Phylogenetic Network of the mtDNA Haplogroup U in Northern Finland Based on Sequence Analysis of the Complete Coding Region by Conformation-Sensitive Gel Electrophoresis

Saara Finnilä,[1,2] Ilmo E. Hassinen,[2] Leena Ala-Kokko,[2,3] and Kari Majamaa[1,2]

Departments of [1]Neurology, [2]Medical Biochemistry, and [3]Collagen Research Unit and Biocenter, University of Oulu, Oulu, Finland

Mutations in mtDNA have accumulated sequentially, and maternal lineages have diverged to form population-specific genotypes. Classification of the genotypes has been made based on differences found in restriction fragment analysis of the coding region or in the sequence of the hypervariable segment I. Both methods have shortcomings, as the former may not detect all the important polymorphisms and the latter makes use of a segment containing hypervariable nucleotide positions. Here, we have used conformation-sensitive gel electrophoresis (CSGE) to detect polymorphisms within the coding region of mtDNA from 22 Finns belonging to haplogroup U. Sixty-three overlapping PCR fragments covering the entire coding region were analyzed by CSGE, and the fragments that differed in their migration pattern were sequenced. CSGE proved to be a sensitive and specific method for identifying mtDNA substitutions. The phylogenetic network of the 22 coding-region sequences constituted a perfect tree, free of homoplasy, and provided several previously unidentified common polymorphisms characterizing subgroups of U. After contrasting this data with that of hypervariable segment I, we concluded that position 16192 seems to be prone to recurrent mutations and that position 16270 has experienced a back mutation. Interestingly, all 22 samples were found to belong to subcluster U5, suggesting that this subcluster is more frequent in Finns than in other European populations. Complete sequence data of the mtDNA yield a more reliable phylogenetic network and a more accurate classification of the haplogroups than previous ones. In medical genetics, such networks may help to decide between a rare polymorphism and a pathogenic mutation; in population genetics, the networks may enable more detailed analyses of population history and mtDNA evolution.

Introduction

The rate of sequence evolution in mtDNA is 10–20 times higher than that in the nuclear genome, and, consequently, any two mtDNAs may differ by 10–66 nt from each other (Zeviani et al. 1998; Chinnery et al. 1999). The mutations have accumulated sequentially along radiating maternal lineages and now characterize human populations in different geographical regions of the world. Restriction fragment length polymorphism (RFLP) studies of mtDNA coding regions have been used to classify such lineages. RFLPs have revealed a number of stable polymorphic sites that define mtDNA haplogroups, most of which have been shown to be continent-specific (Wallace 1994). Ten haplogroups (H, I, J, K, M, T, U, V, W, and X) encompass almost all mtDNAs from European populations (Torroni et al. 1996).

Haplogroup U, defined by 12308A→G, is almost entirely specific to Europeans, and it is found only infrequently in sub-Saharan populations (Torroni et al. 1996). Estimated minimum divergence times have suggested that this haplogroup is ancient, with an estimated age of ~50,000 years (Richards et al. 1996, 1998; Torroni et al. 1996). Comprehensive analyses of European mtDNAs have suggested that haplogroup U is composed of at least five subclusters, termed "U1"–"U5" (Richards et al. 1996, 1998; Macaulay et al. 1999). The frequency of haplogroup U among Europeans is ~7%, but it appears to be several-fold higher in Finland (Richards et al. 1996; Torroni et al. 1996). Furthermore, we have recently identified a subcluster of U that is characterized by 5656A→G and that may be ~30-fold more common in Finland than in two other European populations (Finnilä et al. 1999).

The common set of enzymes used in RFLP allow only ~20% of the mtDNA sequence to be examined (Wallace 1994), and, therefore, a number of polymorphisms may remain undetected—for example, the transitions 12308A→G and 5656A→G. Therefore, sequence analysis of a 300-bp hypervariable segment I (HVS-I) within the D loop has also been used to characterize mtDNA haplogroups. Parallel mutations and back mutations at the hypervariable sites within this segment may com-

Received March 18, 1999; accepted for publication December 17, 1999; electronically published March 1, 2000.

Address for correspondence and reprints: Dr. Kari Majamaa, University of Oulu, Department of Neurology, P.O. Box 5000, FIN-90014, University of Oulu, Finland. E-mail: kari.majamaa@oulu.fi

0002-9297/2000/6603-0024$02.00

plicate the classification of the genotypes, but, generally, a good correlation has been obtained between the RFLP data and HVS-I sequence data (Torroni et al. 1996, 1998; Macaulay et al. 1999).

None of the existing haplogroups is based on the complete sequence of mtDNA, probably because the task would be a laborious one. In recent years, different methods based on gel electrophoresis of PCR-generated products have been developed for detecting single-base mismatches in DNA, including single-strand conformation polymorphism analysis (Thomas et al. 1994), denaturing gradient gel electrophoresis (Gross et al. 1994), low-stringency single-specific primer PCR (Pena et al. 1994), and mutation detection enhancement gel matrix (Alonso et al. 1996). Conformation-sensitive gel electrophoresis (CSGE) is based on the separation of heteroduplexes containing single-base-pair mismatches from homoduplexes in a polyacrylamide gel (Ganguly et al. 1993). Here, we assess the applicability of CSGE for determining mtDNA nucleotide substitutions in the coding sequence and analyze the entire genotype of the mtDNA haplogroup U in 22 healthy Finns. Haplogroup U was selected for the study because of its high frequency in Finland compared with that in other European populations (Richards et al. 1996; Torroni et al. 1996). We also chose it because we have recently shown that mtDNA haplogroup U may be a risk factor for occipital stroke among patients with migraine (Majamaa et al. 1998). The data enabled us to construct a phylogenetic network of haplogroup U that was made on the basis of the complete coding-region sequence and to compare this network with that on the basis of the HVS-I sequence.

Subjects and Methods

Patients and Samples

Blood samples were obtained from 77 Finns who reported that they and their mothers were healthy, without diabetes mellitus, sensorineural hearing impairment, or neurological ailments. Furthermore, we required of the patients that their matrilineal ancestors had been born in northern Finland—that is, in the provinces of Northern Ostrobothnia or Kainuu—before the year 1900. Recent relationships among the subjects were excluded by the time and place of birth of their matrilineal ancestors. After obtaining this information, the patients were anonymized. The research protocol was approved by the Ethics Committee of the Medical Faculty, University of Oulu.

Molecular Methods

DNA extraction.—Total DNA was isolated from the blood cells with a QIAamp blood kit (Qiagen). Platelet mitochondria were isolated from 40 ml of venous blood by fractionating centrifugation (Ausenda and Chomyn 1996), and DNA was purified from the isolated mitochondria with the QIAamp blood kit.

Analysis of mtDNA haplogroups.—The mtDNA haplogroups were determined by restriction digestion to identify the most informative polymorphic sites, and the various mtDNA haplogroups were then defined according to the published criteria (Torroni et al. 1996), except that 12308A→G was detected by *Dde*I digestion of a PCR product amplified in the presence of a mismatched (underlined nucleotide) forward primer 12279–12307 (5′-AAC AGC TAT CCA TTG GTC TTA GGC CC$\underline{\text{T}}$ AA-3′). The cleavage was observed when 12308A→G was present.

PCR for CSGE

Sixty-three pairs of primers were designed for amplifying the coding region (nts 523–16090) of mtDNA according to the Cambridge reference sequence (Anderson et al. 1981). The mean size of the amplified fragments was 354 bp, and the neighboring PCR products were designed to overlap by ⩾80 bp at both ends. The template DNA was amplified in a total volume of 50 μl by PCR in 30 cycles through denaturation at 94°C for 1 min, annealing at a primer-specific temperature for 1 min and extensions at 72°C for 1 min, including a final extension at 72°C for 10 min. The quality of the amplified fragment was estimated visually on a 1.5% agarose gel; then a suitable amount of the PCR product, usually 3–10 μl, was taken for heteroduplex formation. Each amplified fragment of haplogroup U was mixed with the corresponding fragment amplified on a haplogroup H template or haplogroup I template. The amplified fragments were denatured at 95°C for 5 min. The heteroduplexes were subsequently allowed to anneal at 68°C for 30 min.

Conformation-Sensitive Gel Electrophoresis

CSGE was carried out essentially as described earlier (Körkkö et al. 1998). A 1-mm-thick gel with a 36-well comb was prepared with 15% polyacrylamide, a 99:1 ratio of acrylamide to 1,4-bis(acryloyl)piperazine (Fluka, Buchs, Switzerland), 10% ethylene glycol, 15% formamide (Gibco BRL), 0.1% ammonium persulphate, and 0.07% N,N,N′,N′-tetramethylethylenediamine in 0.5 × TTE buffer (44.4 mM Tris, 14.25 mM taurine, 0.1 mM EDTA; pH 9.0). The loading buffer stock for PCR products was a 10 × solution of 30% glycerol, 0.25% bromophenol blue, and 0.25% xylene cyanol FF. The PCR products were analyzed on a standard DNA-sequencing gel apparatus with 0.5 × TTE as the electrode buffer. The gel was preelectrophoresed for 30 min; then, the samples were electrophoresed through the gel at a constant voltage of 400 V overnight at room temperature. After electrophoresis, the gel was stained on the glass plate in 150 μg/liter of ethidium bromide for 5 min; the stain was then removed in water. The gel was then transferred to an ultraviolet transluminator and photographed (Grab-IT Annotating Grabber 2.04.7 [UVP Inc.]).

Sequencing

Selected PCR fragments covering the coding region and the entire D loop of the 22 samples belonging to haplogroup U were analyzed by automated sequencing (ABI PRISM 377 Sequencer with Dye Terminator Cycle Sequencing Ready kit [PE Biosystems]) after treatment with exonuclease I and shrimp alkaline phosphatase (Werle et al. 1994). The primers used for sequencing of the coding region were the same as those used in the amplification reactions for CSGE. The D loop was amplified in two fragments spanning between nts 15714 and 16555 and nts 16449 and 725, respectively, and the sequence was determined between nts 16024 and 576. The reference samples belonging to haplogroup H and I were sequenced between nts 568 and 16400.

Phylogenetic Analysis

The phylogenetic networks were based on the median algorithm (Bandelt et al. 1995).

Results

CSGE Analysis of Amplified Fragments

It was found that 22 of 77 healthy controls belonged to haplogroup U, suggesting a population frequency of 29%. In order to estimate the usefulness of CSGE for analyzing differences within the coding sequence of mtDNA, we examined DNA from these 22 people. Sequence differences within 63 overlapping fragments were screened by CSGE (table 1). Heteroduplexes were allowed to form between an amplified fragment from each person belonging to haplogroup U and the corresponding fragment amplified from a reference DNA sample belonging to either haplogroup H (heteroduplex U/H) or haplogroup I (heteroduplex U/I). All the heteroduplexes obtained from 21 of the PCR fragments migrated in one band, suggesting that there were no sequence differences between DNA from the three haplogroups. Poorly visualized bands on CSGE were observed in the case of heteroduplexes from seven fragments covering regions for 12SRNA, 16SRNA (two fragments), COX1, ATPase, ND4, and ND5 genes in each of the 22 samples. Homologous sequences corresponding to these regions have been shown to exist as pseudogenes in the nuclear genome (Wallace et al. 1997). Use of DNA from isolated platelet mitochondria as the template did not improve the resolution; however, whereas a shift in primer location by 50–100 nt resulted in sharp bands on CSGE and allowed the detection of three more fragments migrating as one band in each of the 22 samples. A total of 24 amplified fragments therefore could be suggested as identical in sequence within the 22 samples representing haplogroup U and the two reference samples representing mtDNA haplogroups H and I.

We found at least one sample that yielded more than one band in either the heteroduplex U/H or heteroduplex U/I among the remaining 39 fragments, suggesting a difference among the sequences (fig. 1).

Sequence Analysis of mtDNA

The two reference DNA samples belonging to haplogroups H and I were sequenced in their entirety, with the exception of the hypervariable segment II (HVS-II) region, and a total of 104 haplogroup U samples from the 63 fragments were sequenced, including at least one sample from each heteroduplex that had migrated in one band and one sample from each heteroduplex that had differed in migration pattern (table 1).

The sensitivity and specificity of CSGE for detecting single-nucleotide changes were then calculated. Sequence data were obtained for 104 U/H and U/I heteroduplexes, 87 of which differed in the nucleotide sequence of the two fragments. In the case of the U/H heteroduplexes, CSGE suggested a similar primary structure in 45 fragments and a dissimilar primary structure in 42 fragments, whereas, in the case of the U/I heteroduplexes, CSGE suggested a similar primary structure in 44 fragments and a dissimilar primary structure in 43 fragments. Three fragments turned out to be false negatives, suggesting that a sensitivity of .96 had been achieved by CSGE in detecting sequence differences in mtDNA. The specificity was found to be 1.0.

CSGE was reproducible, as no interassay variation was found when a given heteroduplex PCR fragment was loaded on separate gels. Furthermore, identical migration was observed in each lane of a gel when multiples of a given heteroduplex PCR fragment were electrophoresed, suggesting that the intraassay variation is negligible.

Phylogenetic Networks for Haplogroup U

In view of the high sensitivity and specificity of CSGE in detecting sequence differences, we assumed that all identical migration patterns in CSGE represented identical sequences. We thus obtained sequence information on the complete coding region covering 15,567 nt of 22 samples belonging to haplogroup U, suggesting that a total of 342,474 nt had been scanned. Using this information, we outlined a novel phylogenetic network for haplogroup U (fig. 2A). The network was found to be

Table 1

Findings in CSGE of 22 Samples Belonging to mtDNA Haplogroup U

		Primer																							
Fragment	Gene	L	H	P11	P15	P16	P17	P19	P20	P22	P28	P30	P32	P37	P40	P41	P43	P47	P49	P54	P60	P64	P65	P70	P76
1	Phe, 12SRNA	523	725	++	++	++1	−+2	++	−+	−+	−+	††3	++	−+	−+	−+	−+	−+	++	++	−+	−+	−+	−+	−+
2	12SRNA	628	950	−−	−−4	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
3	12SRNA	861	1246	−−	−−	−−	−−	−−	−−	−−5	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
4	12SRNA	1137*	1473*	−−	−−	−−	−−	−−	−−	−−6	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
5	12SRNA, Val	1357	1696	−−	−−	−−	−−	−−7	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
6	16SRNA	1615	1900	−+	−+	−+	++8	−+	−+	−+9	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+
7	16SRNA	1811	2179	−−	−−	−−	−−	−−	−−	−−10	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
8	16SRNA	2054*	2440*	−−	−−	−−11	−−12	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
9	16SRNA	2334	2688	−−	−−	−−13	++14	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
10	16SRNA	2586	2965	+−	+−	+−15	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−
11	16SRNA, Leu (UUR)	2866	3263	+−	+−	+−16	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−
12	16SRNA, Leu (UUR)	2996*	3338*	++	++	++17	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
13	Leu (UUR), ND1	3144	3553	−−	−−	−−18	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
14	ND1	3452	3796	−−	−−	−−	−−	−−	−−	−−19	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
15	ND1	3706	4055	−−	−−	−−	−−	−−	−−	−−20	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
16	ND1, Ile, Gln, Met	3969	4324	−−21	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	++22	−−	−−	−−	−−	−−	−−	−−	++23	−−
17	Ile, Gln, Met, ND2	4231	4508	+−24	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−25	+−	+−	+−	+−	+−	+−	+−	+−	+−
18	ND2	4409	4739	++26	++	++	++	++	++	++	++	++	++	++	++	++27	++	++	++	++	++	++	++	++	++
19	ND2	4638	5000	−−	−−	−−	++28	−−	−−	−−	−−	−−	−−	−−29	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
20	ND2	4898	5267	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−30	++31	−−	−−	−−	−−	−−	−−	−−	−−	−−	++32
21	ND2	5161	5493	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−33	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
22	ND2, Trp, Ala, Asn	5379	5680	−−	−−	−−	−−	−−	−−	−−	−−	−−34	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
23	Asn, OL, Cys, Tyr	5548	5917	++	++	++	−−	++	++	−−	−−35	++36	++	−−	−−	++	++	++	++	++	−−	++	−−	++	−−
24	Tyr, COX1	5854	6234	−−	−−	−−	−−	−−	−−37	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
25	COX1	6134	6514	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−38	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
26	COX1	6383*	6778*	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+39	−+40	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+
27	COX1	6690	7079	+−	+−	+−	+−	+−	+−	+−41	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−
28	COX1	7021	7339	−−	−−	−−	−−	−−	−−	−−42	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
29	COX1, Ser	7252	7634	+−	+−	+−	+−	+−	††43	+−44	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−

30	Asp, COX2	7533	7915	++	††45	++	++	++	++	−−	−−	++	++	−−	−−	++	++	++46	++	++	‡‡47	++	−−	++	−−
31	COX2	7813	8203	−−	−−48	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	++49	−−	−−	−−	−−
32	COX2, Lys	8100	8437	−+	−+	−+	−+	−+	−+	−+50	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+
33	Lys, ND6L	8368	8748	−−	−−	−−	−−	−−	−−	−−51	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
34	ATPase	8648	9034	++52	−−53	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
35	ATPase, COX3	8901*	9321*	+−	+−54	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	++55	+−	+−	+−	+−
36	COX3	9211	9595	++	++	++	++	++	++	++	++	++	++	++	++56	++	++	++	++	++	++	++	++	++	++
37	COX3	9485	9875	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	++57	−−	−−	−−	−−58	−−	−−	−−	−−	−−	++59
38	COX3	9772	10107	−+	−+	−+	−+	−+	−+	−+60	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+
39	COX3, Glu	9966	10399	−+	−+	−+	−+	−+	−+	−+61	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+
40	ND3, Arg	10288	10700	−+	−+	−+	−+	−+	−+	−+62	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+
41	ND4L	10590	10942	−−	−−	−−	−−	−−	−−	−−63	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
42	ND4	10840	11110	−−	−−	−−	−−	−−	++64	−−	−−	−−65	−−	−−	−−	++	++66	++	−−	−−	−−	++	−−	++	−−
43	ND4	11010	11390	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−67	−−	−−	−−	−−	−−	−−	−−	−−
44	ND4	11290	11541	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++68	++69	++	++	++
45	ND4	11431*	11814*	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−70	−−71	−−	−−	−−
46	ND4	11711	12100	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−72	−−	−−	−−	−−
47	ND4, His	12000	12386	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++73	++	++	++	++
48	Ser, Leu (CUN), ND5	12217*	12595*	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++74	++75	++	++	++
49	ND5	12500	12854	††	††	††	−+76	††77	††	‡‡78	‡‡	††	††	‡‡	−+	††	††	††	††	††	−+	††	−+	††	−+
50	ND5	12750	13050	−−	−−	−−	−−	−−	−−	−−79	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
51	ND5	12950	13270	−−	−−	−−	−−	−−	−−	−−80	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−81	−−	−−	−−	−−	−−
52	ND5	13172	13570	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−82	−−	−−	−−	−−	−−
53	ND5	13462	13851	++	++	++	††83	++	++	++	++	++	++	++	++	++	++84	++	++	++	++	++	++	++	++
54	ND5	13696	14007	−−	−−	−−	−−	−−	−−	++85	++	−−	−−	++86	−−	−−	−−	−−	−−	−−87	−−	−−	++	−−	−−
55	ND5	13914	14208	−−	−−	−−	−−	−−	−−	−−88	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−
56	ND5, ND6	14080	14400	++	++	++	++	++	++	−−89	−−	++	++	−−	−−	++	++	++90	++	++	−−	++	††91	++	−−
57	ND6	14300	14660	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−92	−−	−−	−−	−−	−−	−−	−−
58	ND6	14560	14820	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−	+−93	+−	+−	+−	+−	+−	+−	+−
59	ND6, Glu, Cytb	14620	15002	+−	+−	+−	+−94	+−	+−	++	++	+−	+−	++95	++96	+−	+−	+−	+−	+−	++	+−	++	+−	++
60	Cytb	14900	15280	−−	−−	−−	−−97	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	++98
61	Cytb	15166	15545	−−	−−	−−	++99	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−100	−−	−−	−−	−−	−−	−−
62	Cytb	15431	15813	−−	−−	−−	++101	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−	−−102	−−	−−	−−	−−	−−	−−
63	Cytb, Thr, Pro	15714	16090	−+	−+	−+	−+103	−+	−+104	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+	−+

NOTE.—The first symbol indicates the migration pattern of the U/H heteroduplexes, and the second symbol indicates that of the U/I heteroduplex. −= Heteroduplex migrates in one band. +, †, and ‡ denote heteroduplex migration in several bands, each symbol indicating a similar migration pattern of the heteroduplex. The numbers 1–104 indicate the fragments that were sequenced. The location of the first nucleotide in the L primer (forward primer) and H primer (reverse primer) is shown. The mean primer length is 22 nt. Fragments that were amplified using a second set of primers to improve resolution are marked by an asterisk.

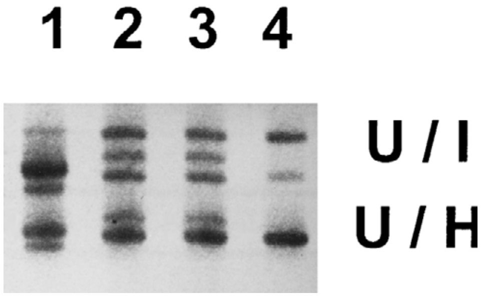

Figure 1 CSGE of PCR fragment 49 (seetable 1). Lane 1, sample P22; lane 2, P20; lane 3, P19; and lane 4, P17. Sequencing of samples indicated the presence of 12630G→A in P22, 12618G→A in P20, and P19 and 12705C→T in reference sample I. The reference sample H and sample P17 were identical to the Cambridge reference sequence.

a perfect tree, one with no homoplasy; this tree provided several previously unidentified common polymorphisms. The median network constructed for HVS-I sequence data (fig. 2B) included three reticulations, but in general, it correlated well with the network obtained for the coding region. Comparison of these two networks suggests that the reticulation involving sample P17 may be resolved by assuming a back mutation at nt 16270. Furthermore, implementation of HVS-II sequence data (table 2) suggests that the largest node in the network has accumulated transitions at nts 16270 and 16189 and—as the last step—at nts 16192 and 217, as well as a length polymorphism in the C tract between nts 568 and 573. Finally, comparison of the two networks led us to favor the assumption that the lineage of P60 consists of 16270C→T and 16256C→T. Inclusion of HVS-II sequence data, however, would create a new, unresolved reticulation of transitions at nts 16192 and 16526.

Four complete mtDNA coding-region sequences containing 12308A→G are available for comparison (Ozawa 1995; Arnason et al. 1996; Ohlenbusch et al. 1998). The sequences of patient P-8 (Ozawa 1995) and patient MS128 (Ohlenbusch et al. 1998) harbor many polymorphisms common with our haplogroup U network; the sequences also share the additional polymorphisms at nts 1811 and 9698 but lack the polymorphisms at nts 3197, 9477, and 13617, suggesting that the genotypes of patients P-8 and MS128 depart from our network (fig. 2A). On the other hand, the two other sequences could be placed in the network. The sequence of patient MS88 (Ohlenbusch et al. 1998) is similar but not identical to our sequence P17 (fig. 2A); it harbors additional polymorphisms at nts 2883, 5097, 7482, 12406, and 14223 (Ohlenbusch et al. 1998). The human sequence X93334 (accession number at The European Molecular Biology Laboratory; Arnason et al. 1996) was found to be identical to the node with 14793A→G as the last mutation, but it diverged thereafter from the network by eight substitutions (fig. 2A). One of these substitutions was at nt 15218, which was also present in our sample P76, whereas none of the remaining seven substitutions were found elsewhere in the network.

Discussion

CSGE proved to be a reproducible, sensitive, and specific method for detecting nucleotide substitutions in mtDNA. Comparison of the CSGE data with the actual sequence data suggested that the sensitivity of CSGE was 96%, similar to that reported for fragments amplified from nuclear DNA templates (Ganguly et al. 1993; Ganguly and Prockop 1995; Körkkö et al. 1998). SSCP may not detect all mutations in mtDNA (Thomas et al. 1994), and its reported sensitivity is 84% (Jordanova et al. 1997). Methods based on low-stringency, single-specific primer PCR (Pena et al. 1994) and single-strand conformation analysis in a mutation detection–enhancement gel matrix (Alonso et al. 1996) have been used for mtDNA genotyping, but no analysis of sensitivity was presented in either case.

We detected three false-negative results among 87 U/H heteroduplexes and 87 U/I heteroduplexes, but we found no false-positive ones. The three false-negative heteroduplexes were located within GC-rich sequences. In the original report on CSGE, 60 of 68 single-base mismatches were detected (Ganguly et al. 1993), the 8 undetected substitutions being either in high-melting-temperature domains or within 50 bp of the nearest end of the PCR product. Therefore, to reduce the probability of failing to detect substitutions located in the terminal parts of the PCR fragments, ~60 bp flanking regions have been suggested (Körkkö et al. 1998). We designed fragments to overlap by at least 80 nt, which was probably more than required, as the shortest distance of an identified substitution from the nearest end of the fragment was 44 bp. The size of the PCR fragment to be analyzed is also critical; originally, PCR products of size 200–800 bp were analyzed on CSGE (Ganguly et al. 1993). Later work has shown that single-base mismatches in the high-melting domains are readily detected by reducing the maximal size of the PCR products to 450 bp (Körkkö et al. 1998). Therefore, we used PCR fragments of size 270–434 bp.

We evaluated the applicability of CSGE for detecting nucleotide substitutions in mtDNA by resolving the polymorphic structure of the mtDNA haplogroup U. Many previously unidentified polymorphisms were found, including six that were found to be in common for all samples. None of these six polymorphisms has been detected in restriction-fragment analysis with the commonly used set of enzymes. The phylogenetic network of haplogroup U turned out to be a perfect tree

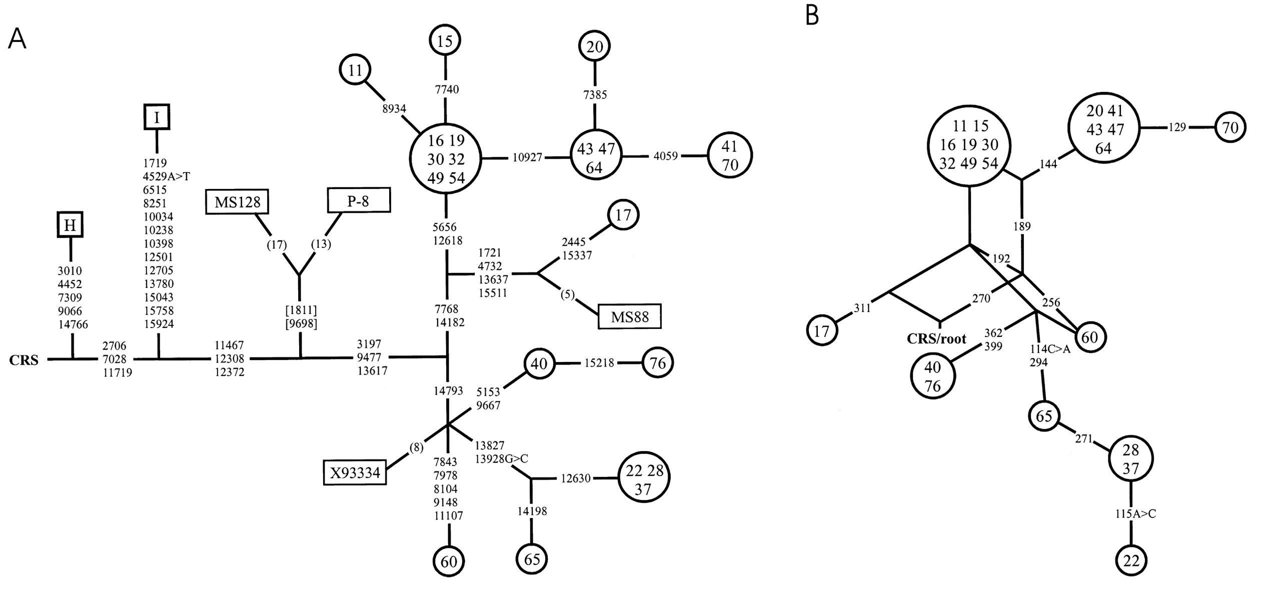

Figure 2 Phylogenetic network of Finnish mtDNA haplogroup U. Among the numbers inside the circles, codes without the prefix "P" of 22 samples belong to haplogroup U. H = reference sample belonging to haplogroup H; I = reference sample belonging to haplogroup I; CRS = Cambridge reference sequence. The substitutions in the networks are transitions, unless otherwise marked. *A,* Network based on variation in the coding sequence. The numbers on the lines connecting the nodes denote polymorphic nucleotides. The following substitutions were common to all haplogroup U samples and the reference samples H and I: 750A→G, 1438A→G, 3106delC, 3423G→T, 4769A→G, 4985G→A, 8860A→G, 9559G→C, 11335T→C, 13702G→C, 14199G→T, 14272G→C, 14365G→C, 14368G→C, and 15326A→G. Neither deletions, insertions, nor length polymorphisms were detected in the coding region. Previously published complete sequences harboring 12308A→G (P-8 [Ozawa 1995], X93334 [Arnason et al. 1996], and MS88 and MS128 [Ohlenbusch et al. 1998]) were positioned in the network. The codes inside rectangles refer to those in the original publications. The numbers in parentheses show the number of divergent nucleotides from the nearest node in the network. The numbers in brackets show two substitutions that were shared by two of the previously published sequences but were not detected in the 22 Finnish samples. The previously published sequences were positioned in the network disregarding the discrepancy in the data on 3106delC and 14272G→C that are not present in X93334 (Arnason et al. 1996) and 11719G→A that is not present in MS88 (Ohlenbusch et al. 1998). The root is not shown in the figure, but it should join to the branch leading to reference sample I (Macaulay et al. 1999). *B,* Median network based on the sequence of the HVS-I (nts 16090–16400). The numbers on the lines connecting the nodes denote polymorphic nucleotides with the first two digits, 16, omitted. An uninterrupted cytosine tract of 10 or 11 residues was detected between nt 16184 and nt 16193 by direct sequencing in samples harboring the motif 16144T→C, 16189T→C, and 16270C→T. Accurate determination of the length of the cytosine tract requires cloning (Bendall and Sykes 1995), and therefore this polymorphism is not included in the network. Reference sample H harbored 16093T→C, 16183delC, and 16189T→C in HVS-I, and reference sample I harbored 16129G→A, 16184A→C, 16223C→T, and 16391G→A. The reference samples are not shown in the network. The root and the CRS coincide and are labeled as CRS/root.

(fig. 2A). All the samples could be positioned in the network, and no back mutations or parallelisms needed to be assumed. The network based on the coding region was concordant with the median network based on HVS-I sequence, in accordance with previous observations (Torroni et al. 1996; Macaulay et al. 1999). Furthermore, comparison of the networks suggested that nt 16192 is a variable site and that nt 16270 has experienced a back mutation.

Interestingly, we found that all of the 22 samples with 12308A→G belonged to subcluster U5 (Richards et al. 1998; Macaulay et al. 1999), suggesting that haplogroup U is quite restricted in its variation in Finland. The major colonization of northern Finland has taken place after the 16th century (Pitkänen 1994), and founder effect by a relatively small number of settlers may confound the extant haplogroup frequencies. However, genetic homogeneity of the Finns—contributed to by the isolation of the Finnish population for geographical, linguistic, and cultural reasons (de la Chapelle 1993; Peltonen et al. 1995)—would imply that the restricted variation found in haplogroup U may be extrapolated to the entire population. The frequency of subcluster U5 in Finns appears to be clearly higher than that in other European populations. An analysis of 67 German mtDNA genotypes has included 11 mtDNAs with 12308A→G, three of which belonged to haplogroup K and four of which belonged to subclusters U4 and U5, respectively, suggesting a frequency of 6% for each subcluster (Hofmann et al. 1997). We found that 22 of 77 samples (29%) belonged to subcluster U5, suggesting a five- to sixfold higher frequency in Finns than in Germans. The other subclusters that have been detected among Europeans (U1–U4) were not found in Finns.

Table 2

Polymorphisms Detected in the Segment between nts 16401 and 574 of mtDNA in 22 Samples Belonging to Haplogroup U

Sample	mtDNA Variant												
Code	16428G→A	16526G→A	73A→G	146T→C	150C→T	152T→C	217T→C	248delA	249A→G	263A→G	303C	311insC	514delCA	568C
P11	−	−	+	−	+	−	+	−	−	+	8	+	−	7
P15	−	−	+	−	+	−	+	−	−	+	8	+	−	7
P16	−	−	+	−	+	−	+	−	−	+	8	+	−	7
P17	−	−	+	−	+	−	−	+	−	+	7	+	−	6
P19	−	−	+	−	+	−	+	−	−	+	7	+	−	7
P20	−	−	+	−	+	−	−	−	−	+	7	+	−	6
P22	−	+	+	−	−	−	−	−	−	+	8	+	−	6
P28	−	+	+	−	−	−	−	−	−	+	8	+	−	6
P30	−	−	+	+	+	−	+	−	−	+	9	+	−	8
P32	−	−	+	−	+	−	+	−	−	+	8	+	−	7
P37	−	+	+	−	−	−	−	−	−	+	7	+	−	6
P40	+	−	+	−	−	−	−	−	+	+	7	+	−	6
P41	−	−	+	−	+	−	−	−	−	+	7	+	−	6
P43	−	−	+	−	+	−	−	−	−	+	8	+	−	6
P47	−	−	+	−	+	−	−	−	−	+	7	+	−	6
P49	−	−	+	−	+	−	+	−	−	+	8	+	−	7
P54	−	−	+	−	+	−	+	−	−	+	8	+	−	7
P60	−	+	+	−	−	−	−	−	−	+	8	+	+	6
P64	−	−	+	−	+	−	−	−	−	+	8	+	−	6
P65	−	+	+	−	−	+	−	−	−	+	7	+	−	6
P70	−	−	+	−	+	−	−	−	−	+	7	+	−	6
P76	−	−	+	−	−	−	−	−	+	+	7	+	−	6

Note.—The absence (−) or the presence (+) of the variant is shown. The number of residues in the cytosine tract beginning at nt 303 and nt 568, respectively, is shown.

The upper limit of the 95% confidence interval for a binomial proportion of 0 of 77 is 4.7, suggesting, however, that a haplotype frequency as high as 6% may have remained undetected by chance.

Most phylogenetic analyses use HVS-I sequence data. However, there are many variable sites within this segment that give rise to a multitude of haplotypes and thus complicate the analysis of phylogenetic relationships and comparisons among populations. The structure of our haplogroup U network was quite straightforward, as it was composed of three distinct branches: an upper branch characterized by 5656A→G and 12618G→A; a lower branch characterized by 14793A→G; and the branch of sample P17 (see figs. 2*A* and 2*B*). The HVS-I sequence suggests that the upper branch is similar to the previously defined subcluster U5b (Richards et al. 1998). This subcluster may be quite specific for the Finns, as we have recently detected that 5656A→G is ~30-fold more frequent in Finns than in two other European populations (Finnilä et al. 1999). Furthermore, part of the upper branch is characterized by the HVS-I motif, including 16144T→C, 16189T→C, and 16270C→T, which has been considered to be fairly specific for the Saami, as it is not found among other Europeans except for the Finns (Sajantila et al. 1995; Lahermo et al. 1996). These findings suggest that the haplotype characterized by the coding region variants 5656A→G and 12618G→A may be highly specific for the Finns and the Saami. On the other hand, the branch of P17 appears to be fairly rare in Finns, whereas its most probable counterpart, subcluster U5a (Richards et al. 1998), is not uncommon among other Europeans. Analysis of our data indicates that U5a is more closely related to subcluster U5b than to subcluster U5a1, the lower branch in our network.

Comparison of our haplogroup U network with four previously published complete mtDNA sequences containing 12308A→G (Ozawa 1995; Arnason et al. 1996; Ohlenbusch et al. 1998) revealed that two of the sequences departed clearly from the network. Indeed, patient MS128 (Ohlenbusch et al. 1998) harbors the substitutions at nts 16224 and 16311 that characterize haplogroup K. This haplogroup has recently been suggested to be a subcluster of haplogroup U (Richards et al. 1998; Macaulay et al. 1999), and the polymorphisms 1811A→G and 9698T→C shared by MS128 and P-8 (Ozawa 1995) may therefore mark the origin of a subcluster that includes at least haplogroup K. This subcluster apparently includes additional branches, as the sequence of patient P-8 did not conform to any of the known motifs in haplogroup U5K (Richards et al. 1998; Macaulay et al. 1999). The two other sequences (Arnason et al. 1996; Ohlenbusch et al. 1998) could be unambiguously placed in the network, suggesting that we have detected all the important polymorphisms that characterize haplogroup U5. One of the sequences (Arnason et al. 1996) differed by five substitutions and the other (Ohlenbusch et al. 1998) by eight substitutions from the nearest node in the network. These differences may indicate differences among populations.

We found CSGE to be an ideal method for screening mutations and polymorphisms in mtDNA, where the high frequency of variants makes sequencing a laborious task and where restriction-fragment analysis lacks sensitivity. Furthermore, the method has the advantages of being nonradioactive and permitting large capacity. Sequence information was obtained with relatively little effort, and the data acquired by CSGE enabled a phylogenetic network to be constructed for the Finnish mtDNA haplogroup U. The effectiveness of CSGE was demonstrated by the identification of several new, informative polymorphisms determining haplogroup U and its subcluster U5. Similar phylogenetic networks are required for the purposes of medical and population genetics. Such networks would help in making a decision between a rare polymorphism and a pathogenic mutation in clinically affected people. Likewise, the networks enable more detailed comparisons among and within populations to be carried out and also enable more accurate phylogenetic relationships to be determined.

Acknowledgments

The expert technical assistance of Ms. Irma Vuoti, Ms. Kaisu Korhonen, and Ms. Anja Heikkinen is acknowledged. This work was supported in part by grants from the Medical Research Council of the Academy of Finland and from the Sigrid Juselius Foundation.

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

European Molecular Biology Laboratory, http://www.psc.edu/general/software/packages/embl/embl.html

References

Alonso A, Martin P, Albarran C, Garcia O, Sancho M (1996) Rapid detection of sequence polymorphisms in the human mitochondrial DNA control region by polymerase chain reaction and single-stranded conformation analysis in mutation detection enhancement gels. Electrophoresis 17: 1299–1301

Anderson S, Bankeir AT, Barrel BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of human mitochondrial genome. Nature 290: 457–465

Arnason U, Xu X, Gullberg A (1996) Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences. J Mol Evol 42:145–152

Ausenda C, Chomyn A (1996) Purification of mitochondrial DNA from human cell cultures and placenta. Methods Enzymol 264:122–128

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57:248–256

Chinnery PF, Howell N, Andrews RM, Turnbull DM (1999) Mitochondrial DNA analysis: polymorphisms and pathogenicity. J Med Genet 36:505–510

de la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857–865

Finnilä S, Hassinen IE, Majamaa K (1999) Restriction fragment analysis as a source of error in detection of heteroplasmic mtDNA mutations. Mutat Res 406:109–114

Ganguly A, Prockop DJ (1995) Detection of mismatched bases in double stranded DNA by gel electrophoresis. Electrophoresis 16:1830–1835

Ganguly A, Rock MJ, Prockop DJ (1993) Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. Proc Natl Acad Sci USA 90:10325–10329

Gross AW, Aprille JR, Ernst SG (1994) Identification of human mitochondrial DNA fragments corresponding to the genes for ATPase, cytochrome C oxidase, and nine tRNAs in a denaturing gradient gel electrophoresis system. Anal Biochem 222:507–510

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. Hum Mol Genet 6: 1835–1846

Jordanova A, Kalaydjieva L, Savov A, Claustres M, Schwarz M, Estvill X, Angelicheva D, et al (1997) SSCP analysis: a blind sensitivity trial. Hum Mutat 10:65–70

Körkkö J, Annunen S, Pihlajamaa T, Prockop DJ, Ala-Kokko L (1998) Conformation sensitive gel electrophoresis for simple and accurate detection of mutations: comparison with denaturing gradient gel electrophoresis and nucleotide sequencing. Proc Natl Acad Sci USA 95:1681–1685

Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, Savontaus ML (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. Am J Hum Genet 58:1309–1322

Macaulay V, Richards M, 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

Majamaa K, Finnilä S, Turkka J, Hassinen IE (1998) Mitochondrial DNA haplogroup U as a risk factor for occipital stroke in migraine. Lancet 352:455–456

Ohlenbusch A, Wilichowski E, Hanefeld F (1998) Characterization of the mitochondrial genome in childhood multiple sclerosis l. Optic neuritis and LHON mutations. Neuropediatrics 29:175–179

Ozawa T (1995) Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. Biochim Biophys Acta 1271:177–189

Peltonen L, Pekkarinen P, Aaltonen J (1995) Messages from an isolate: lessons from the Finnish gene pool. Biol Chem Hoppe Seyler 376:697–704

Pena SD, Barreto G, Vago AR, De Marco L, Reinach FC, Dias Neto E, Simpson AJ (1994) Sequence-specific "gene signatures" can be obtained by PCR with single specific primers at low stringency. Proc Natl Acad Sci USA 91:1946–1949

Pitkänen K (1994) Suomen väestön historialliset kehityslinjat. In: Koskinen S, Martelin T, Notkola IL, Notkola V, Pitkänen K (eds) Suomen väestö. Gaudeamus, Hämeenlinna, Finland, pp 19–63

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 MB, Macaulay VA, Bandelt HJ, Sykes BC (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 ML, Aula P, et al (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Res 5:42–52

Thomas AW, Morgan R, Sweeney M, Rees A, Alcolado J (1994) The detection of mitochondrial DNA mutations using single stranded conformation polymorphism (SSCP) analysis and heteroduplex analysis. Hum Genet 94:621–623

Torroni A, Bandelt HJ, 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: 1137–1152

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 144:1835–1850

Wallace DC (1994) Mitochondrial DNA variation in human evolution and disease. Proc Natl Acad Sci USA 91: 8739–8746

Wallace DC, Stugard C, Murdock D, Schurr T, Brown MD (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. Proc Natl Acad Sci USA 95:14900–14905

Werle E, Schneider C, Renner M, Volker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Res 22:4354–4355

Zeviani M, Tiranti V, Piantadosi C (1998) Reviews in molecular medicine: mitochondrial disorders. Medicine 77:59–72