Phylogenetic Networks for the Human mtDNA Haplogroup T

David A. Pike

Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada

Abstract: We develop phylogenetic networks for mtDNA haplogroup T, based on information stored in the MitoSearch database at www.mitosearch.org. Analysing the structure of the resulting networks, we note that nucleotide 16296 appears to be unstable throughout the haplogroup. We also observe a cluster that does not fall within one of the established subgroups of haplogroup T and so we propose some revision to the haplogroup hierarchy in order to encompass this cluster.

Introduction

The human mitochondrial DNA molecule was first fully sequenced in 1981 by Anderson et al.; this sequence of 16,569 nucleotide base pairs has since become known as the *Cambridge Reference Sequence* and is often referred to by the acronym CRS. Numerous subsequent studies have revealed that mutations within the mtDNA genome are an effective tool with which to delve into aspects of population genetics and human migrations. In this regard, a phylogenetic tree of major mtDNA haplogroups has been developed, whereby each haplogroup is characterised by a particular set of mutational differences as compared to the CRS. A version of this tree that relies on coding region mutations appears in a manuscript by Herrnstadt et al. (2002).

Here we focus our attention on haplogroup T, which was first described as "group 2B" by Richards et al. (1996), who observed that the haplogroup was characterised by a pair of mutations (at nucleotide positions 16126 and 16294) within the first hypervariable region (HVR1) of the noncoding control region of the mtDNA genome. Shortly afterwards Torroni et al. (1996) associated the haplogroup with polymorphic restriction sites within the coding region, but also observed a correlation with HVR1 positions 16294 and 16296 (mutations at 16126 were observed in several samples, but were also lacking in a few others).

Haplogroup T is now generally associated with a number of polymorphisms, at nucleotide positions 16126 and 16294 within the noncoding region of the mtDNA genome and the following positions within the coding region: 709, 1888, 4216, 4917, 8697, 10463, 11251, 13368, 14905, 15452, 15607, and 15928 (Torroni et al. 1996; Macaulay et al. 1999; Finnilä and Majamaa 2001). Additional mutations, such as those at positions 73 and 16519 are common within haplogroup T, but are also found in several other haplogroups (Wilkinson-Herbots et al. 1996; Helgason et al. 2000).

Our inquiry into the structure of the phylogenetic network for haplogroup T stems from the growing number of individuals who are participating in genetic genealogy studies and find themselves to be members of the haplogroup. Searches for information about the haplogroup will, with some effort, reveal that it originated in the Near East approximately 46,500 years ago but is now most prevalent in Europe, where it is found to occur in up to 10% of some subpopulations (Richards et al. 1998; Helgason et al. 2001).

In the overall phylogenetic tree, haplogroup T is closest to haplogroup J, which is characterised by the HVR1 motif 16069–16126 (Torroni et al. 1994; Richards et al. 1996) as well as coding region mutations at 4216, 10398, 11251, 12612, 13708, and 15452 (Torroni et al. 1994; Macaulay et al. 1999; Finnilä and Majamaa 2001). Hence the parent haplogroup JT has the motif 4216–11251–15452–16126, with 16126 being the defining HVR1 mutation. When considering HVR1 mutations, it is therefore the additional mutation at 16294 that defines haplogroup T, whereas haplogroup J is distinguished by the mutation at 16069.

The scientific literature also contains a number of papers from the medical research community, in which attempts to correlate pathological conditions with haplogroup membership are made. For instance, a study conducted in Spain observed a greater rate of occurrence of reduced sperm motility among men in haplogroup T than was found with men in other haplogroups (Ruiz-Pesini et al. 2000). However, a more recent study conducted in Portugal concluded the haplogroup association not to be sound, and noted that care must be taken when attempting to draw conclusions about haplogroups when considering only a regional sampling of data (Pereira et al. 2005). Elsewhere it has

been reported that membership in haplogroup T may offer some protection against Alzheimer Disease (Chagnon et al. 1999; Herrnstadt et al. 2002) and also Parkinson's Disease (Pyle et al. 2005), but the cautionary words of Pereira et al. suggest that further studies may be necessary before reaching firm conclusions.

Searches for information about the haplogroup will also reveal that Russian Tsar Nicholas II was a member of haplogroup T, and that he and his brother, the Grand Duke George Alexandrovitch Romanov, both exhibited heteroplasmy at nucleotide position 16169 (Ivanov et al. 1996).

This information, while interesting, may not satisfy those whose primary interest is genetic genealogy and who are seeking some sense of place within the haplogroup. In this paper we construct a phylogenetic network based on information stored in the MitoSearch database at www.mitosearch.org, a public database designed to assist in the pursuit of genetic genealogy; individuals can enter their own mtDNA haplotype into the database in the hope of making contact with others who share their genetic signature (i.e., with potential relatives who share maternal kinship). In particular, we extract the data pertaining to haplogroup T and then construct a map based on this data set, so that individuals may determine their place within the haplogroup T family.

Subsequent to building phylogenetic networks, we conduct some analysis of the haplogroup and its subgroups. In so doing, we propose a revision to the haplogroup T subgroup hierarchy.

Methodology

The source for the data we use is, as mentioned above, the MitoSearch database found at www.mitosearch.org. For each sample, it contains the results of genetic analysis of the nucleotides in the interval 16001 to 16569, which encompasses the first hypervariable region (HVR1). Several of the database entries also report the results of genetic analysis for nucleotide positions 1 to 574 (this interval includes HVR2).

As of November 15, 2005, the MitoSearch database contained a total of 367 samples that had been classified as belonging to haplogroup T or one of its subgroups. As the majority of these samples had only been tested for mutations within the interval 16001 to 16569, we chose to limit our consideration to this interval alone. Alternatively, we could have opted to work with the minority of samples that had been fully tested for both HVR1 and HVR2, but such a choice would have been contrary to

our motivational goal of presenting a map that could be consulted by genetic genealogists, many of whom only have information for the HVR1 portion of their mtDNA genome.

The MitoSearch database does not store information pertaining to the coding region of the mtDNA genome. Not having coding region data and not utilising HVR2 data may partially inhibit our ability to construct phylogenetic networks in the sense that it is possible that some genetic branching may not be observed if its only evidence is located among mutations in these regions. Fortunately the part of the human mtDNA molecule that we are using (the HVR1 portion) is known to have the highest rate of variation of any part of the mtDNA genome (Greenberg et al. 1983; Kocher and Wilson 1991).

To date, five major subgroups of haplogroup T have been identified, and each is associated with a particular set of HVR1 mutations (Richards et al. 1998; Richards et al. 2000). These motifs, which are in addition to the HVR1 motif 16126–16294 that defines haplogroup T, are listed in Table 1.

Subgroup	Associated HVR1 Mutations
T1	16163–16186–16189
T2	16304
T3	16292
T4	16324
T5	16153

Table 1: Mutation Positions and Subgroups

The MitoSearch database allows for samples to be identified with subgroups and so the 367 haplogroup T samples have been partitioned into seven disjoint subsets (T1 to T5, as well as two others named T and T*); the number of samples in each subset is shown in Table 2.

Subgroup	T	T*	T1	T2	T3	T4	T5
Samples	47	61	76	144	17	7	15

Table 2: Samples in the MitoSearch database

As the intent of the T and T* subsets is perhaps less apparent than for subsets T1 through T5, a short review of haplogroup nomenclature may be useful: Richards et al. (1998) proposed that the star designation (e.g. T*) should be used for each sample that belongs to a haplogroup (e.g. T) but not one of its known subgroups (e.g. not to any of T1 through T5). However, a growing number of genetic genealogists are having their DNA analysed (for instance,

through participation in the Genographic Project) and are being informed that their haplogroup is simply T, rather than a more refined classification (such as T* or one of T1 through T5) even if there is evidence to suggest that a particular subgroup applies. Hence the 47 samples in the T subset may permit a more specific classification than their position in the MitoSearch database suggests. Considering also that the MitoSearch database allows individuals to specify their haplogroup during the data entry process, we cannot assume that the seven data sets contain only samples that are correctly assigned to them. Given these concerns, we process all 367 samples as a single data set and we use the diagnostic motifs listed in Table 1 to determine each sample's subgroup classification.

To comment on the fitness of the 367 samples, nine of them lacked the combination 16126–16294 that classifies haplogroup T. Of these nine, six were identical to the CRS and have likely been mis-entered into the MitoSearch database (either as belonging to haplogroup T or else as having no differences with respect to the CRS) and so we removed these six samples from the data set. One of the other three exceptional samples has the sequence 16294–16304–16519; it appears to have experienced a back-mutation at nucleotide 16126, but otherwise it belongs to subgroup T2. The remaining two other cases share the same set of mutations (16126–16292–16296–16304–16311–16519) and so appear to have had a back-mutation at nucleotide 16294. However, assignment of these two samples to a subgroup would seem to be problematic, given that they possess the motifs for both the T2 and the T3 subgroups.

The number of distinct haplotypes that are represented by the the 361 samples remaining in our data set is 121. Overall the three most frequent haplotypes occur 83, 52, and 20 times and collectively account for 42.9% of the data set.

In the network diagrams that we now describe, each node represents a distinct haplotype, with the size of each node representing the corresponding number of samples. However, node sizes are not linear; rather a logarithmic scale is used for the radii, meaning that the largest nodes represent dominant haplotypes. When a node represents more than a single sample, the node's label begins with the number of samples that share the haplotype, with this number being enclosed in curly braces { and }. This enumeration is omitted if the node represents only one sample. The remainder of each node's label consists of the haplotype for the node. Here we use the format adopted by MitoSearch. Since each nucleotide

under consideration has a position in the 16000s, the leading 16 is omitted, so that positions are reported as one of 001, 002, 003, etc. If a position is followed by the symbol −, then the mutation at that position is a nucleotide deletion. If a letter (one of A, C, G, or T) follows the position, then the mutation is a polymorphism. Insertions are represented in the format xxx.yz, meaning that nucleotide z (one of A, C, G, or T) appears in the yth position after nucleotide 16xxx.

Whenever two nodes represent haplotypes that differ by a single mutation, they have been joined by an edge that is labelled with the position of the mutation (again, dropping the leading 16).

When the overall network was initially mapped, it was found that 32 of the 121 nodes were isolated and not adjacent to any other nodes, meaning that they represented haplotypes that were at least two mutational differences away from any other haplotype in the data set. For these isolated nodes, we checked to see if there were any nodes whose haplotypes were two mutational differences away; whenever this was found to be true we added a dotted edge between the pair of nodes. In this manner, 28 of the 32 originally isolated nodes were able to be placed into a context in the network diagram, in the sense that the dotted edges join nodes that would be close to each other if the intermediate haplotypes had not been absent from our sample data set. We similarly added a dotted edge between the nodes representing the haplotypes 16126–16182–16183–16189–16294–16296–16298–16519 and 16126–16183–16189–16294–16296–16519, so that the cluster centred at the former node could be suitably placed in the network diagram.

Each node is individually coloured to indicate the subgroup to which the corresponding haplotype belongs. For the purpose of determining which colour to use for a given node, we adopted a few simple conventions. In this regard, we first considered whether any of the single mutations that are associated with subgroups T2 through T5 were present, and if so, assigned the haplotype to the corresponding subgroup. However, if the sample met the criteria for more than one of these subgroups, then we arbitrarily gave precedence to the subgroup with the higher numbered diagnostic mutation (so, in essence, T4's 16324 would override T2's 16304, which would override T3's 16292, which would override T5's 16153). In practice, only four samples exhibited such ambiguity: the two already mentioned (and which we assigned to subgroup T2), one with the sequence 16126–16292–16294–16296–16304–16519 that was also assigned to subgroup T2, and one other with the sequence 16126–16292–16294–16296–16324–16519 that we

assigned to subgroup T4.

We also counted the number of mutations that each sample shared with the T1 diagnostic motif 16163–16186–16189. Whenever a haplotype had mutations in at least two of these three positions, we classified it as T1 and coloured its node accordingly. None of the samples with at least two of these three mutations also had mutations associated with subgroups T2 through T5, so we did not have to deal with potentially ambiguous situations with respect to T1.

Each sample not classified by these guidelines as belonging to one of the subgroups T1 through T5 was left unassigned and appears in the "Other" category in the network diagrams.

Analysis and Discussion

The overall phylogenetic network, based on the 361 samples and the corresponding 121 haplotypes is shown in Figure 1. Each node and solid edge has been labelled, but even with a small font the labels may detract from the overall presentation. Hence in Figure 2 we present the corresponding unlabelled network diagram. The node representing the haplotype 16126-16294-16519 has been drawn with a double circle to emphasise that it is the point at which the network connects to the greater human mtDNA phylogenetic network.

There are several obvious features of the phylogenetic network for the haplogroup. One of the most apparent is that subgroups T1 and T2 dominate the haplogroup (or, perhaps more accurately, they dominate the MitoSearch data set). Subgroups T1 and T4 take the form of star-like clusters, whereas T2, T3, and T5 do not exhibit the pattern of a dominant central node with lesser nodes radiating out from it.

As previously noted, the network contains four isolated nodes, indicating that their haplotypes are at least three mutational differences away from any other sample in our data set. These four nodes appear in Figure 1 with one in each of the subgroups T1, T2, T3, and T5. Those in T1 and T2 appear as if they may have suffered from data entry errors. Specifically, the isolated T1 node has a 16163C mutation, yet all of the remaining T1 samples with a mutation at 16163 have the mutation 16163G. The situation with the T2 sample is even more apparent, as it consists of the mutations 16126T, 16241A, 16294C, 16304T, and 16519T, none of which represents a difference when compared to the CRS; likely 16126C, 16294T, 16304C, 16519C and something other than an A at 16241 were intended. The two other isolated nodes, in subgroups T3 and T5, show no obvious signs of

inconsistency and may simply represent haplotypes for which genetic neighbours were not found in the MitoSearch database, which may be a reflection of the relatively small number of samples that belong to the T3 and T5 subgroups.

Another striking aspect of the network is that it is not a tree, but instead it contains many cycles (i.e., reticulations). A closer inspection of the edges in the network reveals that several of these cycles contain edges that correspond to mutations at nucleotide 16296. Moreover, mutations at 16296 account for 13 of the 103 solid edges in the network, more than double that of any other nucleotide.

There is widespread distribution of mutations at 16296. Of the 361 samples in our data set, 183 of them possess a mutation at 16296. We also find that each of the five subgroups contains some haplotypes that have mutations at 16296 as well as some that do not. There is a similar combination among the ungrouped haplotypes.

The situation with nucleotide 16296 appears to warrant some explanation. Since mutations at 16296 are present in each subgroup, it would almost appear that it should be included with 16126 and 16294 as a defining mutation for the T haplogroup. However, its widespread absence would contradict such an inclusion. A lack of consistency with respect to 16296 was previously reported by Richards et al. (2000), who noted that Malyarchuk and Derenko (1999) had suggested that the mutation at 16294 might be an influential factor for instability at 16296 (and, to a lesser extent, also at other nearby nucleotide positions).

Another hypothesis, which we now discuss, is that a situation of heteroplasmy at position 16296 might have developed shortly after the occurrence of the mutation at 16294 that partially defines the T haplogroup. If this heteroplasmy persisted for many generations, beyond the emergence of the mutations that define the T1 through T5 subgroups, then in some lineages the heteroplasmy may have transitioned to a state of homoplasmy with a cytosine nucleotide at 16296 (which matches the CRS and is therefore not reported as a mutation) while others may have fixed to a thymine (which we now detect as a mutation at 16296).

It is interesting to note that a case for persistent heteroplasmy in human mtDNA has been previously reported. In particular, persistent heteroplasmy at nucleotide location 16192 among some members of mtDNA haplogroup U has been detected, and is suspected to be related to a polymorphism at 16189 (Howell and Smejkal 2000). It is therefore conceivable that a similar mechanism is helping the polymorphism at 16294 (which is harboured by members of
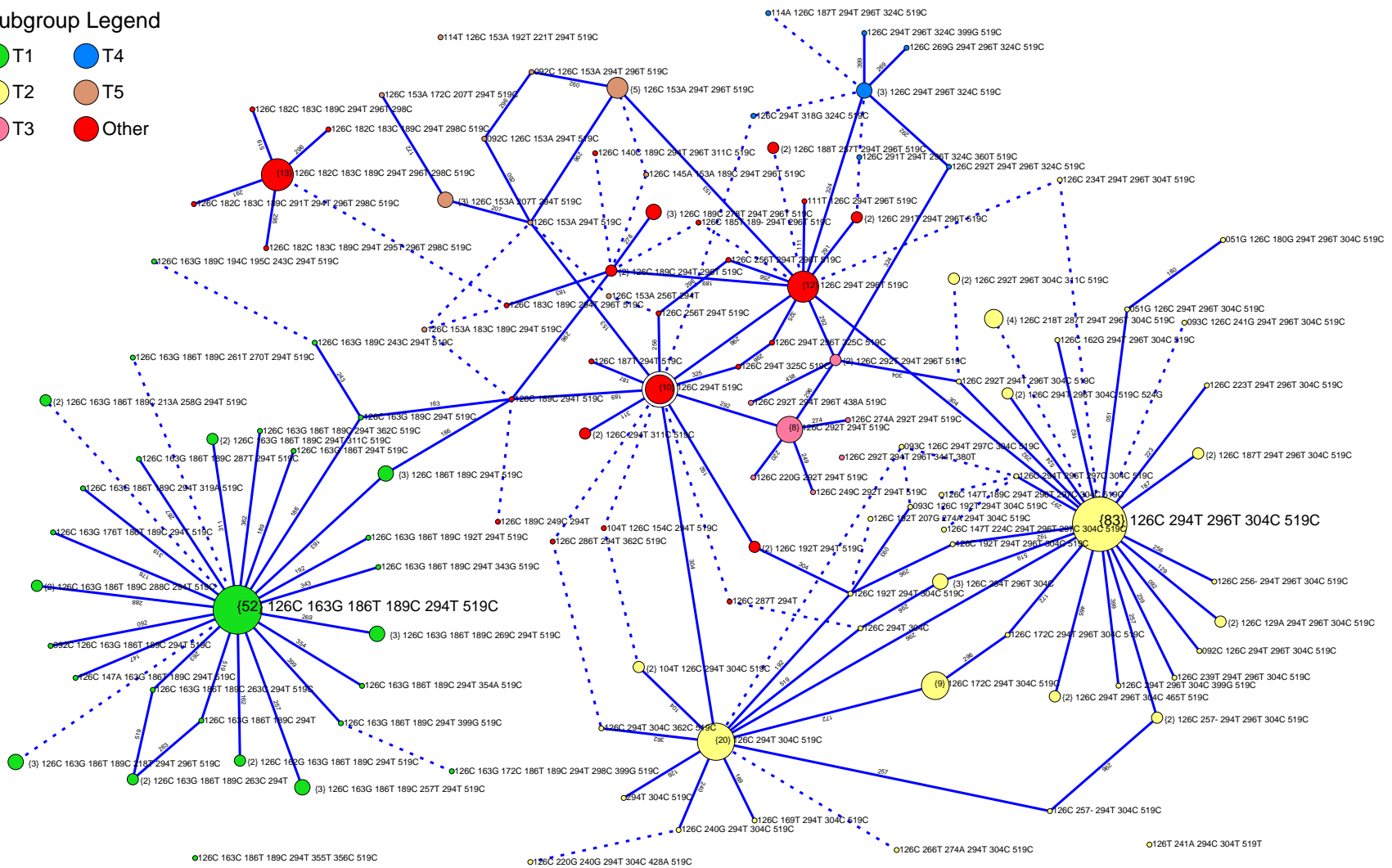
Figure 1: The phylogenetic network

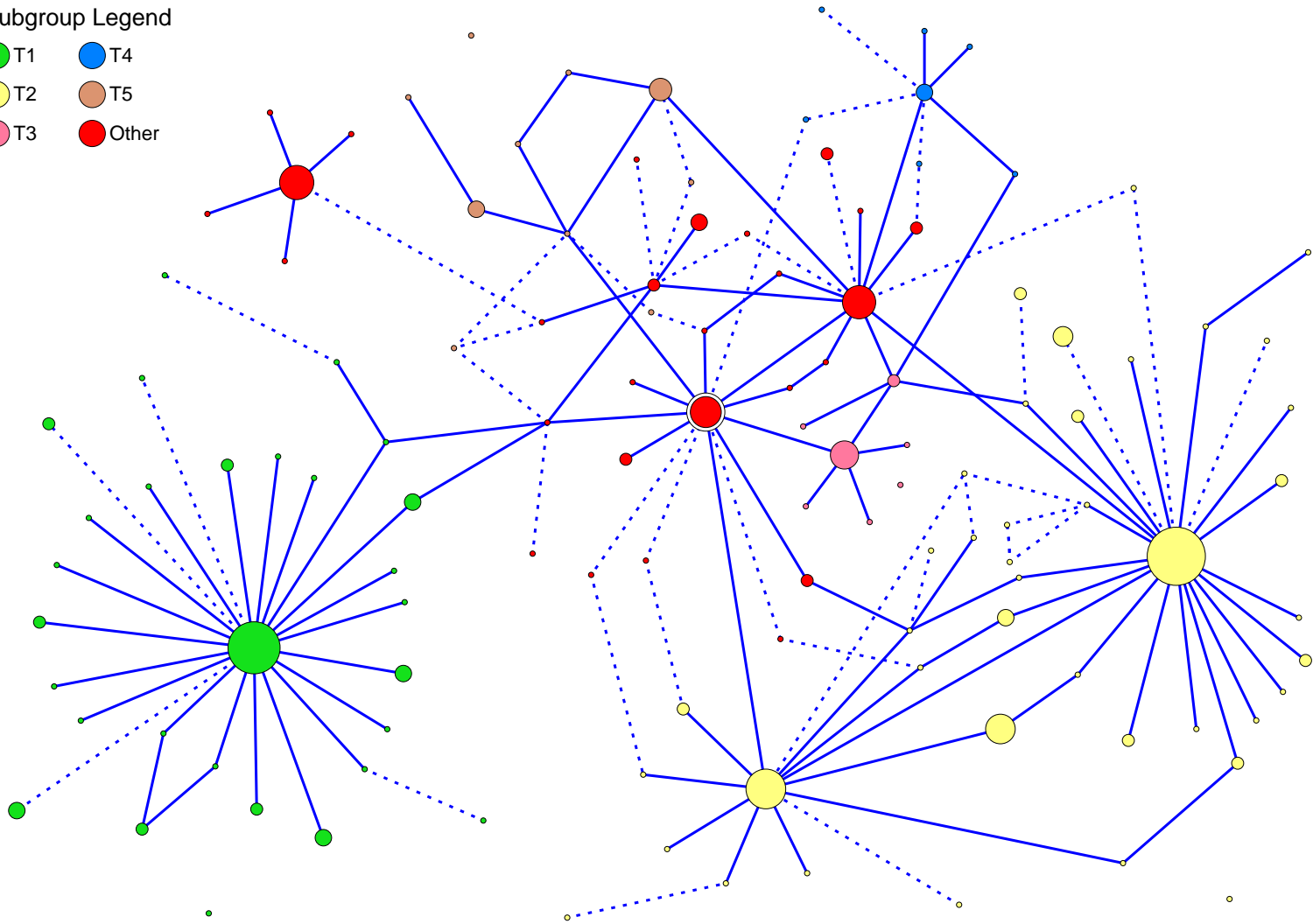Figure 2: The phylogenetic network (unlabelled)

haplogroup T) to produce or sustain heteroplasmy at 16296.

The MitoSearch database presently contains no information about the presence or absence of heteroplasmy and therefore offers no further insight into this question. It would therefore be interesting to see the results of a study that specifically seeks to determine if cases of heteroplasmy at location 16296 occur at an unexpectedly high frequency among the modern-day members of haplogroup T, possibly indicating that 16294 is influencing 16296 or perhaps indicating that a case of ancient heteroplasmy still lingers today.

If we assume that there is inherent instability and/or persistent heteroplasmy at nucleotide 16296 when dealing with haplogroup T, then perhaps the phylogenetic network illustrated in Figure 1 and Figure 2 warrants some simplification. Specifically, we could disregard the presence of mutations at 16296 in our data set (which is an action that we note was also taken by Richards et al. (2000)), and then repeat the process of generating the phylogenetic network. In so doing, we find that the 361 samples now give rise to 108 distinct haplotypes, the most common three of which occur 103, 52, and 22 times (collectively accounting for 49.0% of the data set). The resulting phylogenetic network is shown with labels in Figure 3 and without labels in Figure 4.

For the most part, the revised network has good connectivity and very few cycles, and the subgroups now exhibit a stronger star-like cluster pattern. That the subgroups radiate out from the root node 16126–16294–16519 is to be expected, since this root node represents the ancestral origin of the haplogroup.

We now turn our attention to a conspicuous collection of seventeen samples that form their own star-like cluster, centred at haplotype 16126–16182–16183–16189–16294–16296–16298–16519. This central haplotype is two mutational differences away from a node representing a single instance of the haplotype 16126–16183–16189–16294–16296–16519. Were it not for the dotted edge to the latter node, this small cluster would have appeared orphaned and not joined to the rest of the network. This group is large enough and demonstrates a sufficient star-like pattern to warrant naming. However, before we name this group we give some thought to the evolution of the T1 subgroup.

The path from the T root node to the centre of the T1 cluster begins with an edge that represents a mutation at 16189. Thus it appears that 16189 was the first of the three founding HVR1 mutations to occur in the historical rise of the T1 subgroup. This hypothesis is further supported by the presence of

16189 in a number of haplotypes (including those in our nearly orphaned group) which do not also share one of the other two mutations from the 16163–16186–16189 T1 motif introduced by Richards et al. (1998).

Kivisild et al. (2004) and Palanichamy et al. (2004) have suggested some revision to the T1 subgroup, so that the T1 designation would correspond to the HVR1 motif 16163–16189. Those haplotypes that also contain a polymorphism at 16186 would be designated as belonging to a subgroup named T1a, while those harbouring a mutation at 16243 would be designated as belonging to a subgroup named T1b. Given our hypothesis that 16189 was the first of the T1 mutations to occur, and the accompanying observation of genetic branching subsequent to the advent of the 16189 mutation but prior to the development of either of the 16163 or 16186 mutations, we propose that the T1 hierarchy be slightly further refined. In particular we recommend that the T1 designation should apply to those haplotypes which have the 16189 polymorphism. The supplementary HVR1 motifs for T1a and T1b would therefore be revised to be 16163–16186 and 16163–16243, respectively. The cumulative motifs of 16163–16186–16189 and 16163–16189–16243 for T1a and T1b, respectively, therefore remain unchanged. Incidentally, most of the T1 nodes in our network diagrams belong to the T1a subgroup, whereas only two samples from the MitoSearch data set belong to T1b.

Returning now to our nearly orphaned cluster, we note that each of its seventeen samples has a mutation at 16189, which places the cluster within the revised T1 subgroup. Each haplotype in this cluster also has mutations at each of 16182, 16183, and 16298. Thus we now introduce the T1c designation, along with the supplementary motif 16182–16183–16298.

The new motif specifications for the T1 subgroup are shown in Table 3 and yield a subgroup hierarchy that is consistent with the nomenclature standard outlined by Richards et al. (1998).

Subgroup	Associated HVR1 Mutations
T1	16189
T1a	16163–16186–16189
T1b	16163–16189–16243
T1c	16182–16183–16189–16298

Table 3: Revised HVR1 Motifs for Subgroup T1

Acknowledgements
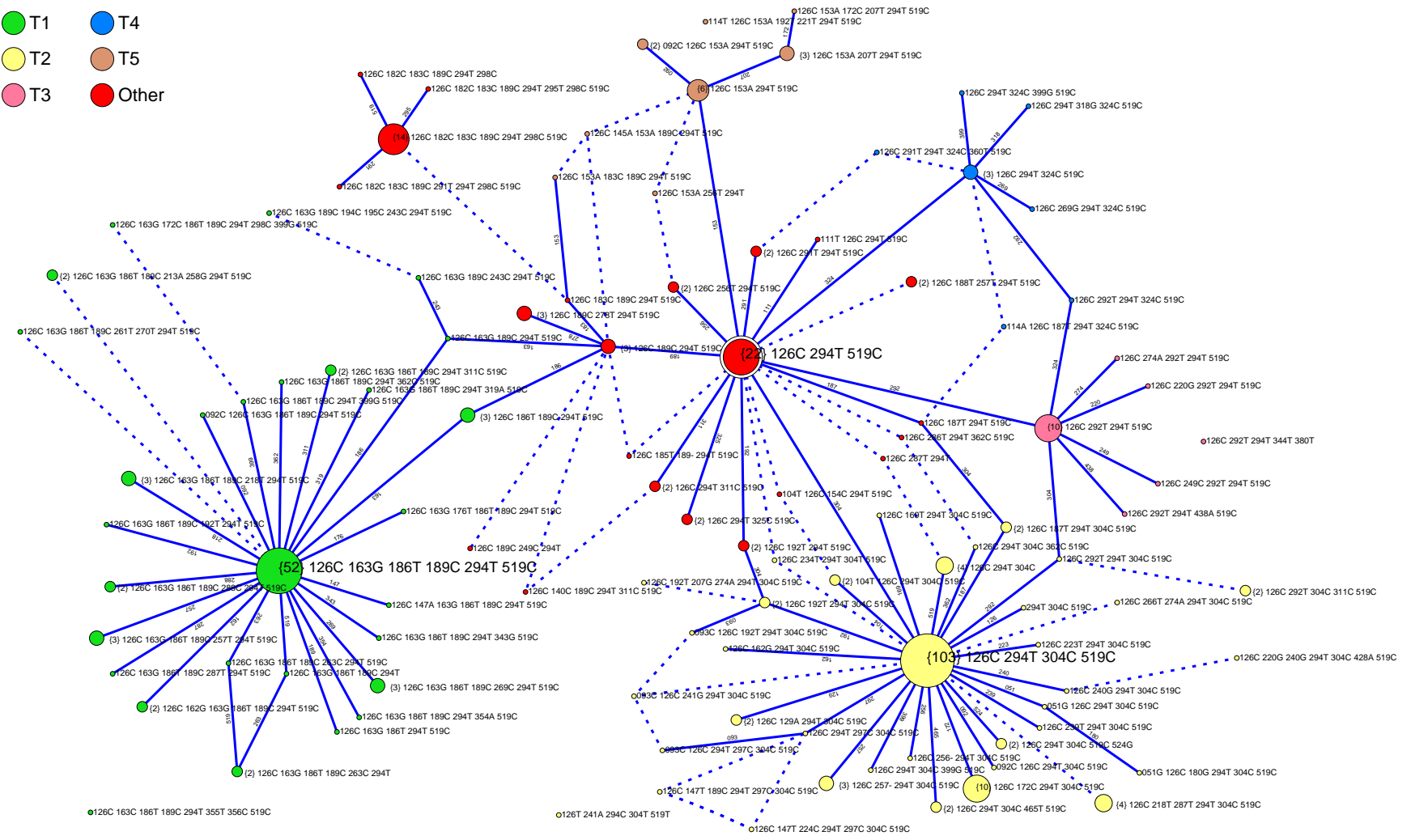
Figure 3: The phylogenetic network, disregarding 16296
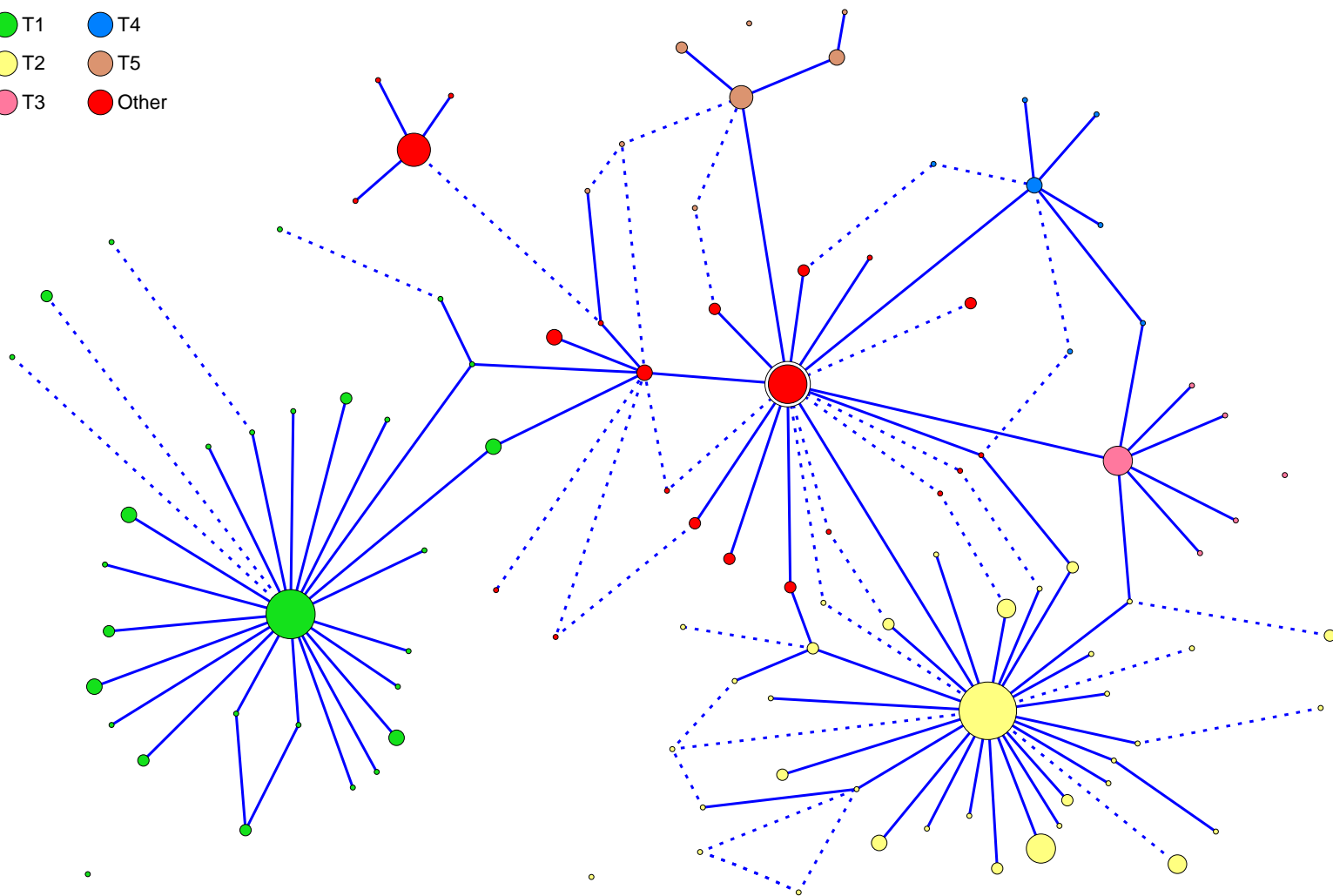
Figure 4: The phylogenetic network (unlabelled), disregarding 16296

Source Data

The raw data extracted from the MitoSearch database and used in the construction of the networks presented in this paper are available online at www.jogg.info/21/T-data.txt.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–465.

Batagelj V, Mrvar A. Pajek – Program for Large Network Analysis. Home page: http://vlado.fmf.uni-lj.si/pub/networks/pajek.

Chagnon P, Gee M, Filion M, Robitaille Y, Belouchi M, Gauvreau D (1999) Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population. Am. J. Med. Genet. 85:20–30.

Finnilä S, Majamaa K (2001) Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population. J. Hum. Genet. 46:64–69.

Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21:33–49.

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B (2001) mtDNA and the Islands of the North Atlantic: Estimating the Proportions of Norse and Gaelic Ancestry. Am. J. Hum. Genet. 68:723–737.

Helgason A, Sigurðardóttir S, Gulcher JR, Ward R, Stefánsson K (2000) mtDNA and the Origin 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, Olefsky JM, Beal MF, Davis RE, Howell N (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, and Smejkal CB (2000) Persistent Heteroplasmy of a Mutation in the Human mtDNA Control Region: Hypermutation as an Apparent Consequence of Simple-Repeat Expansion/Contraction. Am. J. Hum. Genet. 66:1589–1598.

Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat. Genet. 12:417–420.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, Villems R (2004) Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears. Am. J. Hum. Genet. 75:752–770.

Kocher TD, Wilson AC (1991) Sequence Evolution of Mitochondrial DNA in Humans and Chimpanzees: Control Region and a Protein-Coding Region. In S. Osawa and T. Honjo (Eds.), Evolution of Life: Fossils, Molecules, and Culture, pp. 391–413. Springer-Verlag, Tokyo.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A (1999) The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs. Am. J. Hum. Genet. 64:232–249.

Malyarchuk BA, Derenko MV (1999) Molecular instability of the mitochondrial haplogroup T sequences at nucleotide positions 16292 and 16296. Ann. Hum. Genet. 63:489–497.

Palanichamy M, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, Khan F, Wang C-Y, Chaudhuri TK, Palla V, Zhang Y-P (2004) Phylogeny of Mitochondrial DNA Macrohaplogroup N in India, Based on Complete Sequencing: Implications for the Peopling of South Asia. Am. J. Hum. Genet. 75: 966–978.

Pereira L, Gonçalves J, Goios A, Rocha T, Amorim A (2005) Human mtDNA haplogroups and reduced male fertility: real association or hidden population substructuring. Int. J. Androl. 28:241–247.

Pyle A, Foltynie T, Tiangyou W, Lambert C, Keers SM, Allcock LM, Davison J, Lewis SJ, Perry RH, Barker R, Burn DJ, Chinnery PF (2005) Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann. Neurol. 57:564–567.

Richards M, Côrte-Real M, Forster P, Macauley V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (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, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Rienzo AD, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt H-J (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. 62:241–260.

Ruiz-Pesini E, Lapeña A-C, Díez-Sánchez C, Pérez-Martos A, Montoya J, Alvarez E, Díaz M, Urriés A, Montoro L, López-Pérez MJ, Enríquez JA (2000) Human mtDNA Haplogroups Associated with High or Reduced Spermatozoa Motility. Am. J. Hum. Genet. 67:682–696.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC (1996) Classification of European mtDNAs From an Analysis of Three European Populations. Genetics 144:1835–1850.

Torroni A, Lott MT, Cabell MF, Chen Y-S, 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. 55:760–776.

Wilkinson-Herbots HM, Richards MB, Forster P, and Sykes BC (1996) Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. Ann. Hum. Genet. 60:499–508.