup JT (based on 33 haplogroup J and 46 haplogroup T variants) of the European phylogenetic network from the previous study is presented. The transition in the position 3010 was, in addition to haplogroups H, J, U and L2, also discovered from the Asian haplogroup D, confirming its fast-evolving nature. Several polymorphisms proved to be unique to J haplogroup: variations at nps 5633 (C>T), 7476 (C>T), 10172 (G>A), 10499 (A>G), 12612 (A>G), and others. Several new coding region polymorphisms were found to be unique to haplogroup J:

Hypervariable regions from D-loop were not analysed hence as for J haplogroup, valuable information became lost.

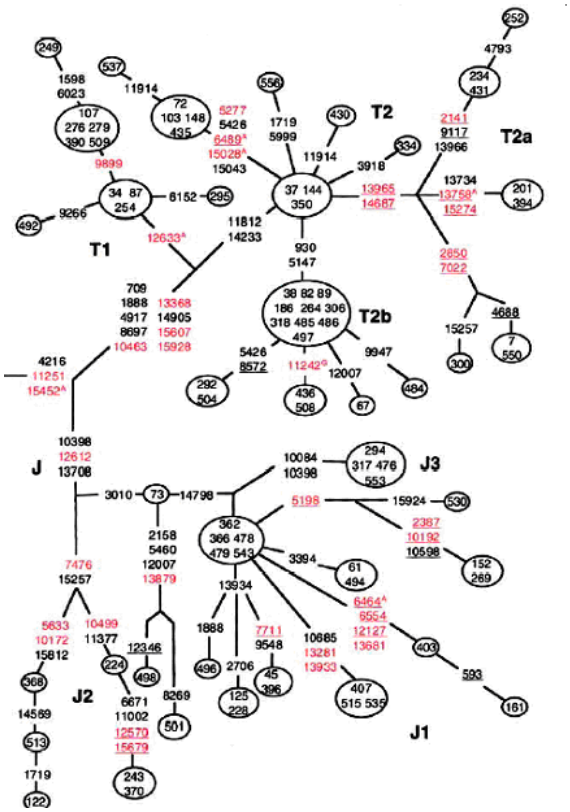


Figure 7. Phylogenetic network of 78 haplogroup JT mtDNA sequences based on coding region variations relative to CRS. Numbers in nodes indicate mtDNA sequences. Nucleotide positions in red are haplogroup specific and appear only on one branch of one haplogroup. Underlined nucleotide positions are novel and are reported here for the first time. Nucleotide substitutions are transitions unless indicated otherwise by suffixes, which denote transversions [figure from (Herrnstadt et al. 2002)].

In the previous survey of haplogroup J topology, according to RFLP, HVSI and HVSII data, major articles that provided useful insights into J haplogroup phylogeny, were cited. There were other, minor studies on haplogroup J that dealt with the issue on the other perspective i.e. associations with LHON: (Ohlenbusch et al. 1998) or associations with aging and longevity: (Rose et al. 2002). These studies provided no new information about the J topology thus they were not included in the current survey of literature. Haplogroup J HVSI polymorphisms described in the literature so far, for example the polymorphisms at np 16145, 16193, 16222 and 16261 may have occurred independently in different clades of haplogoup J and represent fast evolving sites in mtDNA HVSI. In order to clarify haplogroup J tree topology, information from coding and control regions should be combined.

Geographic spread of mtDNA haplogroup J

Average frequency of J haplogroup is the highest in the Near – East (12%) [Richards, 2000 #961; our data) reaching to the highest value, 25%, in the Arabian Peninsula among the Bedouins (Di Rienzo and Wilson 1991). Average haplogroup J frequency starts to decline towards Europe (11%)(Richards et al. 2000), Caucasus 8% (Macaulay et al. 1999; Richards et al. 2000), North – Africa (6%) (Corte-Real et al. 1996; Krings et al. 1999; Pinto et al. 1996; Rando et al. 1998) and becomes practically missing in the East – Asia 1% (Kolman et al. 1996; Yao et al. 2002) and Australia 0% (Huoponen et al. 2001; Redd and Stoneking 1999). In the European populations, J haplogroup is evenly spread (on average 10%), giving occasional peaks i.e. 12% among the Spaniards (Corte-Real et al. 1996; Crespillo et al. 2000; Larruga et al. 2001; Pinto et al. 1996) and 18% among the Italians [Richards, 2000 #961; our data). Frequency pattern can, at least partly, be expounded by the colonization of Europe, North – Africa, and Asia during the Upper Palaeolithic and the Neolithic times. There has been a considerable debate on the respective roles of Palaeolithic (50 000 – 15 000) and Neolithic (7000 – 12 000 YBP) population expansions to the contemporary Europeans' mtDNA gene pool. Studies based on mtDNA haplogroup and haplotype frequencies (Richards et al. 1996; Richards et al. 1998) and the large scale analysis of Y-chromosome markers (Underhill et al. 2001) indicate that European gene pool has mostly Upper Paleolithic origin and the genetic contribution of geographically expanding Near Eastern farming populations during Neolithic was small. Others (Barbujani and Bertorelle 2001; Semino et al. 1996) claim that European gene pool has a largely Near Eastern ancestry dating to the Neolithic period. However, overestimation of the Neolithic share to the contemporary European gene pool by last two authors can be assigned at least partly, to the methodology and sample size of the defining markers used in their works. It is commonly accepted by know, that most of the European haplogroups have likely originated from the Near East, during the Upper Palaeolithic (Richards et al. 1996; Richards et al. 2000). J and T haplogroup (or at least some J and T subhaplogroups) might represent some of the exceptions. Most wide-spread theory states that J haplogroup originates from Near East and was brouht to Europe by farming and herding populations during the Neolithic time. Early agriculturalist expansion across continental space could have consisted, at one extreme, of continuous population

growth along an expansion front, of the type defined by (Ammerman and Cavalli-Sforza 1984) for Neolithic Europe as a wave of advance fueled by “demic diffusion”. At another extreme, one might have a progression of saltatory jumps from one suitable environment to another, as suggested for Neolithic Greece (Van Andel et al., 1995). The first evidence of food production (farming and animal breeding – i.e. the so-called “Neolithic revolution”) dates at around 12 000 YBP in the Levant/Anatolia (designation of the region in the eastern end of the Mediterranean Sea: Israel, Lebanon, Syria and Jordan) (Ammerman and Cavalli-Sforza 1984; Renfrew 1987). The Natufian culture most probably stands for as the first farming – herding society. One has to keep in mind, however, what does mean the word “Neolithic” in the European and in the Near Eastern context. Neolithic is rather a cultural term to which one can not apply a precise timescale. Neolithic time in Near East started about 12 000 YBP, whereas in Europe about 7000 YBP. There might have been several population dispersal waves towards Europe during the Neolithic times and during the post-Neolithic times. Post-Neolithic population expanions have had their effects mainly by “reorganizing” and “redistributing” the genetic material that was present in Europe at the end of Neolithic period because at that point European population was large enough that the small amount of non-European gene flow had an insignificant effect on the European gene pool as a whole (Bellwood, 2001).

The aim of the present study

Haplogroup J is a variety of human mtDNA pool of a special interest for several important reasons, general for the demographic history of western Eurasian populations, as well as to explore deeper some aspects that may have direct relation to molecular medicine. Firstly, haplogroup J is one of the haplogroups that was presumably brought to Europe from the Near East during the Neolithic by the first farming-herding societies – its phylogeography, as far as it was known and interpreted so far, has been a corner-stone for genetic arguments for the spread from Anatolia and/or Levant demic diffusion – itself a consequence of the population growth thanks to the transition to food production in Fertile Crescent and close-by regions (Bellwood, 2001; Salamini et al., 2002). Secondly, it is known to be associated with mitochondrial disease LHON and, probably, with longevity at least in some populations, suggesting a multigene trait - not really surprising for such a complex phenomenon as lifetime expectancy (Brown et al. 2002; Rose et al., 2002). Elucidation of J haplogroup's genealogy would hopefully benefit exploring both of these hypotheses.

MtDNA phylogenies have long been based on fragments of the mtDNA genome: either D-loop sequences or RFLP data. The drawbacks of relying solely on the D-loop polymorphisms is largely assigned to mutational hotspots. RFLP analysis of the coding region, on the other hand, is limited by the recognition of available enzymes. According to the studies performed so far, main clades of J haplogroup's evolutionary tree can be defined by RFLP information but further elucidation is conducted through coding region or HVSI sequencing. However, it is clear that HVSI variation due to fast-evolving nucleotide sites may lead to erroneous results in tree topology and full-genome sequencing is yet too labour-intensive task to complete in a massive scale, involving many hundreds or thousands of samples. Thus, the aim of this work was to make use of:

a) the existing HVSI and "classical" RFLP data;
b) rapidly accumulating new knowledge offered by complete mtDNA sequences;
c) new insights from the information about the variation of HVSII,

in order to work out a detailed strategy to study variation in human mtDNA haplogroup J and to taste as well as to imply this strategy to cover Caucasoid populations – i.e. most of extant populations living in Europe, West Asia and North Africa.

Subjects and methods

Samples

712 samples belonging to haplogroup J according to their RFLP and HVSI data were collected from our mtDNA bank. MtDNA genomes from the following populations were represented. Samples were divided into groups, based on geographic of ethnic affinities.

European samples (417 samples).

Scandinavia: 37 Swedes (from different regions in Sweden, including Gotland). Eastern Europe: 38 Estonians, 9 Latvians and 5 Lithuanians, 28 Ukrainians, 12 Russians from Krasnodar, 6 Hungarians, and 31 Russians. Caucasia: 8 Armenians, 23 Ossetians, 2 Adyges, 6 Russians from Adygea, 5 individuals from Karachaev, 6 Nogays, 2 Kumyks, 6 Lezgines, 2 Abazines, and 4 Kabardins. Volga-Uralic region: 3 Permiac Komis, 6 Zyrian Komis, 2 Chuvashes, 24 Khants, 2 individuals of Mordva-Eryza and 4 individuals of Mordva-Moksha ancestry. Central Europe: 18 Slovaks, and 7 Czechs. Southern Europe: 16 Albanians, 27 Turks, 15 Bosnians, 13 Italians, 10 Greeks, 9 Moldovans, 10 Moldaevian Russians, 9 Gagauzes, 11 Cypriots, and 1 Cretan.

Near – Eastern samples (207 samples).

60 Iranians, 36 Kuwaitis, 44 Saudi-Arabians, 8 Lebanese, 24 Syrians, 12 Jordanians, 7 Yemenitess, 11 Omanis, and 5 Jews (Ashkenazim).

Central and North – Asian samples (45 samples).

9 Kazakh-Shoris, 9 persons from Tuva, 1 Yakutian, 1 Ket, 2 Uighurs, 3 Kalmyks, 5 Arabs from Uzbeckistan, 2 Uzbecks, and 13 Kazakhs.

North – African samples (41 samples).

28 Moroccans and 13 Egyptians.

RFLP analyses were performed to the whole haplogroup J sample (712 individuals). HVSII region was sequenced for 306 samples chosen on random from the European samples data set. Although the whole European J haplogroup was planned to be screened for HVSII variation, due to technical problems some of the European nations (i.e. from Volga - Uralic region and Caucasia) that belong to J haplogroup and are

represented in our mtDNA bank, remained unsequenced. In the current study HVSII variation was screened for Armenian, Ossetian, Sweden, Slovakian, Czech, Bosnian, Ukrainian, Latvian, Lithuanian, Estonian, Hungarian, Turkish, Albanian, Greek, Italian, Cypriot, Kumyk, Jewish, and Moroccan populations (altogether 306 samples). Obtained results are shown in Table 2 (see appendixes). According to the results obtained from HVSII variation and combinig that with the coding region and HVSI polymorphisms, three different median networks were constructed. In Figure 9, European J haplotypes are presented in a median joining network. In addition to the European samples that were screened for HVSII and RFLP polymorphisms in the current study, 17 Finns whose CR and coding region polymorphisms were described previously in literature (Finnilä and Majamaa 2001) were included in data analyses and allocated into European J haplogroup median joining network. In Figure 9, European population was divided between four regions in Europe: Northern Europe (Estonia, Latvia, Lithuania, Sweden and Finland), Eastern and Central Europe (Ukraine, Slovakia, Czech, and Hungary), Balkans (Greece, Crete, Albania, Bosnia, and Italy) and Caucasus (Ossetia and Armenia) rather according to linguistic and genetic than geographic affiliation. In Figure 10, Near Eastern J haplotypes are shown in a median joining network. In Figure 11, a median joining network is presented, showing the ambiguties arising in J haplogroup's topology when HVSII polymorphisms (described according to literature overview as of "slow" substitution rate) are considered as important as HVSI and coding region polynorphisms. Coding region polymorphisms (7789, 10499 and 14798) that are undetectable with restriction enzymes were sequenced for the 306 European samples. J subhaplogroup frequencies were calculated according to the whole J haplogroup sample set (712 individuals) according to their HVSI and RFLP variations. Haplotype diversities and number of halpotypes were also calculated for the whole J haplogroup mtDNA sample (712 individuals) according to their HVSI and RFLP variation.

DNA amplification and sequencing

DNA amplification was performed on 10 – 20 ng of template DNA and was carried out with the thermocycler "Biometra UNO II" usually in total volume of 15-20µl.

Component	Concentration	Concentration in PCR reaction
Tartrazine Buffer or "8,3" buffer or "B" buffer	750 mM Tris-HCl, pH 8.8, 200 mM $(NH_4)_2SO_4$, 0.1% Tween 20; 5% Ficoll 400. 400mM Tri-HCl, pH 8,3, 110µg/ml BSA Courtesy of Solis Biodyne	1/10
$MgCl_2$ dNTP mix (dATP, dCTP, dGTP, dTTP) Taq DNA polymerase (FIREPol) from Solis Biodyne and/or from Juhan Sedman	25 mM $MgCl_2$ 10 mM 2 U/µl	2,5 mM 1 mM 0.125 –0.2 U
F primer R primer Deionized water DNA sample	10 pmol/µl 10 pmol/µl different	~0,2 pM ~0,2 pM 1-3 µl

Oligonucleotide sequences used for PCR and sequencing reactions are listed in Table 1. The cycle profile started with 94°C for 1 min, followed by 35 – 45 cycles of 94°C for 20 s, 52 - 59°C for 15 s and 72°C for 1 min. Number of cycles and annealing temperature depended on primer specifity and mtDNA quality. Incubation with restrictases (0.3-0.5 units per reaction) was done overnight. Sequencing reactions were made using the DYEnamic ET Terminator Cycle Sequencing Kit from Amersham Pharmacia Biotech on a MegaBase 1000 DNA sequencer. PCR primers were destroyed using exonuclease I and free deoxynucleotides were eliminated by shrimp alkaline phosphatase (both from Amersham Pharmacia Biotech). Following reactions were

carried out in 10-µl volume, with using 5µl of purified PCR product and 2µl DYE premix, 1µl of one. As for HVSII region, sequences were obtained between sites 70 and 500. The cycle sequencing profile was 33 to 35 cycles of 94°C for 20s, 50°C for 15 s and 60°C for 1 min. The sequences were aligned and manually checked in SeqLab (GCG Wisconsin Package 10, Genetics Computer Group).

Table 1. Detection of polymorphisms. Amplified sequence or endonuclease restriction site is shown in the first column. The first number refers to the polymorphic position, the second number indicates the restriction site it affects. F and R indicate forward or reverse primers. Number in front of the sequence indicates the position of the 5' nucleotide of the primer.

		Sequences
HVSII, sequencing	F	5' 16029 GCTCTATCACCCTATTAACCAC
	R	
HVSII, sequencing	F	5' 16048 CTCACGGGAGCTCTC
	R	
HVSII, sequencing	F	5' 16453 CCGGGCCCATAACACTTGGG
	R	
HVSII, sequencing	F	5'16483 GTGAACTGTATCCGACATCTGG
	R	
HVSII, sequencing	F	
	R	5' 408 TTGAGGAGGTAAGCTACATA
HVSII, sequencing	F	
	R	5' 580 TTGAGGAGGTAAGCTACATA
HVSII, sequencing	F	5' 16400 CACCATTCTCCGTGAAATCA
	R	5' 755 AGGCTAAGCGTTTTGAGCTG
Sequencing np 7789	F	5' 7128 ACGCCAAAATCCATTTCACT
	R	5' 8075 CGGGAATTGCATCTGTTTTT
Sequencing np 10499	F	5' 10284 CCATGAGCCCTACAAACAACT
	R	5' 10628 GGAGTGGGTGTTGAGGGTTATGAGAGTA
Sequencing np 14798	F	5' 13980 GCATAATTAAACTTTACTTC
	R	5' 14978 AGAATATTGAGGCGCCATTG
3010; 3008 *Bsh*1236I	F	5' 2981 ACGACCTCGATGTTGGATCAGGACATCGC
	R	5' 3168 GAAGGCGCTTTGTGAAGTAGG
4216; 4216 *Nla* III	F	5' 4057 TCCGAACTCTACTCG
	R	5' 4251 GGGAATGCTGGAGAT
5633; 5633 *Alu*I	F	5' 5424 CATACAAAACCCACCCCATTCCTC
	R	5' 5660 CTAGTAAGGGCTTGGCTTAA
7476; 7474 *Alu*I	F	5' 7169 TCTAACTTTCTTCCC
	R	5' 7817 GGGCGATCAGGACTA
10398; 10394 *Dde*I	F	5' 10284 CCATGAGCCCTACAAACAACT
	R	5' 10484 GTAAATGAGGGGCATTTGGTA
12612; 12603 *Msi*I	F	5' 12541 GCCACAACCCAAACA
	R	5' 12812 CGGGCGTATCATCAACTG
13708; 13704 *Bst*OI	F	5' 13583 CCTCCCTGACAAGCGCCTATAGC
	R	5' 13843 CTAGGGCTGTTAGAAGTCCT
15257; 15257 *Xmi*I	F	5' 15149 TGAGGCCAATATCATTCTGAGGGG
	R	5' 15701 GGCGAAATATTATGCTTTGT

Detection of polymorphisms in haplogroup J samples and data analysis

All the samples belonging to J haplogroup were collected from our mtDNA bank according to their HVSI and RFLP motifs. Information about polymorphic sites in mtDNA coding region was obtained from published complete sequences and RFLP studies (Finnilä and Majamaa 2001; Helgason et al. 2000; Herrnstadt et al. 2002; Maca-Meyer et al. 2001; Macaulay et al. 1999; Richards et al. 1996; Richards et al. 1998; Torroni et al. 1994; Torroni et al. 1997). The whole haplogroup J sample (712 individuals) was screened for the polymorphism at np 13708 and 10398 that are two of the many polymorphisms specific to J haplogroup (Finnilä and Majamaa 2001), in order to exclude the possible earlier misidentifications or mixing sample numbers. All the samples were additionally tested for the polymorphisms at nps 3010 and 7476, in order to divide the samples between two major clades of J haplogroup (Torroni et al. 1997; Finnilä et al. 2001). The samples whose 3010 and 7476 RFLP data or/and 13708 and 10398 RFLP data was controversial were additionally screened for the polymorphisms at np 4216 and 12612, the former defines the common root of JT haplogroup (Macaulay et al. 1999; Richards et al. 1998) and the latter is known to be specific to haplogroup J (i.e. (Finnilä and Majamaa 2001)). The samples harbouring polymorphism at np 16193 according to their HVSI data and loss of *Alu* I restriction site at np 7474, were tested for the polymorphism at np 5633 (Finnilä and Majamaa 2001). The samples harbouring *Alu* I restriction site loss at np 7474, were additionally screened for *Xmi* I (an isoscisomer of *Acc* I) restriction site loss at np 15257. The samples harbouring polymorphism at np 16231 in their HVSI region and *Alu* I restriction site loss at np 7474 (thus belonging to J2 subhaplogroup), were screened by sequencing for polymorphisms at np 7789 and 10499, in order to increase the phylogenetic resolution of the J2 branch.

The second hypervariable region was amplified and sequenced for 306 samples from European populations. Problems arose with the HVSII sequncing primers. Namely, DNA from different populations was of different quality, so in order to obtain HVSII sequences 8 different sequencing primers were tested and combined with each other when sequencing HVSII of every population. HVSII of 82 Near Eastern samples (11 Egyptians, 19 Iranians, 4 Jordanians, 15 Kuwaitis, 2 Lebanese, 18 Saudi – Arabians and 13 Syyrians) was sequenced by Karmen Mae. According to her HVSII sequencing information median joining network for Near Eastern J haplogroup samples was drawn

(see Figure 10). In order to obtain true pictures of J haplogroup spread in Europe and in Near East, I added the Turks, the Cypriots and the Jews (Ashkenazim) to the Near Eastern median joining network (again see Figure 10). I did it for classical reasons, since Anatolia as well as Cyprus may be regarded as Near East. HVSII of 21 Maroccans was screened for polymorphisms by Erwan Pennarun and the Moroccans were allocated into the European median joining network.

Networks were constructed manually according to reduced median and median joining principles (Bandelt et al. 1999; Bandelt et al. 1995). Nucleotide positions were divided into three classes of transition rates, fast (146, 150, 152, 189, 16092, 16093, 16129, 16189, 16304, 16311 and 16362), intermediate (72, 93, 95, 143, 182, 185, 194, 195, 198, 200, 204, 16171, 16172, 16209, 16223, 16278, 16291 and 16293) and slow (remainder of the positions between 16090 and 16365 and the remainder of the positions between 73 and 500) (Hasegawa et al. 1993; Ingman et al. 2000; Stoneking 2000; Wakeley 1993). Classes weights 1, 2 and 4, were assigned respectively. Transversions were weighted 8 except for those adjacent to the HVSII polycytosine tract from nps 302 to 309 and HVSI polycytosine tract from nps 16184 to 16193, which were ignored as resulting length variation (Bendall et al. 1996).

The time to the most recent common ancestor of each cluster was estimated using ρ, the average transitional distance from the putative root haplotype. For calibration 1 transitional step between nps 16090 and 16365 was taken equal to 20180 years (Forster et al. 1996). Standard deviation σ was calculated as in (Saillard et al. 2000).

Haplotype diversity was calculated using the formula $D = n/(n-1)*(1-\Sigma x_i^2)$, where n is the total number of samples and x_i is the frequency of i-th HVSI haplotype (Nei 1987).

Results and discussion

In the current study HVSII variation was screened for 306 mtDNA samples of mostly European origin. Polymorphisms at np 7789, 10499 and 14798 were screened by mtDNA sequencing of 306 mtDNA samples. RFLP analysis was performed to coding region polymorphisms at np 3010, 7476, 5633, 10398, 4216, 12612 and 15257 where necessary to mtDNA samples belonging to J haplogroup (712 individuals). RFLP analysis provided us information about the subdivision of J haplogroup samples between two major branches J1 and J2, and in addition to that, J2b. Skeleton network for haplogroup J was constructed according to RFLP data. Further analysis of haplogroup J topology was conducted using the information obtained from the control region variation – by sequencing HVSI and HVSII. 306 HVSI and HVSII haplotypes as well as RFLP data is shown in Table 3 (see appendixes). Nucleotide positions that according to literature were stated as fast evolving (i.e. HVSII positions 146, 150 and 152) were excluded from data analysis. Transition (G>A) at np 263 represents a set of rare polymorphisms in the CRS mtDNA and was excluded from the study. Transition (A>G) at np 73 is not haplogroup J specific but has shared anchestry in haplogroups U, K, T, J, I, W, X and was not included in the present study. Three individuals - Turk 271, Turk 289 and Greek 88 - were excluded from the analysis. The Turks showed controversial HVSI, HVSII and coding region polymorphisms while the Greek was probably mistaken as haplogroup J sample according to its HVSI polymorphism and it apparently belonged to haplogroup H because it carried HVSI polymorphism 16069 that is also common to H haplogroup. Three J haplogroup median joining networks are presented in the current study. European J haplogroup is presented in Figure 9, Near Eastern J haplogroup is presented in Figure 10. In the third graph (Figure 11), J haplogroup HVSII polymorphisms are combined with HVSI and coding region polymorphisms.

Topology of haplogroup J

Root of the median network constructed of J haplotypes is defined by several coding region and HVSI polymorphisms at nps 4216, 10398, 12612, 13708, and 16069. Haplogroup J is further divided into two major branches according to *Bsh*

12306I restriction site at np 3008 and *Alu* I restriction site at np 7474 losses. Loss of *Bsh* 12306I restriction site at np 3008 defines J1 subhaplogroup whereas loss of *Alu* I restriction site at np 7474 defines J2 subhaplogroup. In addition to the loss of *Bsh* 12306I restriction site at np 3008 is subhaplogroup J1 characterized by HVSII polymorphism at np 462. Subhaplogroup J2 is additionally defined by loss of *Acc* I restriction site at np 15257. According to the current study out of 712 haplogroup J samples 86% (613 individuals) belonged into J1 subhaplogroup, 14% (99 individuals) belonged to subhaplogroup J2. Subhaplogroup J1 is divided into 3 subclusters J1a, J1b and J1c. Polymorphism at np 16222 is one of the polymorphisms defining J1a subcluster. Polymorphism at np 14798 is one of the polymorphisms defining J1b subcluster, polymorphism at np 16193 is one of the polymorphisms defining subcluster J1c. J haplogroup samples can be divided between these three J1 subclusters only when their HVSI and coding region polymorphisms are combined and compared. As subhaplogroup J1 is comparatively well represented in all the populations included in this study, then subhaplogroup J2 is on the contrary, less abundant.

Especially confusing in constructing median network for haplogroup J were HVSI polymorphisms at nps 16145 and 16261. An interesting phenomenon was observed. Namely, HVSI polymorphisms 16145 and 16261 seem to be coupled in J haplogroup. Both of them occur as a couple on the branches leading to J1a and J2a subclusters. That was clear already from earlier studies, in particular from complete mtDNA sequence data (Finnilä and Majamaa 2001). In addition, as Figure 9 suggests that the combination 16145 and 16261 has also formed inside J1b subcluster. In case we regard HVSII polymorphisms the defining ones (as it is shown in Figure 11) it is also possible to allocate the subset of J1b linegaes carrying HVSI 16261 into a separate, the fourth subcluster according their HVSII polymorphism at np 188. It is absolutely clear that assuming neutral evolution, probablility of arising this tandem of transitions independently, four times in one single haplogroup, is extremely low ($p<0.01$). In order to study the evolutionary background of these two polymorphisms, the following haplogroup J HVSI lineages 16069-16126-16145 and 16069-16126-16261 and 16069-16126-16145-16261 were selected from the data set and analysed. The task was easier for J1 subhaplogroup since the sample size for this subcluster is much larger. The results for J2a subcluster are statistically less representative because of smaller sample

size. Firstly, hypothesis was considered that 16145-16261 motifs have not formed as parallel mutations on both J1 and J2 branches. As it was mentioned in the literature overview, transition (G>A) at np 3010 has occurred several times in mitochondrial phylogeny and can be excluded from the analyses for being recurrent. However, one may notice that inside haplogroup J, the transition (G>A) at np 3010 has not reverted nor shown any parallel mutation. Furthermore, an even stronger argument in favour of independent origins of 16145, 16261 tandem motifs in J1 and J2 is provided by the fact that the two sub-clades under consideration, differ, in addition, by three mutations that are stable according to mtDNA sequence databases – namely at nps 462 in HVSII and 7476 plus 15257 in the coding region of mtDNA genome (see Figure 9). Even more convincingly, the 16145, 16261 motif in J2a clade is always seen at the background of coding region mutations at nps. 7789 and 10499, not reported to occur within J1 clade. Taken together, irrespective of a low probability of an independent origin of this tandem HVSI signature in haplogroup J, its occurence at least twice in the phylogenesis of this haplogroup appears to be a more parsimonious solution to explain the topology of haplogroup J, than to assume that four (462, 3010, 7476, 15257) or even six (the former plus 7789 and 10499) mutations, nearly all within coding region, have arisen independently twice. When polymorphisms 16145 and 16261 were mutational hotspots it would be easier to explain the emerging picture. But they are not. Hence, the question remains: why are they mutating in haplogroup J in tandem. A possible explanation is that these polymorphisms somehow contribute to fitness and, therefore, are under positive selection, being, i.e. somehow functionally connected to so far unknown conserved regulatory element in HVSI in the specific context of haplogroup J. One may speculate further and to suggest that this context should be somehow embedded into an array of mutations (a sub-fraction of them) that defines haplogroup J, namely nps 489, 16069, 4216, 10398, 12612 and 13708 (see Figure 9).

Haplogroup J may play a role in successful ageing and it provides a background for some mitochondrial diseases (see Literature overwiew, page 12), thus one may to assume connections between haplogroup J polymorphisms and mitochondrial functionality. But there are no known regulatory elements in the vicinity of these two nucleotide sites in the HVSI of mtDNA molecule reported so far, not to add that their speculative positive selection is exclusively context-specific – i.e. manifests itself only

in one particular haplogroup of the human mtDNA pool. Neither has this assumed selection led to the fixation of the 16145, 16261 motif in haplogroup J, suggesting that even if true, this selective advantage cannot be high. Alternatively, other coding or/and control region mutations in the remaining branches of haplogroup J that do not encompass this tandem HVSI motif, may have pleiotropic effects that diminish a putative selective effect of the motif under discussion. Summing up, our current knowledge is insufficient to provide an answer to the question whether an apparently independent occurrence of 16145, 16261 tandem motif in two limbs of haplogroup J topology – and only in haplogroup J – is a stochastic event of a very low probability, or has been selected for. Unfortunately, any straightforward experimental approach to cast light to the dilemma does not seem to be achievable. So the puzzle remains unsolved.

Information about the polymorphisms in HVSII

According to the present study, HVSII variation provides with little or even no help in elucidating J haplogroup's topology. All of the HVSII polymorphisms are presented in Figure 11, except for those (nps 146, 150, 152 and 189) that were previously according to literature overview classified as evolving at "fast rates". In case we consider the fact, that some polymorphisms in HVSI are according to previous studies rather unique, and that HVSII polymorphisms on the whole are "mutational hotspots" we may say that HVSII variation instead of providing with increase in phylogenetic resolution creates "noise" that rather hinders than helps to resolve J haplogroup's genealogy. J haplogroup's topology becomes confused for several subsequent reversions and parallel mutations in all the branches. Still, I noticed a tendency that certain HVSII polymorphisms characterize certain J subclades, at least on the whole. For example, polymorphisms at nps 185, 188 and 228 were only found among the mtDNA J haplogroup samples that had concentrated into J1b subclade. In J1b subclade, however these polymorphisms had experienced several reversions. Or HVSII polymorphisms at np 242 that is common to J1a, but once again, experiences reversions within J1a. The same goes for HVSII polymorphism at np 215 in J2a clade. Polymorphism at np 195 had experienced several reversions as well as parallel mutations in all haplogroup J subclades and can be considered as one of the most inappropriate polymorphisms to reconstruct J haplogroup's phylogeny. At the first

glance, some polymorphisms in HVSII seemed to defining, i.e. 3 Turkes formed a lineage acoording to HVSII polymorphism at np 217 (compare Figure 10 and Figure 11) and the Ossetians according to HVSII polymorphism at np 199, still these polymorphisms do not improve the phylogenetic resolution and can be considered either random or population specific. Nevertheless, three polymorphisms in HVSII at nps 295 defining the root of JT subchaplogroup, at np 319 defining J2a clade and at np 462 defining J1 subhaplogroup were found to be sufficiently stable within J haplogroup, showing no reversions or parallel mutations and can be used as polymorphic markers in evolutionary studies.

Subhaplogroup J1

J1 subhaplogroup can be further divided into 3 subclusters J1a, J1b and J1c according to their most common haplotypes (see Figure 9 and Figure 10). Here we try to combine data from HVSI, HVSII and coding region variation in order to divide haplogroup J samples between J1 subclusters.

Subclade J1a is evenly spread in Europe as well as in Near East. Most of the Caucasian nations analysed in the current study – the Armenians, the Ossetians, and the Kumyks have converged into J1a subcluster and form a "Caucasian – specific" subclade J1a1. "Caucasian" J1a1 subclade shares certain lineages with the Near Eastern J1a1 i.e. 16362, 16274, 16218 and some lineages with Northern Europeans i.e. 16274, still clearly being distinct from Near Eastern or European J1a1 lineages. Most common HVSI lineages for J1a subcluster are 16069-16126-16145-16222-16261, and 16069-16126-16145-16172-16222-16261. HVSI polymorphism at np 16222 is present in nearly all of the J1a lineages and it has been considered as one of the defining mutation for J1a subcluster so far. According to this study, transition at np 16222 may have underwent several reversions inside J1a subcluster as well as parallel mutations within haplogroup J. Thus, in order to reduce ambiguities in haplogroup J evolutionary tree branching order, HVSI polymorphism at np 16222 was considered undefining in J1a subcluster (see Figures 9 and 10). Nevertheless it is presented in Figure 11, in order to demonstrate less parsimonious solution for building an evolutonary tree. Another possible characteristic for defining subcluster J1a arose from determination of HVSII nucleotide variation. Polymorphism at np 242 combined with the HVSI polymorphism

at np 16172 seemed to be specific to several populations belonging to subcluster J1a. Unfortunately, certain lineages like the Ossetians and the Armenians lack both the HVSI polymorphism at np 16172 as well as HVSII polymorphism at np 242. Besides, HVSII polymorphism at np 242 has underwent reversion in J1a subcluster and so might have done the HVSI polymorphism at np 16222. Therefore markers additional to 16222 and 242 that would define also the Caucasian share of J1a subcluster would be benefitial to have at hand. All in all additional polymorphic markers besides polymorphisms at np 242, 16172 and 16222 are required in order to specify the topology of J1a subcluster. According to literature, several coding region polymorphisms have been reported to define J1a subcluster (Finnilä and Majamaa 2001). Still it is not known yet wether these coding region polymorphisms coincide with the formation of transition at np 16222.

Subclade J1b is defined by the coding region polymorphism at np 14798. It could also be defined by HVSII region polymorphisms at np 185, 188 and 228, unfortunately all of these polymorphisms have underwent several reversions as well as parallel mutations in subclade J1b. There is considerable number of European 16069-16126 HVSI lineages that hence can not be further divided into separate subclusters and remain into the central node of J1b. Using additional polymorphic markers from the mtDNA coding region would increase the phylogenetic resolution of J1b subclade, this remains as future objective. Coding region polymorphism (T>C) at np 14798 was found to be a stable marker, experiencing no reversions nor parallel mutations in subclades of J haplogroup and it can be considered as one of the trustworthy defining markers of J1b subclade. Both European and Near Eastern J1b subclades show star-like topology, the European subclade being more divaricated than the Near Eastern one.

Subcluster J1c is characterized by HVSI polymorphism at np 16193. HVSI polymorphism at np 16193 is common all over mtDNA phylogeny and has experienced parallel mutations in all haplogroup J subclusters (except for J2a) and is also one of the defining mutations leading to J2b branch. So, it remains a poor characteristic to define distinct J haplogroup's subclusters. In order to separate J1c and J2b subclusters (as well as the other J subclusters harboring 16193 polymorphism) additional RFLP or HVSII markers were required. So, 16069-16126-16193 lineages that had been allocated to J1 branch according to the *Bsh* 1236I restriction site loss at np 3008, were screened for

coding region polymorphism at np 14798 as well as HVSII polymorphisms 185 and 228. Result of this screening showed that although the polymorphism at np 16193 has experienced parallel mutations in haplogroup J, it has to be considered defining and monophyletic in case of subcluster J1c since we lack any other polymorphic mtDNA markers to elucidate the topology of this subclade. Subcluster J1c is more diversified in Near East than in Europe but still shows no star-like topology. As for the European J1c subcluster, it consists mostly of Caucasian nations' lineages. Once again, in addition to J1a subcluster, the Caucasian nations studied in the present paper, form a distcinct cluster different from the Europeans as well as the Near Easteners. In J1c the Caucasians harbor private HVSI lineage 16193-16320, that has not been detected yet in the Near Eastern populations as well as in the European populations. Polymorphic markers besides 16193 would be useful to define J1c subcluster, albeit according to literature overview, there is only one person so far, belonging to J1c subcluster whose full mtDNA molecule has been sequenced (Maca-Meyer et al. 2001). As a future perspective, full mtDNA sequences of individuals belonging to J1c subcluster would support the further elucidation of J1c topology.

Subhaplogroup J2

Subhaplogroup J2 is defined by *Alu* I restriction site at np 7474 and *Acc* I restriction site at np 15257 losses. Subhaplogroup J2 consists of two subclusters J2a and J2b. Subclade J2a is chracterized by several coding region polymorphisms (7789 and 10499), HVSI polymorphisms (16231) and HVSII polymorphisms (215 and 319). Subcluster J2b is characterized by formation of *Alu* I restriction site at np 5633, HVSI polymorphisms at np 16193 and 16278. There are no HVSII polymorphisms specific to J2b subcluster. J2b subcluster is further divided into J2b1, J2b2 and J2b3 subclusters according to their most common haplotypes.

Subclade J2a provided with a challenge to resolve phylogenetic connections between the HVSI polymorphism at np 16231 and HVSII polymorphism at np 319 and coding region polymorphisms at nps 7789 and 10499. Unfortunately this could not be done. These polymorphisms remained attached with each other. There was one Estonian who harbored loss of the polymorphism at np 16231, indicating revesion at this nucleotide position since all other polymorphisms characteristic to J2a subclade were

present. Subclade J2a seems to be the clade that can very well be described by HVSI polymorphisms like 16231 and coding region polymorphisms like 7789 and 10499 and HVSII polymorphism 319. Among the J subclusters all of these polymorphisms are specific to J2a subclade only. J2a subclade is homogenously spread in Europe, the Caucasian nations lack J2a subclade. Regrettably there is no information available so far about Near Eastern J2a subclade haplotypes.

Subcluster J2b1 is characterized by the formation of *Alu* I restriction site at np 5633. Subcluster J2b1 haplotypes are virtually absent in Europe (only 8 Moroccans are allocated into this cluster). In Near East however, subcluster J2b1 is more abundant and diverse. Most of the lineages allocated into J2b subcluster comprise in addition to coding region polymorphism at np 5633 HVSI polymorphism at np 16193 and form J2b2 subclade. J2b2 subclades are present both in Europe and in Near East. According to HVSI polymorphism at np 16278 the J2b2 subclade was further classified into subclade J2b3. Both of these HVSI polymorphisms (16193 and 16278 that define J2b2 and J2b3 subclades respectively) show no reversions within J2b subcluster. Nevertheless, polymorphism at np 16278 has been characterized as being abundant among the Spaniards belonging to J haplogroup (Crespillo et al. 2000) and according to the current study, has experienced one parallel mutation in European J1b subcluster and one parallel mutation in Near Eastern J1a1 subcluster. As for the polymorphism at np 16193 the situation is alike. It has not experienced reversions inside J2b subcluster, but is rich in parallel mutations in other haplogroup J subclusters. Another problem arises with relative scarcity of the samples belonging to J2 subhaplogroup. According to some studies, J haplogroup LHON mutations are preferably associated with J2 subclade (Finnilä and Majamaa 2001). Still, attributing the scarcity of individuals belonging J2 subclade to LHON expression would be a premature hypothesis since some analyses, quite on the contrary, indicate that polymorphisms at nps 4216 and 13708 may answere for LHON expression (Brown et al. 2002). In conclusion, we can say that most of the problems in studying phylogeny of subhaplogroup J2 arise from the scarcity of the representatives of this branch of J haplogroup.

In order to specify J haplogroup's topology further on, additional polymorphic markers within human mtDNA genome are required. New information about the coding region polymorphisms would be most helpful.

Geographic spread of J haplotypes

132 different haplotypes were detected among 417 European samples according to their HVSI and RFLP polymorphisms. All of the haplotypes were homogenously distributed across Europe, showing no clustering to specific geographical regions. Only the Ossetians and the Albanians seem to have experienced population bottleneck or founder effect during their demographic history when haplogroup J is concerned. The Ossetians and the Albanians exhibited exceedingly similar haplotypes. 109 different haplotypes were detected among 207 Near Eastern samples, this is about twice as much as for this sample size in Europe. Haplogroup J diversity in the Near East is higher than that in Europe (0.987 and 0.899 respectively) probably reflecting the Near Eastern origin for (at least) some J subhaplogroups as proposed before (Richards et al. 1996; Richards et al. 2000). Haplogroup J diversities across Europe show in some cases distinct pattern. Namely, populations in the outskirts of Europe tend to be less diverse. For example, haplogroup J diversity among the Swedes in the Scandinavia is 0.705, diversity of haplogroup J is rising towards Central Europe 0.917 and Southern Europe 0.958 and declining in the Eastern Europe 0.852. There are some populations that show considerable low haplotype diversity in Europe, i.e. as mentioned above, the Albanians and the Ossetians (0.635 and 0.685 respectively), who harbor distinctive HVSI and HVSII sequence patterns that comprise most of their J haplotypes. Low haplotype diversity implies on the fact that these populations have experienced founder effect or population bottleneck in their demographic history. The Ossetian as well the Albanian populations were probably established by a small group of founders who harbored specific J haplotypes. Or during their demographic history have experienced reduction in population effective size (according to J haplogroup) magnifying the influence of genetic drift acting upon the population, leading to fixation of certain alleles. The Ossetians are almost entirely concentrated into J1a1 and J1c subclusters. The Albanians are concentrated into J1a and J1b subclusters.

Populations in the Balkans peninsula (Greece, Albania, and Creta) show distinctive J haplogroup patterns. When the Albanian population harobored uniform HVSI polymorphisms, then the Greeks exposed interesting and unique HVSII variation. For example, HVSII polymorphism at np 203 that was not detected in other European

populations. Additionally, J2 subclade is represented at high frequency (55%) among the Greeks, that of course, might indicate sampling effect.

Estimation of coalescence times

Coalescence times (state the beginning of population expansion in the certain region) for all the major nodes within J haplogroup that expressed more or less star-like topology were calculated (see Table 2). In some cases, when the branching pattern was being far from star-like for both Near Eastern as well as European haplogroup J graphs (as for J1c subcluster), the coalescence time was not estimated. When comparing two median joining networks, constructed of the European and Near Eastern J haplotypes (Figure 9 and Figure 10, respectively), we can see that J1b subclade harbors the most star-like topology in the European median joining network. Thus, the time estimate dating the subclade J1b population expansion will be more precise than that calculated i.e. for J2b2 subclade in Europe. As for J1 branch, J1b subclade is much younger in the European populations than in the Near Eastern ones. According to the present study, subclusters J1b as well as J1a2 date to the Neolithic times in Europe. More difficulties arise with the "Caucasian" share of J1 subcluster, J1a1. In the European median joining network subcluster J1a1 looks star-like, well, in a way. Coalescence time estimate obtained for the European J1a1 subcluster becomes, for some reason, unexpectedly high. Whereas the tip of J1a subclade, J1a2 in Europe, remains much younger. So, J1a1 is older in Caucasus than in Near East. There are some long branches protruding from the central node of Caucasian J1a1 subcluster (i.e. 16193-16362, 16189-16235 and 16189-16187) that may account for the old age of the Caucasian J1a1. When evaluating the coalescence time for this subcluster, the Ossetians, who have experienced population bottleneck or founder effect during their demographic history were excluded from the calculations for statistical reasons. Still, the time of population expansion becomes about 27000 years. Phenomenon alike the one described here, has been reported earlier, in case of subhaplogroup T1, whose expansion time was also evaluted to be about of the same magnitude in Caucasus (Metspalu et al. 1999). According to the coalescence times obtained for J1a1 subcluster we may assume that the population expansion in Caucasus occurred during the Upper Palaeolithic. Human dispersals might have started from Caucasus region , towards Near East and then, later on, during the

Table 2. Coalescence times calculated for the major nodes within J haplogroup in Europe and in Near East.

Subclade of J haplogroup	Coalescence time in Europe	Coalescence time in Near East
J1a1	27300 (standard error +/- 8000 years)	17700 (standard error +/- 2500 years)
J1a2	7700 (standard error +/- 3500 years)	-
J1b	5000 (standard error +/- 2200 years)	23300 (standard error +/- 4300 years)
J2a	19200 (standard error +/- 6900 years)	-
J2b1	-	15000 (standard error +/- 5000)
J2b2	161600 (standard error +/- 8100 years)	16000 (standard error +/- 5700 years)
J2b3	5800 (standard error +/- 2900 years)	-

Neolihtic, towards Europe. Nevertheless, subcluster J1a1 haplotype diversities in Caucasus are lower than those in Near East, increasing credibility to the fact that J1a1 subcluster still may descend from Near East. However, haplotype diversity is tightly connected with the number of samples under investigation, therefore, we can attribute the low haplotype diversity in Caucasus to small sample size. However, as discussed before, when dating population expansions by coalescence time estimates one has to be cautious, due to probable errors in time estimates that may arise from non-starlikeness of an evolutionary tree.

In case of J2 subhaplogroup, the time estimates evaluating the population expansions, become older compared to those of J1 subhaplogroup. That will apparently account for Europe as well as for Near East although in the current study there were no Near Eastern haplotypes belonging to J2a subclade represented. Therefore, J2 subhaplogroup in Europe can be considered as “old”, probably descending from the population expansions during the Upper Palaeolithic. The expansion of the major clades in J1 subhaplogroup (J1a and J1b) has started around Last Glacial Maximum (about 20 000 YBP) in Near East, J2b subcluster expanded about at the same time or later, during the early Natufian, when extreme aridity in Anatolia was over and the Caucasus glaciers started to melt. In conclusion, the picture emerges showing that J haplogroup (as a whole) is older in Near East than in Europe. There has been a remarkable gene flow

from the Near East into Europe during European prehistory. All the European major haplogroup J clusters date back to the Neolithic, the only exception is the Caucasian J1a1, that might rather originate from Caucasus than from Near East. Due to its location between the Near East, Europe and Central Asia, genetics of the Caucasus region populations may have an important role in the reconstruction of the dispersal routes of modern humans, including from the very beginning of the colonization of Eurasia. On the other hand, a possibility exists that some lineages that are by now widely spread in Near East, might have emerged from there.

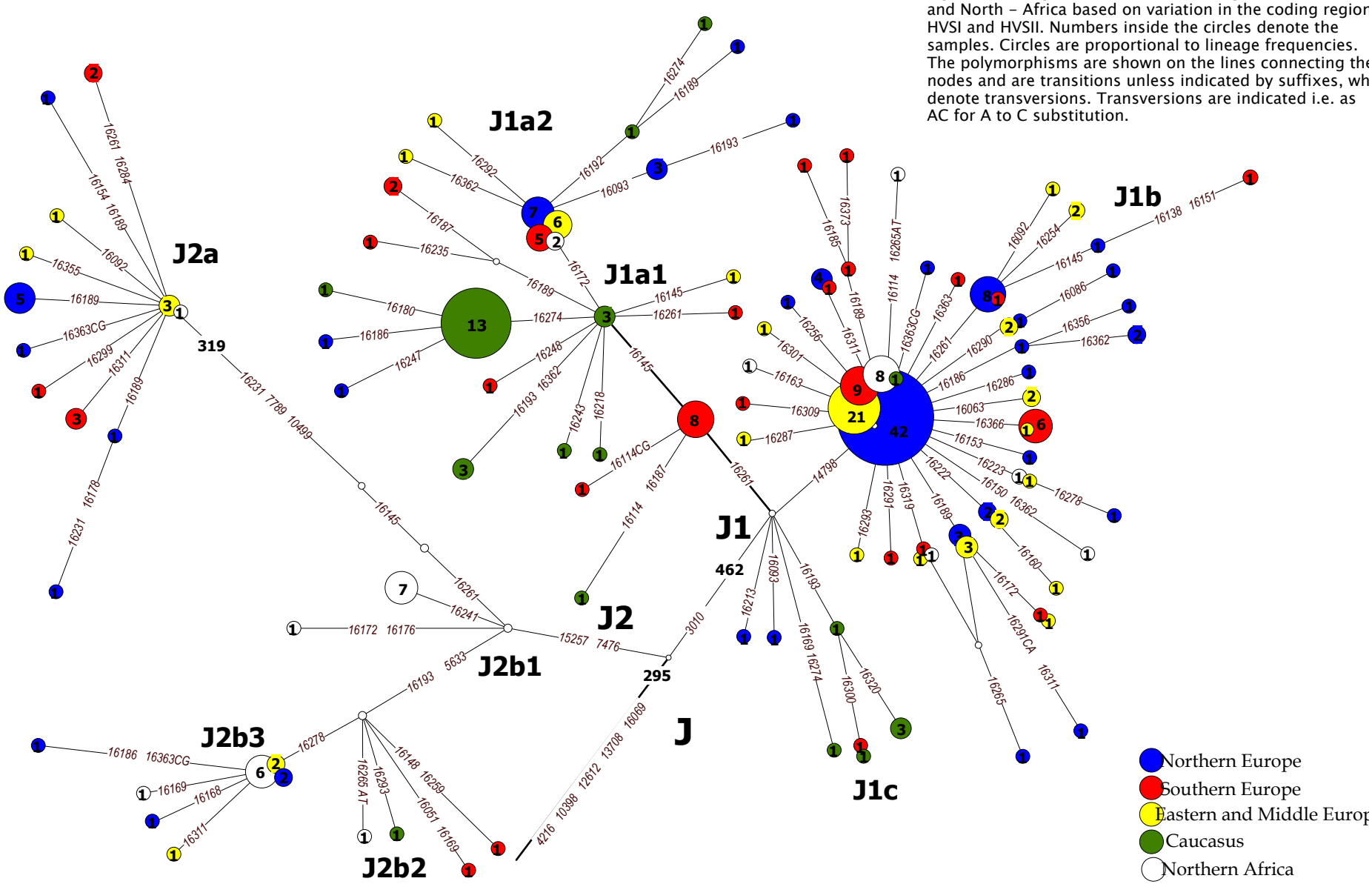
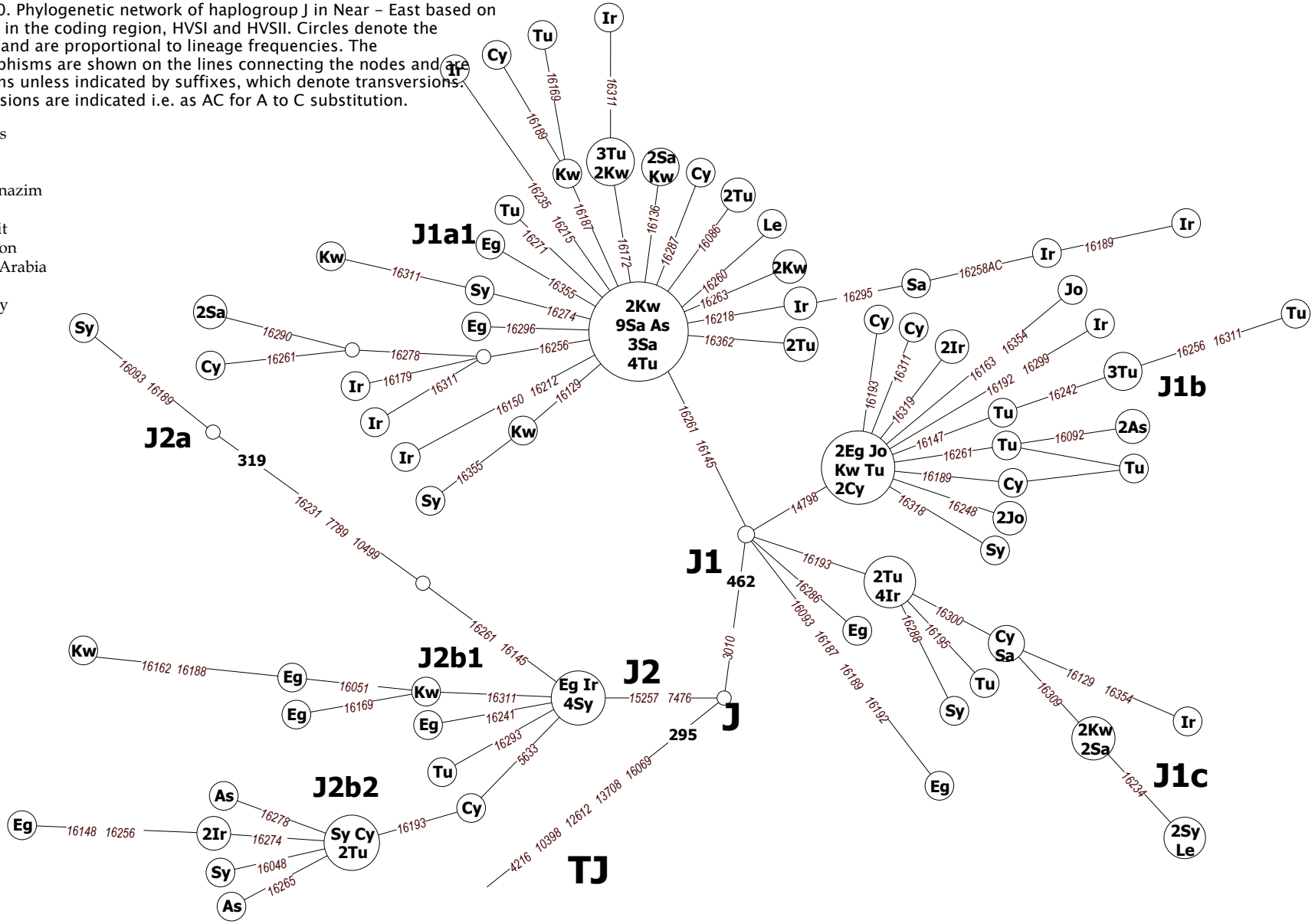


Figure 9. Phylogenetic network of haplogroup J in Europe and North – Africa based on variation in the coding regior HVSI and HVSII. Numbers inside the circles denote the samples. Circles are proportional to lineage frequencies. The polymorphisms are shown on the lines connecting the nodes and are transitions unless indicated by suffixes, wh denote transversions. Transversions are indicated i.e. as AC for A to C substitution.

Figure 10. Phylogenetic network of haplogroup J in Near – East based on variation in the coding region, HVSI and HVSII. Circles denote the samples and are proportional to lineage frequencies. The polymorphisms are shown on the lines connecting the nodes and are transitions unless indicated by suffixes, which denote transversions. Transversions are indicated i.e. as AC for A to C substitution.

Cy- Cyprus
Eg- Egypt
Ir- Iran
As- Ashkenazim
Jo- Jordan
Kw- Kuwait
Le- Lebanon
Sa- Saudi Arabia
Sy- Syria
Tu- Turkey



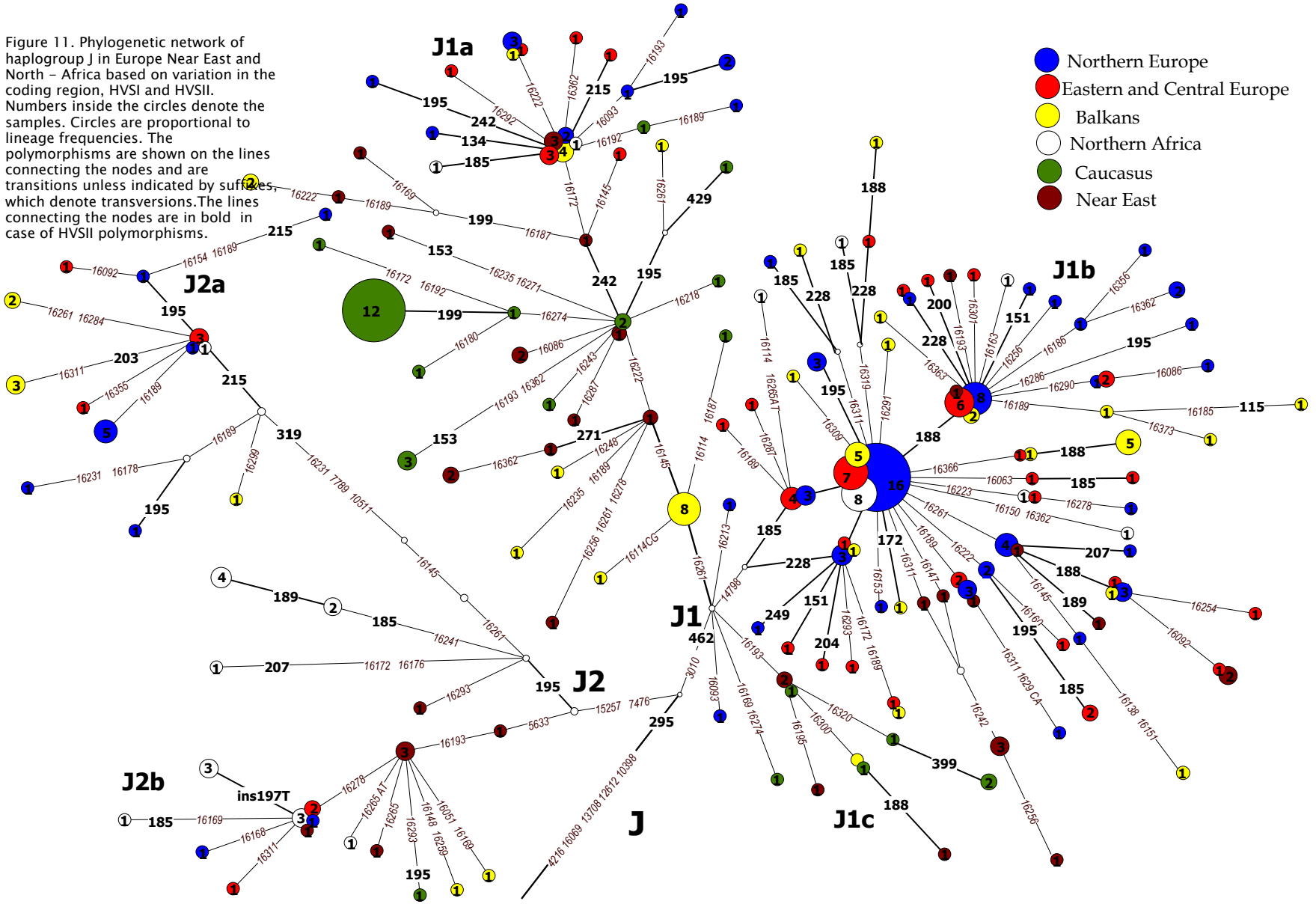


Figure 11. Phylogenetic network of haplogroup J in Europe Near East and North – Africa based on variation in the coding region, HVSI and HVSII. Numbers inside the circles denote the samples. Circles are proportional to lineage frequencies. The polymorphisms are shown on the lines connecting the nodes and are transitions unless indicated by suffixes, which denote transversions.The lines connecting the nodes are in bold in case of HVSII polymorphisms.

Summary

The current study is in concordance with the previous statements referring to the Near Eastern ancestry of some European J subclusters dating back to the Neolithic times. This probably accounts for the vast majority of European J subclusters, especially for subcluster J1b that is being the most frequent and diverse among the European J lineages. Still, J1a1 subcluster in Near East may originally descend from the Caucasus that according to time estimates dating the population dispersal, is about 27 000 years old in the Caucasus region.

According to the present study, in order to refine mtDNA J haplogroup's genealogy, information about polymorphisms in the mtDNA coding region as well as from HVSI and HVSII was combined. . The most parsimonious evolutionary tree for J haplogroup was presented according to HVSI, HVSII and coding region polymorphisms.According the results of the current thesis, conclusion can be drawn that HVSII region contains little information in resolving J haplogroup's topology. Although several HVSII polymorphisms are specific to J haplotypes on the whole, they can not be used in order to increase the phylogenetic resolution of J haplogroup's topology because of a high number of substitutions that occur at these nucleotide positions. Three polymorphisms were found to be of use in evolutionary studies. Polymorphism at np 295 is one of the polymorphisms defining the root of J haplogroup; polymorphism at np 319 is one of the polymorphisms specific to J2a subclade and polymorphism at np 462 is one of the polymorphisms that defines the J1 subhaplogroup

Kokkuvõte (Summary in Estonian)

Inimese mitokondri genoom on 16569 aluspaari pikkune DNA rõngasmolekul. Enamuses inimese rakkudest on kuni mitu tuhat mitokondriaalse DNA (mtDNA) koopiat. MtDNA-d iseloomustab unilineaarne päritavus, ta pärandub järglastele ainult emapoolselt vanemalt ega rekombineeru. Mitokondriaalse DNA võib jagada kaheks fuktsionaalselt erinevaks osaks: kodeerivaks ja mittekodeerivaks piirkonnaks. Mittekodeeriv ala omakorda jagatakse kaheks hüpervarieeruvaks piirkonnaks. Mittekodeeriv osa evolutsioneerub umbes kümme korda kiiremini kui kodeeriv osa, mille evolutsiooni aeglustavad funktsionaalsed piirangud. Mitterekombineeruvus, kiire ja aeglase evolutsiooniga piirkondade olemasolu ning suur koopiaarv teevad mitokondriaalsest DNA-st sobiva tööriista inimpopulatsioonide populatsioonide päritolu ja ajaloo uurimisel. MtDNA nö. võidukäik inimkonna demograafilise ajaloo uurimisel algas 1980-ntatel aastatel, kui sekveneeriti esimene inimese mtDNA täisjärjestus.

Käesoleva töö eesmärgiks oli suurendada mtDNA haplogupp J fülogeneesi lahutusvõimet, uurida J haplogrupi topoloogiat ja vanust nii Lähis-Idas kui ka Euroopas. Selleks analüüsiti nii eelnevalt määratud J haplogrupi täisjärjestusi kui ka I hüpervarieeruva piirkonna polümorfisme ja vastavalt kirjanusest saadud andmetele, otsustati täpsemalt uurida II hüpervarieeruva ala polümorfset mustrit. Käesolevas töös uuriti täpsemalt 306 Eurooplase ja 207 Lähis-Idast pärineva J haplogruppi kuuluva indiviidi mtDNA II hüpervarieeruva regiooni polümorfismide mustrit ja kombineerides seda nii I hüpervarieeuva piirkonna kui ka kodeeriva ala polümorfismidega, esitati vastavalt saadud andmetele kõige “säästlikum” J haplogrupi evolutsiooniline puu. Selgus, et II hüpervarieeruv piirkond üldiselt ei ole piisavalt informatiivne selleks, et seda edspidi J haplogrupi topoloogia uurimisel kasutada. Sellest hoolimata leiti II hüpervarieeruvast piirkonnast kolm polümorfismi (295, 319, 462), mida saab J haplogrupi puhul lugeda defineerivateks. Vastavalt analüüsitud I hüpervarieeruva piirkonna polümorfismidele jagati J haplogrupp J1 ja J2 alamhaplogruppideks, need alamhaplogrupid omakorda alamklastriteks.

J haplogrupi vanus Lähis-Idas on tunduvalt suurem kui Euroopas. Enamus Euroopa J haplogrupi alamklastreid pärineb ilmselt neoliitilisest ekspansioonist Lähis-Idast. Seda seostatakse nö. “neoliitilise revolutsiooniga”, maaviljeluse ja karjakasvatamise algusega Lähis-Idas, mis tõi kaas rahvaarvu suurenemise vastavas piirkonnas ning populatsiooni ekspansiooni Euroopa suunas. Samas J2 alamhaplogrupp Euroopas võib pärineda

neoliitikumi eelsest paleoliitilisest ekspansioonist. Vastavalt J haplogrupile koonduvad Kaukaasia rahvad põhiliselt J1a1 alamklastrisse, mille vanuseks käesolevas töös hinnati ligikaudu 27 000 aastat. Sama alamklastri vanuseks Lähis-Idas aga ligikaudu 18 000 aastat. Seega hoopis Kaukaasia võib olla mõnede J hapogrupi liinide algkoduks.

Kokkuvõttes võib öelda, et töö tulemusena saavutati J haplogrupi senisest märksa sügavam fülogeneetiline resolutsioon ja saadud tulemused on nüüdsest peale kasutatavad fülogeograafilises analüüsis. Siiski nõuab J haplogrupi evolutsioonilise puu topoloogia detailiseerimine suurt edasist tööd: kodeeriva ala täielikku polümorfismide andmestikku ja selle käigus leitud uute mutatsioonide uurimist nende fülogeograafilise informatiivsuse selgitamiseks.

Acknowledgements

It has been a privilege to have Professor Richard Villems and Dr. Ene Metspalu as supervisors. I thank them both for the professional guidance in this interesting field of science. Many thanks to Ille Ilpus, Hela Help, Jaan Lind and Jüri Parik for technical assistance. Special thanks to Professor Elsa Khusnutdinova for access of large set of samples, Marina Bermisheva, Oleg Balanovsky, Larissa Tamba, Marina Gubina, Vladislava Gusar, Lovorka Barać, Marijana Perečić, Svetlana Cvietan, Erwan Pennarun, and Doron Behar. I would like to express my gratitude to all the co-workers in our laboratory whose previous work has supplied me with J haplogroup HVSI sequences: Katrin Kaldma, Ene Metspalu, Toomas Kivisild, Kristiina Tambets, Jüri Parik, Sirle Laos and Karmen Mae.

Publications

Tambets K, Rootsi S, Kivisild T, Help H, **Serk P**, Loogvali EL, Tolk HV, Reidla M, Metspalu E, Pliss L, Balanovsky O, Pshenichnov A, Balanovska E, Gubina M, Zhadanov S, Osipova L, Damba L, Voevoda M, Kutuev I, Bermisheva M, Khusnutdinova E, Gusar V, Grechanina E, Parik J, Pennarun E, Richard C, Chaventre A, Moisan JP, Barac L, Pericic M, Rudan P, Terzic R, Mikerezi I, Krumina A, Baumanis V, Koziel S, Rickards O, De Stefano GF, Anagnou N, Pappa KI, Michalodimitrakis E, Ferak V, Furedi S, Komel R, Beckman L, Villems R.

The western and eastern roots of the Saami--the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes.
Am J Hum Genet. 2004 Apr;74(4):661-82

References

Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444-61

Ammerman AJ, Cavalli-Sforza LL (1984) The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-65

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13:494-504

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139-52

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37-48

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743-53

Barbujani G, Bertorelle G (2001) Genetics and the population history of Europe. Proc Natl Acad Sci U S A 98:22-5.

Bendall KE, Macaulay VA, Baker JR, Sykes BC (1996) Heteroplasmic point mutations in the human mtDNA control region. Am J Hum Genet 59:1276-87

Bellwood P (2001) Annu Rev Anthropol 30: 181 - 207

Brown MD, Starikovskaya E, Derbeneva O, Hosseini S, Allen JC, Mikhailovskaya IE, Sukernik RI, et al (2002) The role of mtDNA background in disease expression: a new primary LHON mutation associated with Western Eurasian haplogroup J. Hum Genet 110:130-8.

Brown WM (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci U S A 77:3605-9

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31-6

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57:133-149

Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, et al (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331-50

Crespillo M, Luque JA, Paredes M, Fernandez R, Ramirez E, Valverde JL (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. Int J Legal Med 114:130-2

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Elson JL, Samuels DC, Turnbull DM, Chinnery PF (2001) Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. Am J Hum Genet 68:802-6.

Excoffier L, Smouse PE (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. Genetics 136:343-59.

Finnilä S, Majamaa K (2001) Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population. J Hum Genet 46:64-9

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935-945

Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A 77:6715-9

Hartl DL, Clark AG (1989) Principles of population genetics. Sinauer Associates, Inc, Sunderland, Mass. U.S.A., pp 682

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol 37:347-354

Hasegawa M, Kishino H, Hayasaka K, Horai S (1990) Mitochondrial DNA evolution in primates: transition rate has been extremely low in the lemur. J Mol Evol 31:113-21

Hauswirth WW, Clayton DA (1985) Length heterogeneity of conserved displacement loop sequence in human ,mitochondrial DNA Nuc Ac Res 13: 8093-104

Helgason A, Sigurdadottir S, Gulcher J, Ward R, Stefanson K (2000) mtDNA and the origins of the Icelanders: deciphering signals of recent population history. Am J Hum Genet 66:999-1016

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, 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.

Howell N, Bindoff LA, McCullough DA, Kubacka I, Poulton J, Mackey D, Taylor L (1991) Leber hereditary optic neuropathy: indentification of the same mitochondrial ND1 mutation in six pedigrees. Am J Hum Genet 49:939-950

Huoponen K, Schurr TG, Chen Y, Wallace DC (2001) Mitochondrial DNA variation in an aboriginal Australian population: evidence for genetic isolation and regional differentiation. Hum Immunol 62:954-69.

Ingman M, Gyllensten U (2001) Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. J Hered 92:454-61.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708-713

Johns DR, Neufeld MJ (1992) Cytochrome b mutations Leber hereditary optic neuropathy. Biochem Biophys Res Commun 181:1358-1364

Johns DR, Heher KL, Miller NR, Smith KH (1993) Leber's hereditary optic neuropathy: clinical manifestations of the 14484 mutation. Arch Ophthalmol 11:495-498

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, et al (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331-1334

Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737-1751

Kolman C, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321-34

Krings M, Salem AE, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, et al (1999) mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? Am J Hum Genet 64:1166-76.

Larruga JM, Diez F, Pinto FM, Flores C, Gonzalez AM (2001) Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. Eur J Hum Genet 9:708-16.

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic

mitochondrial lineages delineate early human expansions. BMC Genet 2:13

Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, et al (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232-249

Maddison D (1991) African origin of human mitochondrial DNA reexamined. Syst Zool 40:355-363

Metspalu E, Kivisild T, Kaldma K, Parik J, Reidla M, Tambets K, Villems R (1999) The Trans-Caucasus and the Expansion of the Caucasoid-Specific Human Mitochondrial DNA. In: Papiha S, Deka R, Chakraborty R (eds) Genomic Diversity: Application in Human Population Genetics. Kluwer Academic / Plenum Publishers, New York, pp 121-134

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York, pp 145-63

Ohlenbusch A, Wilichowski E, Hanefeld F (1998) Characterization of the mitochondrial genome in childhood multiple sclerosis. I. Optic neuritis and LHON mutations. Neuropediatrics 29: 179-9

Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60:321-30

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23:437-441

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet 62:531-50

Redd AJ, Stoneking M (1999) Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet 65:808-28

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk H, Parik J, et al (2003) Origin and Diffusion of mtDNA Haplogroup X. Am J Hum Genet. 73:1178-90

Renfrew C (1987) Archaeology and Language: the Puzzle of Indo-European Origins. Jonathan Cape, London

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, et al (1996) Paleolithic and neolithic lineages in the European mitochondrial gene

pool. Am J Hum Genet 59:185-203

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, et al (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251-1276

Richards MB, Macaulay VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241-60

Rose G, Passarino G, Carrieri G, Altomare K, Greco V, Bertolini S, Bonafè M, Franceschi C, De Benedictis G (2001) Eur Jour of Hum Genet 9:701-707

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718-726

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R Martin W (2002) Genetics and geography of wild cereal domestication in Near East Nat Rev Genet 3:429-41

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, et al (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613-23

Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. Am J Hum Genet 59:964-8

Sigurđardóttir, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) Am J Hum Genet 66:1599-1609

Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029-32.

Stringer CB, Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. Science 239:1263-8

Templeton AR (1992) Human origins and analysis of mitochondrial DNA sequences. Science 255:737

Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, et al (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137-52

Torroni A, Chen YS, Semino O, Santachiara-Beneceretti AS, Scott CR, Lott MT, Winter M, et al (1994) mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico [see comments]. Am J Hum Genet 54:303-18

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, et al (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835-50

Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC (1994) mtDNA and the origin of Caucasians: identification of ancient Caucasian- specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet 55:760-76

Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. Proc Natl Acad Sci U S A 91:1158-62

Torroni A, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, Carrara F, Carducci C, et al (1997) Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. Am J Hum Genet 60:1107-21

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley R, Oefner PJ, et al (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet. 65:43-62

Van Andel T, Runnels C (1995) The earliest farmers in Europe. Antiquity 69:481-500

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503-7

Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37:613-23

Wallace D (1995) Mitochondrial DNA variation in human evolution, degenerative disease, and aging. Am J Hum Genet 57:201-23

Wolpoff M, Wu XZ, Thorne AG (1984) Modern Homo sapiens origins: A general theory of hominid evolutioninvolving the fossil evidence from East Asia. In: Smith FH, Spencer F (eds) The origin of modern humans: A world survey of the fossil evidence. Alan L. Riss, New York, pp 411-483

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635-51.

Appendixes

Table 3. European mtDNA haplotypes according to their HVSI, HVSII and coding region variation analysed in the current study. In case of HVSI, the first two digits, 16, are omitted, + and – characters refer to formation or loss of nucleotide at subsequent position, creating a restriction site for corresponding enzyme. CRS refers to Cambrige Reference Sequence.

				-3010	-7476	-13704	+56633	7789	-10398	+42161	+12612	10499	14798	-152257	
POPULATION	SAMPLE	HVSI	HVSII	*Bsh*1230 6I	*Alu*I	*Bst*OI	*Alu*I	CRS-C	*Dde*I	*Nla*III	*Msl*I	CRS-A	CRS-T	*Xmi*I	J sub haplogroup
Armenia	4	69-114-126-187-261	73-185- 263-295-ins316C-462	-	+	-			-				T		J1a
Armenia	100	69-126-145-218-222-261	73-263-295-ins316C-462	-	+	-			-				T		J1a
Armenia	72	69-126-193	73-263-295-ins316C-462	-	+	-			-				T		J1c
Armenia	40	69-126-193-300	73-152-263-295-ins317C-462	-	+	-			-				T		J1c
Armenia	83	69-126-145-222-261-274	73-263-295-ins317C-462-489	-	+	-			-				T		J1a
Armenia	11	69-126-145-222-261	73-263-295-ins316C-462	-	+	-			-				T		J1a
Armenia	75	69-126-145-222-261	73-195-263-295-ins310C-ins317C-429-462	-	+	-			-				T		J1a
Armenia	88	69-126-145-222-261	73-263-295-ins316C-462	-	+	-			-				T		J1a
Ossetia	27	69-126-145-172-192-222-261-274	73-146-263-295-ins310C-ins316C-462 -523-524delCA	-	+	-			-				T		J1a
Ossetia	54	69-126-145-193-222-261-362	73-153-263-295-ins317C-459-462	-	+	-			-				T		J1a
Ossetia	81	69-126-193-320	73-152-263-295-ins310C-ins317C-462	-	+	-			-				T		J1c
Ossetia	130	69-126-145-222-261-274	73-199-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Ossetia	145	69-126-145-222-261-274	73-199-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Ossetia	21	69-126-145-222-243-261	73-263-295-ins317C-462	-	+	-			-				T		J1a
Ossetia	150	69-126-193-320	73-152-263-295-ins 310C-ins316C-399-462-ins574C	-	+	-			-				T		J1c
Ossetia	151	69-126-193-320	73-152-263-295-ins 310C-ins316C-399-462-ins574C	-	+	-			-				T		J1c
Ossetia	138	69-126-145-193-222-261-362	73-152-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Ossetia	188	69-126-145-193-222-261-362	73-152-263-295-ins316C-462	-	+	-			-				T		J1a
Ossetia	64	69-126-145-222-261-274	73-199-263-295-ins316C-462	-	+	-			-				T		J1a
Ossetia	65	69-126-145-222-261-274	73-199-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Ossetia	66	69-126-145-222-261-274	73-199-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Ossetia	69	69-126-145-222-261-274	73-199-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Ossetia	70	69-126-145-222-261-274	73-199-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a

Ossetia	71	69-126-145-222-261-274	73-199-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Ossetia	85	69-126-145-222-261-274	73-199-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Ossetia	86	69-126-145-222-261-274	73-199-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Ossetia	88	69-126-145-222-261-274	73-199-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Ossetia	93	69-126-145-222-261-274	73-199-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Ossetia	99	69-126-145-222-261-274	73-199-263-295-ins310C-ins316C-462-490	-	+	-			-				T		J1a
Ossetia	101	69-126-145-222-261-274	73-199-263-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Sweden	21	69-126-145-172-222-261	73-242-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Sweden	92	69-126	73-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	28	69-126-222	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	76	69-126-222	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	90	69-126	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	86	69-126-261	73-146-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	3	69-126-261	73-146-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	15	69-126-261	73-146-185-207-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	55	69-126-261	73-146-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	65	69-126-261	73-146-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	84	69-126-189	73-185-228-263-295-ins310C-ins316C-462	-	+	-			+				C		J1b
Sweden	64	69-126-145-189-231-261	73-150-152-263-295-ins310T-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Sweden	23	69-126-145-172-189-192-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Sweden	55	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	68	69-126	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	55	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	68	69-126	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	44	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	32	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	49	69-126	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	31	63-69-93-126	73-263-295-ins316C-462	-	+	-			-				T		J1
Sweden	104	69-126-223-278	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	36	69-126-286	73-185-188-195-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	52	69-126-311	73-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	34	69-126-261	73-146-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	88	69-126-145-231-261	73-ins150T-152-195-215-263-295-ins310T-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Sweden	70	69-126	73-185-195-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	77	69-126	73-185-195-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	13	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	50	69-126	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	112	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	58	69-126	73-228-263-295-ins316C-462	-	+	-			-				C		J1b
Sweden	61	69-126	73-185-228-263-295-inc316C-462	-	+	-			-				C		J1b
Sweden	84	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	139	69-126	73-185-228-263-ins310C-ins316C-462	-	+	-			-				C		J1b
Sweden	16	69-126	73-185-195-228-263-295-ins316C-462-482	-	+	-			-				C		J1b

Slovakkia	46	69-126-189	73-151-185-228-263-295-ins310C-ins316C-462	-	+	-			+				C			J1b
Slovakkia	87	69-126-290	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C			J1b
Slovakkia	79	69-126-223	73-146-185-228-263-295-ins316C-462	-	+	-			-				C			J1b
Slovakkia	33	69-126-160-222	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C			J1b
Slovakkia	86	69-126-293	73-228-263-295-ins316C-462	-	+	-			-				C			J1b
Slovakkia	27	69-126-301	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C			J1b
Slovakkia	111	69-126-311-319	73-185-263-295-ins310C-ins316C-462	-	+	-			+				C			J1b
Slovakkia	138	69-92-126-261	73-185-228-263-295- ins316C-462	-	+	-			-				C			J1b
Slovakkia	64	69-92-126-145-231-261	73-150-152-215-263-295- ins311T- ins316C-319	+	-	-	-	A	-			G	T	-		J2a
Slovakkia	12	69-126-193-278	73-150-263-295-ins310C-ins316C	+	-	-	+		-				T	-		J2b
Slovakkia	94	69-126-366	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C			J1b
Slovakkia	60	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C			J1b
Slovakkia	63	69-126	73-185-188-228-263-295-ins316C-462	-	+	-			-				C			J1b
Slovakkia	93	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C			J1b
Slovakkia	133	69-126	73-185-188-200-228-263-295-ins317C-462	-	+	-			-				C			J1b
Slovakkia	107	69-126-145-231-261	73-150-152-195-215-263-295-ins311T-ins317C-319	+	-	-	-	A	-			G	T	-		J2a
Slovakkia	125	69-126-145-231-261	73-150-152-195-215-263-295-ins310C-ins317C-319	+	-	-	-	A	-			G	T	-		J2a
Slovakkia	130	69-126-145-231-261	73-150-152-195-215-263-295-ins310T-ins316C-319	+	-	-	-	A	-			G	T	-		J2a
Czech	9	69-126-222-261	73-242-263-295- ins310C-ins316C-462	-	+	-			-				T			J1a
Czech	30	69-126-145-172-222-261-292	73-146-242-263-295-ins316C-462	-	+	-			-				T			J1a
Czech	23	69-126	73-185-263-295-ins316C-462	-	+	-			-				C			J1b
Czech	43	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C			J1b
Czech	51	69-126	73-185-188-228-263-295-ins316-462	-	+				-				C			J1b
Czech	26	69-126-145-172-222-261	73-242-263-295-ins316C-462	-	+	-			-				T			J1a
Czech	40	69-126-145-172-222-261	73-146-242-263-295-ins316C-462	-	+	-			-				T			J1a
Bosnia	B1	69-126-172-189	73-228-263-295-ins316C-462	-	+	-			-				C			J1b
Bosnia	B119	69-126-189	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C			J1b
Bosnia	B333	69-126-311	73-185-263-295-ins316C-462	-	+	-			-				C			J1b
Bosnia	B138	69-126-291	73-185-228-263-295-ins316C-462	-	+	-			-				C			J1b
Bosnia	B313	69-114CG-126-261	73-263-295-ins310C-ins316C-462	-	+	-			-				T			J1a
Bosnia	B39	69-126-145-231-261-299	73-150-152-195-263-295-ins311T-ins316C-319	+	-	-	-	A	-			G	T	-		J2a
Bosnia	B331	69-126-145-222	73-195-263-295-ins316C-462	-	+	-			-				T			J1a
Bosnia	B247	69-126-366	73-185-228-263-295-ins310C-ins316C-429-462	-	+	-			-				C			J1b
Bosnia	B84	69-126	73-146-185-188-228-263-295-ins316C-462	-	+	-			-				C			J1b
Bosnia	B208	69-126-145-172-261	73-242-263-295-ins310C-ins316C-462	-	+	-			-				T			J1a
Bosnia	B336	69-126-145-172-261	73-242-263-295-ins310C-ins316C-462	-	+	-			-				T			J1a
Bosnia	B332	69-126-145-172-222-261	73-242-263-295-ins316C-327-462	-	+	-			-				T			J1a
Bosnia	B34	69-126-145-172-222-261	73-242-263-295-ins316C-327-462	-	+	-			-				T			J1a
Bosnia	B334	69-126	73-172-185-228-295-ins310C-ins316C-462-482	-	+	+			-	+	+		C			J1b
Bosnia	B237	69-126	73-146-185-228-263-295-ins316C-462	-	+	-			-				C			J1b
Ukraine	138-01	63-69-126	73-228-263-295-ins316C-462	-	+	-			-				C			J1b
Ukraine	93-01	69-126-222	73-195-228-263-295-ins316C-462	-	+	-			-				C			J1b
Ukraine	122	69-126-254-261	73-185-228-263-295-ins310C-ins316C-462-489	-	+	-			-				C			J1b

Ukraine	93-01	69-126-222	73-195-228-263-295-ins316C-462-489	-	+	-			-				C		J1b
Ukraine	102-02	69-126-172-189	73-228-263-295-ins316C-462	-	+	-			+				C		J1b
Ukraine	78-01	69-126-189	73-151-185-228-263-295-ins310C-ins316C-462	-	+	-			+				C		J1b
Ukraine	25-02	69-126-287	73-185-263-295-ins316C-368-462	-	+	-			-				C		J1b
Ukraine	131-02	69-126-145-231-261-355	73-150-152-215-195-263-295-ins311T-ins316C-319-430	+	-	-	-	A	-			G	T	-	J2a
Ukraine	161-02	69-126-145-172-222-261-362	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Ukraine	52-01	69-126-193-278-311	73-150-152-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Ukraine	104-01	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	111-01	69-126	73-146-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	126-01	69-126	73-185-263-265-ins316C-462	-	+	-			-				C		J1b
Ukraine	13-01	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	68-02	69-126	73-146-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	9-Jan	69-126	73-204-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Ukraine	93-02	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	97-02	69-126	73-151-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Ukraine	24	63-69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	70	69-126-189	73-150-185-263-295-ins316C-462	-	+	-			+				C		J1b
Ukraine	29	69-126-290	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	116	69-126-145-172-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Ukraine	122	69-126-254-261	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Ukraine	39	69-126	73-185-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	50	69-126	73-185-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	119	69-126	73-228-263-295-ins316C-462	-	+	-			-				C		J1b
Ukraine	135	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Latvia	14	69-126-153	73-185-228-263-295-ins316C-462-482	-	+	-			-				C		J1b
Latvia	86	69-126-189-291CA-311	73-185-228-263-295-ins310C-ins316C-462	-	+	-			+				C		J1b
Latvia	44	69-126-145-189-231-261	73-150-152-195-215-263-295-ins303C-ins310C-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Latvia	29	69-126	73-185-228-249-263-295-ins316C-462	-	+	-			-				C		J1b
Latvia	41	69-126	73-150-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Latvia	20	69-126	73-185-263-295-ins316C-462	-	+	-			-				C		J1b
Latvia	27	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Latvia	51	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Latvia	32	69-126	73-146-151-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Lithuania	50	69-126	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Lithuania	49	69-126-290	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Lithuania	18	69-126-145-189-231-261	73-150-152-195-215-263-295-ins310C-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Lithuania	47	69-126-193-278	73-150-263-295-ins310C-ins316C	+	-	-	-		-				T	-	J2b
Lithuania	15	69-86-126-290	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Estonia	304	69-126-186	73-185-188- 228-263-295-ins310C- ins316C-462	-	+	-			-				C		J1b
Estonia	150	69-126-186-356	73-185-188-228-263-295- ins310C-ins316C-462	-	+	-			-				C		J1b
Estonia	276	69-126-145-178-189-261	73-150-152-195-215-263-ins310C-ins316C-319	+	-	-	-	A	+			G	T	-	J2a
Estonia	495	69-126-145-231-261	73-150-215-263-295- ins310T-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Estonia	415	69-126-145-154-189-231-261	73-150-215-263-295- ins310C-ins316C-319	+	-	-	-	A	-			G	T	-	J2a

Estonia	308	69-93-126-145-172-193-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Estonia	519	69-126-168-193-278	73-146-150-263-295-ins310C-316insC	+	-	-	+	G	-				T	-	J2b
Estonia	155	69-126-189	73-185-228-263-295-ins316C-462	-	+	-			+				C		J1b
Estonia	453	69-126-189	73-185-228-263-295-ins316C-462	-	+	-			+				C		J1b
Estonia	43	69-126-213	73-263-295-ins310C- ins316C-462	-	+	-			-				T		J1
Estonia	524	69-126-145	73-146-150-185-188- 228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	199	69-126-186-362	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Estonia	306	69-126-186-362	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Estonia	29	69-261	73-146-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	273	69-261	73-146-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	33	69-126-145-189-231-261	73-150-152-195-215-263- 295-ins310C-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Estonia	258	69-126-145-189-231-261	73-150-152-195-215-263-295-ins310C-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Estonia	454	69-126-145-189-231-261	73-150-152-195-215-263-295-ins310C-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Estonia	61	69-126-145-172-261	73-242-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Estonia	267	69-126-145-172-261	73-242-263-295-ins316-462	-	+	-			-				T		J1a
Estonia	369	69-126-145-172-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Estonia	126	69-93-126-145-172-222-261	73-242-263-295- ins316C-462	-	+	-			-				T		J1a
Estonia	162	69-93-126-145-172-222-261	73-195-242-295-ins310C-ins317C-462	-	+	-			-				T		J1a
Estonia	166	69-93-126-145-172-222-261	73-195-242-263-295-ins317C-462	-	+	-			-				T		J1a
Estonia	52	69-126-145-172-222-261	73-134-242-263-295-ins316C-462	-	+	-			-				T		J1a
Estonia	136	69-126-145-172-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Estonia	424	69-126-145-172-222-261	73-195-263-295-ins316C-462	-	+	-			-				T		J1a
Estonia	51	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	154	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	182	69-126	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	224	69-126	73-146-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	253	69-126	73-185-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	270	69-126	73-185-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	300	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	332	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	372	69-126	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Estonia	452	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Estonia	piia	69-126-256	73-150-185-188-228-263-295-ins316C-462	-	+	-	-		-				C		J1b
Hungary	BuCa018	69-126-145-222-261-172	73-215-242-263-295-ins316C-327-462	-	+	-			-				T		J1a
Hungary	BuCa220	69-126-145-172-261	73-242-263-ins310C-ins316C-309-462	-	+	-			-				T		J1a
Hungary	BuCa039	69-126-193-278	73-150-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Hungary	BuCa008	69-126	73-185-188-263-295-ins303C-ins310C-ins316C-462	-	+	-			-				C		J1b
Hungary	BuCa015	69-126	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Hungary	BuCa153	69-126	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Turkey	44	69-126	73-185-228-263-295-ins316C-462-482	-	+	-			-				C		J1b
Turkey	289	69-126-147	73-185-228-263-295-ins316C-462	+	+	+			-	+	+		C	+	?
Turkey	108	69-126-293	73-150-195-263-295-ins310C-ins316C	+	-	-	-	G	-				T	-	J2
Turkey	272	69-126-147-242-256-311	73-152-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b

Turkey	166	69-126-145-169-187-222-261	73-146-199-242-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	70	69-126-145-222-235-261-271	73-153-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	349	69-126-193-195	73-146-152-263-295-ins316C-462	-	+	-			-				T		J1c
Turkey	46	69-126-261	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Turkey	359	69-126-261	73-185-189-228-263-295-ins316C-462	-	+	-			-				C		J1b
Turkey	2	69-126-145-261	73-150-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	294	69-126-145-261	73-146-263-271-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	260	69-126-145-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Turkey	122	69-126-147-242-311	185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Turkey	303	69-126-147-242-311	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Turkey	347	69-126-147-242-311	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Turkey	160	69-126-145-261-362	73-152-263-271-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	246	69-126-145-261-362	73-263-271-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	393	69-86-126-145-222-261	73-150-263-266-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	394	69-86-126-145-222-261	73-150-263-266-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Turkey	183	69-126-193	73-152-263-295-is310C-ins316C-462	-	+	-			-				T		J1c
Turkey	255	69-126-193	73-150-152-263-295-ins316C	+	-	-	+		-				T	-	J2b
Turkey	317	69-126-193	73-150-152-263-295-ins316C	+	-	-	+		-				T	-	J2b
Turkey	356	69-126-193	73-150-152-263-295-ins310C-ins316C-462	-	+	-			-				T		J1c
Turkey	116	69-126-145-172-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Turkey	287	69-126-145-172-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a
Turkey	319	69-126-145-172-222-261	73-146-242-263-295-ins316C-462	-	+	-			-				T		J1a
Turkey	271	69-126-145-172-222-261	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		?
Albania	12	51-69-126-169-193	73-150-152-263-295-ins316C	+	-	-	+		-				T	-	J2b
Albania	55	69-126	73-185-188-228-263-ins310C-ins316C-462	-	+	-			-				C		J1b
Albania	190	69-126-145-172-261	73-242-263-295-310C-316C-462	-	+	-			-				T		J1a
Albania	144	69-126-366	73-185-188-228-263-295-ins310C-ins316C-329-462	-	+	-			-				C		J1b
Albania	149	69-126-366	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Albania	41	69-126-366	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Albania	89	69-126-366	73-185-188-228-262-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Albania	117	69-126-366	73-150-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Albania	150	69-126-261	73-263-295-ins310C-ins316C-462-489	-	+	-			-				T		J1a
Albania	158	69-126-261	73-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Albania	169	69-126-261	73-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Albania	172	69-126-261	73-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Albania	93	69-126-261	73-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Albania	94	69-126-261	73-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Albania	50	69-126-261	73-263-295-ins310C-ins311C-ins316C-462	-	+	-			-				T		J1a
Albania	34	69-126-261	73-263-ins310C-ins316C-462	-	+	-			-				T		J1a
Greece	65	69-126-145-189-235-261	73-263-295-ins310C-ins316C-462	-	+	-			+				T		J1a
Greece	42	69-CRS	73-228-263-295-ins316C-462	-	+	-			-				C		J1b
Greece	88	69-CRS	73A-263-ins310C-ins311C-ins316C-462	+	+	+			-	-	-			+	H
Greece	26	69-126-145-187-189-261	73-199-242-263-295-ins310C-ins316C-462	-	+	-			+				T		J1a

Greece	38	69-126-145-187-189-261	73-199-242-263-265-295-ins310C-ins316C-462	-	+	-			+				T		J1a
Greece	58	69-126-145-231-284	73-150-152-195-215-263-295-ins310C-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Greece	83	69-126-145-231-284	73-152-195-215-263-295-ins310C-ins311C-ins316C-319-441	+	-	-	-	A	-			G	T	-	J2a
Greece	64	69-126-145-231-261-311	73-150-152-195-203-215-263-295-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Greece	74	69-126-145-231-261-311	73-152-195-203-215-228-263-295-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Greece	100	69-126-145-231-261-311	73-150-152-195-203-215-263-295-ins316C-319	+	-	-	-	A	-			G	T	-	J2a
Creta	119	69-CRS	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Italy	1	69-126-319	73-185-188-263-ins316C-462	-	+	-			+				C		J1b
Italy	39	69-126-309	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Italy	103	69-126-185-189	73-115-185-188-228-263-295-ins310C-ins316C-462	-	+	-			+				C		J1b
Italy	32	69-126-189-373	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			+				C		J1b
Italy	163	69-126-363	73-146-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Italy	37	69-126-138-145-151-261	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Italy	45	69-126-193-300	73-152-263-295-ins310C-ins316C-462	-	+	-			-				T		J1c
Italy	35	69-126-148-193-259	73-146-150-152-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Italy	136	69-126-145-248-261	73-263-295-ins316C-462	-	+	-			-				T		J1a
Italy	11	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Italy	43	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Italy	123	69-126	73-146-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Italy	147	69-126	73-185-188-228-263-ins310C-ins316C-462	-	+	-			-				C		J1b
Cyprus	81	69-126-145-256-278	73-199-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Cyprus	73	69-126-189	73-185-228-263-295-ins310C-ins316C-462	-	+	-			+				C		J1b
Cyprus	108	69-126-311	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Cyprus	140	69-126-193-300	73-152-188-263-295-ins310C-ins 316C-462	-	+	-			-				T		J1c
Cyprus	6	69-126-145-187-189-261	73-199-242-263-295-ins310C-ins316C-462	-	+	-			+				T		J1a
Cyprus	111	69-126-145-222-261-287	73-263-295-ins310C-ins316C-462	-	+	-			-				T		J1a
Cyprus	25	69-126-193	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Cyprus	37	69-126-193	73-150-152-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Cyprus	50	69-126	73-185-188-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Cyprus	89	69-126	73-150-152-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Cyprus	117	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Kumyk	35kum	69-126-193-293	73-150-152-195-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Kumyk	77kum	69-126-169-274	73-171-208-263-294-295-ins310C-ins316C-462	-	+	-			-				T		J1
Jews	1363	69-92-126-261	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Jews	1387	69-126-145-222-261	73-152-263-295-ins310C-316C-462	-	+	-			-				T		J1a
Jews	1583	69-126-193-278	73-150-152-195-263-295-ins310C-ins312T-ins316C	+	-	-	+		-				T	-	J2b
Jews	1823	69-126-193-265	73-150-152-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Jews	4332	69-92-126-261	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Marocco	75B	69-126-241	73-185-195-263-295-ins316C-462	+	-	-	-	G	-			A	T	-	J2a
Marocco	88E	69-126-241	73-150-185-189-195-263-295-ins316C-462	+	-	-	-	G	-			A	T	-	J2a
Marocco	56E	69-126-223	73-185-228-263-295-ins316C-482-709	-	+	-			-				C		J1b
Marocco	61E	69-126-145-231-261	73-150-152-195-215-263-295-ins316C-319-462-513	+	-	+	-	A	-	+	+	G	T	-	J2a
Marocco	75E	69-126-145-172-222-261	73-242-263-295-ins316C-462	-	+	-			-				T		J1a

Marocco	91E	69-126-169-193-278	73-150-152-185-263-295-ins316C-462	+	-	-	+		-				T	-	J2b
Marocco	1F	69-126-193-278	73-150-152-ins197T-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Marocco	5F/AB	69-126	73-185-228-263-295-ins316C	-	+	-			-				C		J1b
Marocco	E38	69-126-319	73-143-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Marocco	A176	69-126-163	73-185-188-228-263-295-ins316C-462	-	+	-			-				C		J1b
Marocco	C76	69-114-126-265AT	73-185-263-295-ins316C-462	-	+	-			-				C		J1b
Marocco	A163	69-126-150-362	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Marocco	C15	69-126-172-176	73-150-195-207-263-295-ins316C	+	-	-	-		-				T	-	J2
Marocco	A158	69-126-145-172-222-261	73-185-242-263-295-ins316C-462	-	+	-			-				T		J1a
Marocco	A48	69-126-193-265AT	73-146-150-152-195-263-295-ins316C	+	-	-	+		-				T	-	J2b
Marocco	D4	69-126-241	73-150-185-189-195-263-295-ins316C	+	-	-			-				C	-	J1b
Marocco	F13	69-126-241	73-150-152-185-195-263-295-ins310C-ins316C	+	-	-			-				C	-	J1b
Marocco	A12	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Marocco	A123	69-126	73-185-228-263-295-ins316C-463	-	+	-			-				C		J1b
Marocco	B38	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Marocco	B56/BA	69-126	73-185-228-263-295-ins316C-462	-	+	-			-				C		J1b
Marocco	C42/BA	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Marocco	E51	69-126	73-185-228-263-295-ins310C-ins316C-462	-	+	-			-				C		J1b
Marocco	B8	69-126-193-278	73-150-152-197insT-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Marocco	C55/BA	69-126-193-278	73-150-152-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b
Marocco	C87	69-126-193-278	73-150-152-263-295-ins316C	+	-	-	+		-				T	-	J2b
Marocco	D41	69-126-193-278	73-150-152-263-295-ins316C	+	-	-	+		-				T	-	J2b
Marocco	D23	69-126-193-278	73-150-152-197insT-263-295-ins310C-ins316C	+	-	-	+		-				T	-	J2b