*Ann. Hum. Genet.* (1999), **63**, 489–497
*Printed in Great Britain*



# Molecular instability of the mitochondrial haplogroup T sequences at nucleotide positions 16292 and 16296

B. A. MALYARCHUK AND M. V. DERENKO
*Institute of Biological Problems of the North, 685000 Magadan, Russia*

(Received 4.8.99. Accepted 14.10.99)

SUMMARY

The mitochondrial haplogroup T, characterized by the nucleotide motif 16126C–16294T in the hypervariable segment I (HVS I), is one of the most frequent among Europeans. It has been shown that this haplogroup includes the only well-resolved subgroup, T1, but that other HVS I sequences cannot be differentiated into subgroups due to possible homoplasies at nucleotide positions 16292, 16296 and 16304, leading to the reticulations in the topology of phylogenetic networks. To study the problem of molecular instability at these positions, we have performed an analysis of 159 previously published West Eurasian HVS I sequences belonging to haplogroup T, together with 12 new HVS I sequences of Eastern Slavs. These 12 sequences represent 16.9% of a total of 71 samples analysed and identified as haplogroup T mtDNAs by RFLP analysis in this study. A search for rare point mutations associated with different combinations of nucleotides 16292T, 16296T and 16304C within the haplogroup T sequences, and specific to certain populations or a group of closely related-by-descent populations, was performed. This analysis revealed 11 marker mutations, each of which was characteristic for a certain group of linguistically or geographically close individuals – the Adygei, Germans, Kazakhs and linguistic isolates of the Eastern Italian Alps. The occurrence of these rare population-specific polymorphisms in association with various combinations of mutations at positions 16292 and 16296 on the haplogroup T background provides evidence of molecular instability at these nucleotide positions. Molecular instability in the haplogroup T HVS I sequences is also suggested by multiple independent losses of the haplogroup T diagnostic nucleotide variants in different populations. The results of the present study suggest that identical haplogroup T HVS I sequence types might have arisen independently in different human populations.

## INTRODUCTION

The maternal mode of inheritance of the human mitochondrial genome and the high rate of base substitutions in this genome allow the use of mitochondrial DNA (mtDNA) polymorphisms for inferring the pattern of prehistoric female migrations and peopling of different regions of the world. Human mtDNA accumulates mutations at a rate 10–16 times higher than the nuclear genome (Larsson & Clayton, 1995), but mitochondria are relatively deficient in DNA-repair mechanisms (Lightowlers *et al.* 1997; Bogenhagen, 1999). It has been suggested that human cells contain a high copy number of mtDNA molecules – 1000–100000 copies of mtDNA in each cell (Larsson & Clayton, 1995) – in order to protect them against the accumulation of mutant alleles during their lifetime (Chinnery & Samuels, 1999). Despite the high copy number, recombination of mtDNA

Correspondence: Dr Boris A. Malyarchuk, Genetics Laboratory, Institute of Biological Problems of the North, Portovaya str., 18, 685000 Magadan, Russia. Tel/Fax: 7 41322 34463.
E-mail: rosa@online.magadan.su

molecules is rare or absent (Howell *et al.* 1996; Eyre-Walker *et al.* 1999; Hagelberg *et al.* 1999).

By the use of direct sequencing of the fast-evolving control region hypervariable segment I (HVS I) and analysis of the entire molecule of mtDNA by means of a standard set of restriction enzymes in human populations, it was shown that maternal genealogy has marked continent-specific features (Wallace, 1995). For instance, virtually all West European mtDNAs belong to eight haplogroups of mitochondrial sequence types (H, V, U, J, T, I, W, X), which are determined by certain RFLP and HVS I nucleotide motifs of cluster-diagnostic mutations (Richards *et al.* 1996; Torroni *et al.* 1996; Macaulay *et al.* 1999). However, one of the problems of identification of the HVS I sequence types is homoplasy at sites in the mtDNA control region. For example, HVS I motif 16189C–16356C is not sufficient to dissect cluster H from U4; at least, information on nucleotide position (np) 00073 in HVS II is required (00073A – for H, 00073G – for U4). Therefore, the high rate of base substitutions in the mtDNA control region often make it difficult to resolve recurrent mutations affecting the same nucleotide positions.

The unequal mutation rate at different nucleotide positions in the mtDNA control region was demonstrated by Hasegawa *et al.* (1993) and Wakeley (1993). Recently, Richards *et al.* (1998) suggested that the list of nucleotide positions in the HVS I can be divided into three weight classes with fast, intermediate and slow transition rates, for the purposes of phylogenetic analysis. The mechanism of such differences between rates of base substitutions is still unclear, but the influence of nucleotide context on the mutagenesis intensity seems to be obvious. Length polymorphism in the polycytosine tract between nps 16184 and 16193 is one of the most well-known cases of molecular instability in HVS I, and is caused by the thymine-cytosine transition at np 16189 (Bendall & Sykes, 1995). As a consequence of this type of instability, many phylogenetically unrelated sequences of HVS I may have identical types of poly-C tract. The possibility of parallel and reverse mutations occurring within and between haplogroups of HVS I sequences is a real problem in the classification of mtDNAs.

Haplogroup T with the 16126C–16294T HVS I motif is one of the most widespread sequence types among Europeans (Richards *et al.* 1996). The molecular structure of the haplogroup T mtDNAs in various human populations of West Eurasia has been well studied by means of combined RFLP of coding regions and HVS I sequencing (Malyarchuk, 1995, 1997; Richards *et al.* 1996, 1998; Torroni *et al.* 1996). It has also been shown that haplogroup T includes the only well-resolved subgroup, T1 (Malyarchuk, 1997; Richards *et al.* 1998). Other HVS I sequence types in haplogroup T form a set of sequences which cannot be differentiated in subgroups because the branching order of these sequences cannot be resolved. This ambiguity in the phylogenetic relationships between non-T1 HVS I sequences is caused by three mutations at nps 16292, 16296, 16304, causing reticulations in the topology of phylogenetic networks (Fig. 8 in Richards *et al.* 1998). The problem of HVS I sequence identification within haplogroup T is probably due to parallelism at nps 16292, 16296 and 16304. To study the problem of molecular instability of haplogroup T sequences we present data on the diversity of haplogroup T mtDNAs in the Eastern Slavonic population, and the results of analysis of published data sets of haplogroup T HVS I sequences in the various populations of West Eurasia.

## subjects and methods

### *Subjects and mtDNA analyses*

DNA samples from 53 Russians, 13 Ukrainians and 5 Belorussians were obtained from maternally-unrelated individuals. To determine the haplogroup T mtDNAs among these 71 individuals, we typed a set of RFLPs that were diagnostic of haplogroup T as a whole ($+13366$*Bam*HI/$-13367$*Ava*II, $+15606$*Alu*I, $-15925$*Msp*I) and its subgroup T1

(−12629*Ava*II, additionally). RFLP typing was performed by restriction endonuclease analysis of PCR amplified mtDNA fragments using the same primer pairs and amplification conditions as described by Torroni *et al.* (1994, 1996). Then, we sequenced HVS I between nps 16000 and 16400 for samples assigned to the haplogroup T. Nucleotide sequences were determined by use of the Sanger dideoxy chain-termination method and Sequenase enzyme with amplification primers L15926 and H16498 (Di Rienzo & Wilson, 1991) as sequencing primers.

HVS I sequences sharing the motif 16126C–16163G–16186T–16189C–16294T were defined as subgroup T1 sequences; the other HVS I sequences with the haplogroup T-diagnostic motif 16126C-16294T were identified as belonging to the subgroup T* (Richards *et al.* 1998).

*Comparative analysis of mtDNA data*

We compared mtDNA HVS I sequences of a total of 2093 samples from West Eurasia, North Africa and Central Asia retrieved from the published data: 42 Middle Easterners (Di Rienzo & Wilson, 1991); 96 Turks (Calafell *et al.* 1996; Comas *et al.* 1996; Richards *et al.* 1996); 45 Israeli Druze and 50 Adygei (Macaulay *et al.* 1999); 85 Algerians and 54 Portuguese (Corte-Real *et al.* 1996); 18 Berbers and 56 Canary Islanders (Pinto *et al.* 1996); 89 Spanish (Corte-Real *et al.* 1996; Pinto *et al.* 1996); 105 Basque (Bertranpetit *et al.* 1995; Corte-Real *et al.* 1996); 30 Bulgarians (Calafell *et al.* 1996); 69 Sardinians (Di Rienzo & Wilson, 1991); 49 Tuscans (Francalacci *et al.* 1996); 29 Finns, 33 Danes, 107 Northern Germans, 49 Bavarians (Richards *et al.* 1996); 74 Swiss (Pult *et al.* 1994); 101 Austrians (Parson *et al.* 1998); 200 Southern Germans (Lutz *et al.* 1998); 34 Volga Finnic, 47 Estonians, 83 Karelians, 50 Finns, 32 Swedes (Sajantila *et al.* 1995); 261 British (Piercy *et al.* 1993; Richards *et al.* 1996); and 205 subjects from central Asian admixed populations of East Asian and European ancestry: 55 Kazakh, 95 Kirghiz, 55 Uighur (Comas *et al.* 1998). After haplogroup T HVS I sequences from these data were combined, 159 individual mtDNA sequences and 62 distinct sequence types were obtained. In addition, HVS I sequence data on 70 individuals from the Eastern Italian Alps (Stenico *et al.* 1996), 407 Africans (Watson *et al.* 1997) and 538 Asians (Shields *et al.* 1993; Horai *et al.* 1996; Kolman *et al.* 1996; Derenko & Shields, 1997) were incorporated into the analysis.

## RESULTS AND DISCUSSION

The mtDNA restriction endonuclease analysis of the Russian, Ukrainian and Belorussian samples revealed that twelve individuals, representing 16.9% of the total samples analysed, fell into the mtDNA haplogroup designated T (Table 1). This group was defined by a stable haplogroup T-specific motif of mutations (+13366*Bam*HI/−13367*Ava*II, +15606*Alu*I, −15925*Msp*I) located in the coding regions of the mtDNA (*MTND5*, *MTCYB* and $tRNA^{Thr}$ respectively). Nine sequence types of HVS I, defined by 13 polymorphic nucleotide sites, were observed among the 12 samples analysed. Only two sequence types were observed in more than one subject (ES3 and ES8), and these were also shared by the Ukrainians and Belorussians (ES3) and the Russians and Ukrainians (ES8). The subgroup T1 was represented by two mtDNA HVS I sequence types (ES8 and ES9) and was defined by the combination of nucleotides 16163G–16186T–16189C in HVS I and −12629*Ava*II site in the *MTND5* gene.

The majority of the Eastern Slavonic HVS I sequence types, belonging to haplogroup T, have been described previously in West Eurasian populations. Sequence types ES8, ES3, ES2 and ES1 appear to be the four most frequent haplotypes in West Eurasians, and comprise 22.0, 13.8, 8.2 and 5.0%, respectively, of 159 haplogroup T HVS I sequences analysed. Sequence types ES5, ES6 and ES9 are unique to the Russians. The remaining sequence types, ES4 and ES7, were found in other Caucasian populations at low frequencies (less than 2.0%).

Phylogenetic analysis of the combined RFLP haplotypes and HVS I sequences, performed by

Table 1. *RFLP haplotypes and HVS I sequence variation of haplogroup T mtDNAs of Eastern Slavonic samples*

| Subgroup/ sequence type | RFLP haplotype[a] | HVS I haplotype[b] | No. of mtDNAs, by population[c] | | |
|---|---|---|---|---|---|
| | | | RUS | UKR | BEL |
| T* | | | | | |
| ES1 | **+13366m/−13367b, +15606a, −15925i** | **126 294** | 1 | . | . |
| ES2 | **+13366m/−13367b, +15606a, −15925i** | **126 294** 296 | 1 | . | . |
| ES3 | **+13366m/−13367b, +15606a, −15925i** | **126 294** 296 304 | . | 2 | 1 |
| ES4 | **+13366m/−13367b, +15606a, −15925i** | 093 **126 294** 296 304 | 1 | . | . |
| ES5 | **+13366m/−13367b, +15606a, −15925i** | 111 **126 294** 296 304 311 325 | 1 | . | . |
| ES6 | **+13366m/−13367b, +15606a, −15925i** | **126** 153 **294** 362 | 1 | . | . |
| ES7 | **+13366m/−13367b, +15606a, −15925i** | **126** 189 292 **294** | 1 | . | . |
| T1: | | | | | |
| ES8 | **−12629b, +13366m/ −13367b, +15606a, −15925i** | **126 163 186 189 294** | 1 | 1 | . |
| ES9 | **−12629b, +13366m/ −13367b, +15606a, −15925i** | **126 163 186 189** 271 **294** | 1 | . | . |
| Total | | | 8 | 3 | 1 |

States diagnostic of haplotype clusters are shown in bold.

[a] Restriction sites are numbered from the first nucleotide of the recognition sequence. A plus sign (+) indicates the presence of a restriction site, and a minus sign (−) indicates the absence of a restriction site. The restriction enzymes used in the analysis are designated by the following single-letter codes, as described by Torroni *et al.* (1996): a = *Alu*I, b = *Ava*II, i = *Msp*I, m = *Bam*HI. A slash (/) separating states indicates the simultaneous presence or absence of restriction sites that can be correlated with a single-nucleotide substitution.

[b] nps (−16,000) between 16000 and 16400 that are different from the Cambridge reference sequence (CRS; Anderson *et al.* 1981).

[c] RUS, Russians, UKR, Ukrainians, BEL, Belorussians.

means of the network method of Bandelt *et al.* (1995) reveals a high simplicity of the network, with two well defined clusters, T1 and T*, and with no uncertainty in the branching order of the HVS I sequence types. However, the phylogeny of haplogroup T sequences becomes complicated when additional information derived from the database of mtDNA HVS I in East European populations is adding. The inclusion in the analysis of 23 haplogroup T sequences from East European populations, such as Estonians, Karelians, Volga Finnic and Adygei, introduced considerable ambiguity in the phylogeny, manifested by reticulations in a network (not shown). This ambiguity is connected mostly with positions 16292, 16296 and 16304, which determine the nucleotide combinations in the most frequent haplogroup T sequence type and its derivates. Because of the high substitution rate of the HVS I control region, it is unclear whether nucleotides 161292, 16296 and 16304 determine a subset of haplogroup T sequences (i.e. define its subgroups) or identical combinations of these nucleotides in different sequence types arose independently because of instability at these nucleotide positions. To investigate this question, a search for rare point mutations associated with different combinations of nucleotides 16292T, 16296T, 16304C within the haplogroup T HVS I sequences, and specific to certain populations or groups of closely related populations, was performed. This approach assumes that rare mutations associated with the abovementioned nucleotide variants, and the haplo-

Table 2. *Analysis of molecular instability in haplogroup T HVS I sequence types in several populations*

| Sample origin | Marker mutation | HVS I sequence type | Haplogroup T subgroup | Unstable nucleotide position |
|---|---|---|---|---|
| Adygei[a] | **078T** | AD08 – **78** 126 <u>292</u> 294 296 | T* | <u>292</u> |
| | | AD20 – **78** 126 294 296 | T* | |
| Kazakh[b] | **093C** | KAZ98 – **93** 126 163 186 189 294 | T1 | <u>296</u> |
| | | KAZ99 – **93** 126 163 186 189 294 <u>296</u> | T1 | |
| Germans[c] | **298C** | GER109 – 126 163 186 189 **298** | T1 | <u>294</u> <u>296</u> |
| | | GER29 – 126 189 <u>294</u> <u>296</u> **298** | T* | |
| | | GER13 – 126 163 172 186 189 <u>294</u> **298** 399 | T1 | |
| Ladins[d] | **079T** | LS2 – **79** 126 294 296 | T* | <u>292</u> |
| | | LS3 –**79** 126 145 <u>292</u> 294 296 | T* | |
| | | LS8 – 69 **79** 85 126 244 294 296 | T* | |
| Ladins[d] | **085T** | LS8 – 69 79 **85** 126 244 294 <u>296</u> | T* | <u>296</u> |
| | | LC13 – **85** 106 126 163 186 189 294 | T1 | |
| Ladins[d] | **106A** | LC13 – 85 **106** 126 163 186 189 294 | T1 | <u>296</u> |
| | | LS5 – **106** 126 145 184 280 294 <u>296</u> 344 | T* | |
| Ladins[d] | **145A** | LS5 – 106 126 **145** 184 280 294 296 344 | T* | <u>292</u> |
| | | LS10 – 126 **145** <u>292</u> 294 296 | T* | |
| | | LS3 – 79 126 **145** <u>292</u> 294 296 | T* | |
| Ladins/Germans[d] | **221T** | LC14 – 126 163 186 189 **221** 294 | T1 | <u>296</u> |
| | | GJ20 – 126 184 **221** 280 294 <u>296</u> 308 344 | T* | |
| Ladins/Germans[d] | **280G** | GJ20 – 126 184 221 **280** 294 <u>296</u> 308 344 | T* | <u>296</u> |
| | | LS5 – 106 126 145 184 **280** 294 <u>296</u> 344 | T* | |
| | | GJ22 – 126 184 **280** 294 <u>296</u> 308 344 | T* | |
| | | GJ27 – 126 **280** 294 343 | T* | |
| Ladins[d] | **299G** | LC15 – 126 186 189 294 <u>296</u> **299** | T1 | <u>296</u> |
| | | LC16 – 126 163 186 189 246 294 **299** | T1 | |
| Ladins/Germans[d] | **343G** | GJ27 – 126 280 294 **343** | T* | <u>296</u> |
| | | LS6 – 126 294 <u>296</u> **343** | T* | |

Marker (population-specific) mutations are shown in bold, unstable nucleotide positions are underlined. HVS I sequences from the following studies were analysed: [a] Macaulay *et al*. 1999, [b] Comas *et al*. 1998, [c] Richards *et al*. 1996 (GER109) and Lutz *et al*. 1998 (GER13 and GER29), [d] Stenico *et al*. 1996.

group T-diagnostic polymorphisms, are not a result of site-directed population-specific mutagenesis, and therefore such marker mutations could be useful in the search for unstable nucleotide positions within haplogroup T sequences.

The analysis of 192 haplogroup T HVS I sequences, comprising 159 published West Eurasian sequences, 12 Eastern Slavonic mtDNAs presented here, and 21 HVS I sequences characterized by T-diagnostic motif 16126C-16294T and retrieved from the study of isolated localities in the Eastern Italian Alps (Stenico *et al*. 1996), reveals 11 marker mutations, each of which was characteristic for a certain population or group of geographically close individuals, and showed local patterning of geographic distribution (Table 2).

The variant 16078T, associated with subgroup T* HVS I sequences, has been detected in the Adygei population from the North Caucasus (Macaulay *et al*. 1999). This variant appears to be stable, since it has never been seen in other published HVS I sequences on the haplogroup T background including the 16126C and 16294T mutations. The comparison of HVS I sequences AD08 and AD20 showed the only difference between them is at position 16292.

The next sequences, KAZ98 and KAZ99, which are found in the Kazakh population from central Asia (Comas *et al*. 1998), belong to subgroup T1. These HVS I sequences are characterized by the common marker variant 16093C and by nucleotide instability at position 16296. Although the variant 16093C has occurred multiple times in different mtDNA haplogroups of Eurasians, in the case of the present study it can be used as a marker variant for revealing

recurrent mutations within subgroup T1 sequences, since among 56 West Eurasian T1 sequences the variant 16093C has been found only in the Kazakh sequence types.

The variant 16298C has been observed in several German-speaking populations – in the Germans from Northern and Southern Germany (Richards *et al*. 1996; Lutz *et al*. 1998) and in the Austrians (Parson *et al*. 1998). This mutation has arisen multiple times in human mtDNA, being found in Caucasian haplogroup V (Torroni *et al*. 1998) and Asian haplogroup C (Torroni *et al*. 1993), where it appears to be a haplogroup-diagnostic mutation. However, to date, the variant 16298C has only been detected on the haplogroup T background in Germans. A comparison of these HVS I sequences, therefore, reveals the instability at nps 16294 and 16296.

Many of the HVS I sequences belonging to the haplogroup T mtDNAs were identified in a study of linguistic isolates (German-speaking, Italian-speaking and Ladin-speaking samples) of the Eastern Italian Alps (Stenico *et al*. 1996). It was shown that 75 % of the Ladin sequences fall into haplogroup T (Stenico *et al*. 1996). The evolutionary relationships between most of these sequence types seems to be too ambiguous to obtain a fully resolved (without reticulations) phylogenetic tree. The problem of resolving the Alpine mtDNA phylogeny may be partially explained by the possibility of sequencing errors. However, the Alpine mtDNA data are useful in clarifying the problem of molecular instability in haplogroup T HVS I sequences. Among the HVS I data set of individuals inhabiting the Italian Alps we have found eight marker variants (16079T, 16085T, 16106A, 16145A, 16221T, 16280G, 16299G, 16343G) (Table 2). Three of these (16145A, 16221T, 16343G) have arisen several times in different mtDNA haplogroups, being found in Caucasian haplogroup J (16145A and 16221T), which has a common origin with haplogroup T (Richards *et al*. 1996), and in subgroup U3 (16343G) of haplogroup U (Richards *et al*. 1998). However, none of these variants was detected in any of the other 171 West Eurasian HVS I sequences analysed which belonged to haplogroup T. Therefore, the occurrence of these marker variants in the East Italian Alpine HVS I sequences, in association with various combinations of mutations at positions 16292, 16294 and 16296, gives additional evidence for molecular instability at positions 16292 and 16296 within the haplogroup T sequences (Table 2).

Overall, the data presented here allow us to conclude that within haplogroup T HVS I sequences there is considerable instability at nucleotide positions 16292 and 16296. We have not found any evidence for HVS I molecular instability at position 16304 because all of the possible marker mutations (16172C, 16189C, 16266T, 16311C) are characterized by dispersed geographic distribution. However, one cannot exclude the possibility that some haplogroup T mtDNAs containing variant 16304C were generated by independent mutational events at this site; additional sequence information is required to clarify this.

It is interesting that three marker mutations in the Alpine mtDNAs (16085T, 16106A, 16221T) occurred on two different haplotype backgrounds – T1 and T* – which are distinguished by three mutations at positions 16163, 16186, 16189 (Table 2). Thomas *et al*. (1998) observed a similar pattern in the HVS I sequences of two Scottish individuals with a 9 bp deletion in the MTCO2-$tRNA^{Lys}$ intergenic region of mtDNA. They have shown that the HVS I sequences of these 9 bp deleted individuals (Glasgow 34 and Glasgow 19) share two C–G transversions at positions 16306 and 16332 (Table 3) and have 00073A status in the HVS II. The two transversions common to individuals Glasgow 19 and Glasgow 34 were not detected in any other 9 bp deleted sequences or in the GenBank-EMBL sequence database (Thomas *et al*. 1998). We should note that the HVS I sequences of Glasgow 34 and German individual GER29 (Table 2), as well as Austrian individual AUT14 (Parson *et al*. 1998), share an identical combination of nucleotides at positions 16126, 16189, 16294, 16296, 16298, but the HVS I sequences of the Scottish individuals Glasgow 19 and Glasgow 34 differ by

Table 3. *Multiple losses of the haplogroup T diagnostic nucleotide variants in different populations*

| Sample origin | Marker mutations | HVS I sequence |
|---|---|---|
| Scottish[a] | **306G-332G** | Glasgow 34 – 126 189 294 296 298 **306G 332G** |
| | | Glasgow 19 – **306G** 327 **332G** |
| Ladins/Italians[b] | **069T-085T** | LS8 – **69** 79 **85** 126 244 294 296 |
| | | IS42 – **69 85** |
| | | IN50 – **69 85** 311 |
| Ladins[b] | **106A** | LS5 – **106** 126 145 184 280 294 296 344 |
| | | LC13 – 85 **106** 126 163 186 189 294 |
| | | LC18 – **106** |

Marker mutations are shown in bold, haplogroup T-diagnostic nucleotide variants (T* and T1) are underlined. HVS I sequences from the following studies were analysed: [a] Thomas *et al.* 1998, [b] Stenico *et al.* 1996.

six nucleotides, two of which (16126C and 16294T) appear to be diagnostic for the haplogroup T. The sharing of the two rare transversions and the 9 bp deletion, however, suggest a common origin for the HVS I in these two individuals, although formally their sequences belong to two different mitochondrial haplogroups (T for Glasgow 34 and, possibly, H for Glasgow 19). Thomas *et al.* (1998) have suggested that the large number of differences between these two Scottish sequences could occur if the three adjacent changes at positions 16294, 16296, and 16298 were linked by a single mutational event. However, another model – recombination between different mtDNA types – is also possible. For instance, a recent study of individuals from one isolated locality (a small island in Melanesia) (Hagelberg *et al.* 1999) presented evidence suggesting that recombination is a more likely explanation for the multiple occurrence of extremely rare mutational events in separate mtDNA lineages.

The Alpine mtDNA data (Stenico *et al.* 1996) contains several examples of multiple instability in the HVS I sequences belonging to the haplogroup T. Some of these cases are listed in Table 3. If we assume that certain mutations with a higher prevalence in some restricted geographic locations, such as 16069T–16085T in sequence types LS8, IS42, IN50 and 16106A in sequence types LS5, LC13, LC18, are identical-by-descent, then we can conclude that the losses of the haplogroup T diagnostic mutations (16126C–16294T) in the sequence types IS42, IN50 and LC18 were generated by a single mutational event.

Molecular instability in the haplogroup T HVS I sequences is therefore suggested by two lines of evidence: (1) multiple independent losses of the haplogroup T diagnostic nucleotide variants in different populations; (2) the occurrence of nucleotide changes at positions 16292 and 16296 in the HVS I sequence in individuals sharing rare population-specific polymorphisms. The results of the present study allow us to suggest that the identical haplogroup T HVS I sequence types might have arisen independently in different human populations. If this is correct then this should be taken into account in subsequent phylogenetic analyses, with respect to nucleotides 16292 and 16296.

There are other instances of molecular instability in the HVS I of the human mitochondrial genome. Howell *et al.* (1995, 1996) have found multiple HVS I mutations on the haplogroup J background in pedigrees with Leber's Hereditary Optic Neuropathy (LHON) within a span of 40–80 bp of one another. Moreover, Bendall & Sykes (1995) have identified another case of instability associated with heteroplasmic length variation in HVS I sequences. The mutation at position 16189 (from T to C) is frequently accompanied by heteroplasmic variation of 3 to 13 bp in the length of the polycytosine tract between positions 16184 and 16193. This length variation may result from instability of the poly-C tract due to the loss of the 16189T variant. Interestingly, additional mutations (at positions 16186 and 16192) which interrupt the poly-C tract appear to decrease or abolish the length variation (Bendall & Sykes, 1995; Marchington *et al.* 1996).

It is possible that certain nucleotides, or combination of nucleotides, within the mtDNA control region may alter the rate at which additional mutations occur within this region (Howell *et al.* 1996). The influence of the neighboring nucleotides is therefore one of the possible explanations for molecular instability in the mtDNA HVS I. In order to investigate this question with respect to the haplogroup T instability, we performed an extensive survey of previously published HVS I sequence data sets, including Europeans, Asians and Africans, for the 16294T variant. The survey revealed that this nucleotide variant appears to be diagnostic only for subsets of the HVS I sequences belonging to African haplogroups L1i and L2 (Watson *et al.* 1997). However, the occurrence of the 16294T variant in these sequences does not lead to instability at nps 16292 and 16296, as observed on the haplogroup T background, and it is clear therefore that the C-to-T transition at np 16294 is not the sole determinant of the haplogroup T mtDNA instability. However, it does seem likely that an important factor in generating the instability at np 16296 (and perhaps also at 16292) is the mutation at position 16294 in haplogroup T – the possibility exists that the 16294 mutation promotes back-mutation on these variants, but does not affect the forward rate.

Despite the significant amounts of experimental data that have been collected to date, the molecular mechanisms involved in the induction of the instability in the mtDNA are still unclear. Further investigations are necessary to determine the main factors involved in the maintenance or loss of molecular stability on the different mitochondrial haplogroups backgrounds.

This research was supported by grants from the Russian Fund for Basic Research (99 06 80430) and State Program Frontiers in Genetics. The authors would like to thank the anonymous reviewer for useful comments.

## REFERENCES

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., *et al.* (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Bandelt, H.-J., Forster, P., Sykes, B. C. & Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753.

Bendall, K. E. & Sykes, B. C. (1995). Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am. J. Hum. Genet.* **57**, 248–256.

Bertranpetit, J., Sala, J., Calafell, F., Underhill, P. A., Moral, P. & Comas, D. (1995). Human mitochondrial DNA variation and the origin of Basques. *Ann. Hum. Genet.* **59**, 63–81.

Bogenhagen, D. F. (1999). Repair of mtDNA in vertebrates. *Am. J. Hum. Genet.* **64**, 1276–1281.

Calafell, F., Underhill, P., Tolun, A., Angelicheva, D. & Kalaydjieva, L. (1996). From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann. Hum. Genet.* **65**, 35–49.

Chinnery, P. F. & Samuels, D. C. (1999). Relaxed replication of mtDNA: a model with implications for the expression of disease. *Am. J. Hum. Genet.* **64**, 1158–1165.

Comas, D., Calafell, F., Mateu, E., Perez Lezaun, A. & Bertranpetit, J. (1996). Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol. Biol. Evol.* **13**, 1067–1077.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., *et al.* (1998). Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am. J. Hum. Genet.* **63**, 1824–1838.

Corte-Real, H. B. S. M., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., *et al.* (1996). Genetic diversity in the Iberian peninsula determined from mitochondrial sequence analysis. *Ann. Hum. Genet.* **60**, 331–350.

Derenko, M. V. & Shields, G. F. (1997). Mitochondrial DNA sequence diversity in three North Asian aboriginal population groups. *Mol. Biol.* (*Moscow*) **31**, 665–669.

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

Eyre-Walker, A., Smith, N. H. & Smith, J. M. (1999). How clonal are human mitochondria? *Proc. R. Soc. Lond. B.* **266**, 477–483.

Francalacci, P., Bertranpetit, J., Calafell, F. & Underhill, P. A. (1996). Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am. J. Phys. Anthropol.* **100**, 443–460.

Hagelberg, E., Goldman, N., Lio, P., Whelan, S., Schiefenhovel, W., Clegg, J. B. & Bowden, D. K. (1999). Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc. R. Soc. Lond.* **266**, 485–492.

Hasegawa, M., Di Rienzo, A., Kocher, T. D. & Wilson, A. C. (1993). Toward a more accurate time scale for the human mitochondrial DNA tree. *J. Mol. Evol.* **37**, 347–354.

Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., *et al.* (1996). mtDNA polymorphism in east Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.* **59**, 579–590.

Howell, N., Kubacka, I., Halvorson, S., Howell, B., McCullough, D. A. & Mackey, D. (1995). Phylogenetic analysis of the mitochondrial genomes from Leber Hereditary Optic Neuropathy pedigrees. *Genetics* **140**, 285–302.

Howell, N., Kubacka, I. & Mackey, D. A. (1996). How rapidly does the human mitochondrial genome evolve? *Am. J. Hum. Genet.* **59**, 501–509.

Kolman, C. J., Sambuughin, N. & Bermingham, E. (1996). Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142**, 1321–1334.

Larsson, N. G. & Clayton, D. A. (1995). Molecular genetic aspects of human mitochondrial disorders. *Annu. Rev. Genet.* **29**, 151–178.

Lightowlers, R. N., Chinnery, P. F., Turnbull, D. M. & Howell, N. (1997). Mammalian mitochondrial genetics: hereditary, heteroplasmy and disease. *Trends Genet.* **13**, 450–455.

Lutz, S., Weisser, H.-J., Heizmann, J. & Pollak, S. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int. J. Legal Med.* **111**, 67–77.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., *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.

Malyarchuk, B. A. (1995). Association between nucleotide substitutions in cytochrome b gene and the control region in human mitochondrial DNA types. *Russ. J. Genet.* **31**, 846–848.

Malyarchuk, B. A. (1997). Origin of the Caucasoid BamHI-3/MspI-4 marker of mitochondrial DNA. *Russ. J. Genet.* **33**, 1424–1429.

Marchington, D. R., Poulton, J. Sellar, A. & Holt, I. J. (1996). Do sequence variants in the major non-coding region of the mitochondrial genome influence mitochondrial mutations associated with disease? *Hum. Mol. Genet.* **5**, 473–479.

Parson, W., Parsons, T. J., Scheithauer, R. & Holland, M. M. (1998). Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence analysis to a forensic case. *Int. J. Legal Med.* **111**, 124–132.

Pinto, F., Gonzalez, A. M., Hernandez, M., Larruga, J. M. & Cabrera, V. M. (1996). Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann. Hum. Genet.* **60**, 321–330.

Piercy, R., Sullivan, K. M., Benson, N. & Gill, P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Leg. Med.* **106**, 85–90.

Pult, I., Sajantila, A., Simanainen, J., Georgiev, O., Schaffner, W. & Paabo, S. (1994). Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biol. Chem. Hoppe Seyler* **375**, 837–840.

Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., *et al.* (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**, 185–203.

Richards, M. B., Macaulay, V. A., Bandelt, H.-J. & Sykes, B. C. (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* **62**, 241–260.

Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M. L., *et al.* (1995). Genes and languages in Europe – an analysis of mitochondrial lineages. *Genome Res.* **5**, 42–52.

Shields, G. F., Schmiechen, A. M., Frazier, B. L., Redd, A., Voevoda, M. I., Reed, J. K. & Ward, R. H. (1993). mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am. J. Hum. Genet.* **53**, 549–562.

Stenico, M., Nigro, L., Bertorelle, G., Calafell, F., Capitanio, M., Corrain, C., *et al.* (1996). High mitochondrial sequence diversity in linguistic isolates of the Alps. *Am. J. Hum. Genet.* **59**, 1363–1375.

Thomas, M. G., Cook, C. E., Miller, K. W. P., Waring, M. J. & Hagelberg, E. (1998). Molecular instability in the COII-tRNA$^{Lys}$ intergenic region of the human mitochondrial genome: multiple origins of the 9-bp deletion and heteroplasmy for expanded repeats. *Phil. Trans. R. Soc. Lond. B* **353**, 955–965.

Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Morral, P., Sellito, D., *et al.* (1998). MtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* **62**, 1137–1152.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., *et al.* (1996). Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**, 1835–1850.

Torroni, A., Lott, M. T., Cabell, M. F., Chen, Y. S., Lavergne, L. & Wallace, D. C. (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.

Wakeley, J. (1993). Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J. Mol. Evol.* **37**, 613–623.

Wallace, D. C. (1995). Mitochondrial DNA variation in human evolution, degenerative disease and aging. *Am. J. Hum. Genet.* **57**, 201–223.

Watson, E., Forster, P., Richards, M. & Bandelt, H-J. (1997). Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* **61**, 691–704.