ELSEVIER

Gene 238 (1999) 211–230

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Mitochondrial DNA variation in human evolution and disease

Douglas C. Wallace *, Michael D. Brown, Marie T. Lott

Center for Molecular Medicine, Emory University School of Medicine, 1462 Clifton Rd., Atlanta, GA 30322, USA

Received 21 June 1999; accepted 6 July 1999

Abstract

Analysis of mitochondrial DNA (mtDNA) variation has permitted the reconstruction of the ancient migrations of women. This has provided evidence that our species arose in Africa about 150 000 years before present (YBP), migrated out of Africa into Asia about 60 000 to 70 000 YBP and into Europe about 40 000 to 50 000 YBP, and migrated from Asia and possibly Europe to the Americas about 20 000 to 30 000 YBP. Although much of the mtDNA variation that exists in modern populations may be selectively neutral, studies of the mildly deleterious mtDNA mutations causing Leber's hereditary optic neuropathy (LHON) have demonstrated that some continent-specific mtDNA lineages are more prone to manifest the clinical symptoms of LHON than others. Hence, all mtDNA lineages are not equal, which may provide insights into the extreme environments that were encountered by our ancient ancestor, and which may be of great importance in understanding the pathophysiology of mitochondrial disease.

Keywords: Leber's hereditary optic neuropathy; Mitochondrial disease; Mitochondrial DNA; mtDNA mutation

1. Introduction

The mitochondrial DNA (mtDNA) is a small circular genome located within the mitochondria in the cytoplasm of the cell. The mitochondrion arose as a symbiont of the proto-eukaryotic cell about 1.5 to 2 billion years before present (YBP), and subsequently gave up most of its genes to the cell's nucleus. Currently, mtDNA codes for only 13 polypeptides, all of which are essential subunits for the mitochondrial energy-generating enzymes of oxidative phosphorylation (OXPHOS). Today, each human cell contains hundreds of mitochondria and thousands of mtDNAs. Thus, mitochondrial genetics is a population genetics, both at the level of the intracellular colony of symbionts and at the level of human populations.

The mitochondria provide much of the energetic needs for our cells. Mitochondrial energy is generated by OXPHOS in which the hydrogen derived from the carbohydrates and fats in our diets is oxidized by the oxygen we breathe to give water. The released energy is stored as an electrochemical gradient across the mitochondrial inner membrane ($\Delta\psi$), which is utilized by the mitochondrial ATP synthase to generate ATP for work or is short circuited (uncoupled) to generate heat for thermal regulation.

The mtDNA is strictly maternally inherited. This is because the cytoplasmic location of the mitochondria and mtDNA dictates that the mitochondria and mtDNAs are transmitted from one generation to the next through the oocyte cytoplasm. The sperm appears to make no genetic contribution to the mtDNA.

The mtDNAs also have a very high mutation rate. As a result, the human population currently harbors a high level of population-specific mtDNA polymorphisms. Analysis of the population-specific mtDNA polymorphism has permitted the reconstruction of human pre-history. Moreover, deleterious mtDNA mutations arise frequently, and result in mitochondrial disease. Hence, analysis of maternally inherited diseases has opened a new field of human genetics, and recently demonstrated that some mtDNA diseases show a strong continental bias. In one case, this is now understood as the synergistic interaction between a pathogenic mtDNA mutation and a specific European mtDNA lineage.

Thus various mtDNA lineages are qualitatively different, and hence can be differentially acted on by selection. Analysis of this mtDNA variation may ultimately tell us something about the pathophysiology of mtDNA disease.

* Corresponding author. Tel.: +1-404-727-5624; fax: +1-404-727-3949.

0378-1119/99/$ – see front matter
PII: S0378-1119(99)00295-4

2. Mitochondrial DNA structure and genetics

The mtDNA is a 16 569 nucleotide pair (np), closed circular molecule that codes for a small (12S, MTRNR1) and large (16S, MTRNR2) rRNA as well as 22 tRNAs. These structural RNAs are used to translate the 13 polypeptides that are subunits of the OXPHOS enzyme complexes, including: seven of the approximately 43 polypeptides of complex I (MTND1, 2, 3, 4L, 4, 5, and 6), one of the 11 polypeptides of complex III (MTCYTB), three polypeptides of complex IV (MTCOI, COII, and COIII), and two polypeptides of complex V (MTATP6 and 8) (Fig. 1).

The human mtDNA is strictly maternally inherited (Giles et al., 1980; Case and Wallace, 1981), due in large measure to the fact that the mammalian egg contains about 100 000 mitochondria and mtDNAs, whereas the sperm contains only in the order of 100 mtDNAs (Michaels et al., 1982; Chen et al., 1995a). The sperm mtDNAs are contributed to the zygote at fertilization and will persist in inter-specific crosses (Gyllensten et al., 1991), but are selectively eliminated in intra-specific crosses (Kaneda et al., 1995). This has recently been correlated with the observation that the mitochondria in the sperm mid-piece is ubiquinated (Hopkin, 1999) and hence may be labeled for immediate destruction within the oocyte cytoplasm.

When a mutation arises in a cellular mtDNA, it creates a mixed intracellular population of mutant and normal molecules known as heteroplasmy. When a cell divides, it is a matter of chance as to whether the mutant mtDNAs will be partitioned into one daughter cell or another. Thus, over time, the percentage of mutant mtDNAs in different cell lineages can drift toward either pure mutant or normal (homoplasmy), a process known as replicative segregation (Wallace, 1986).

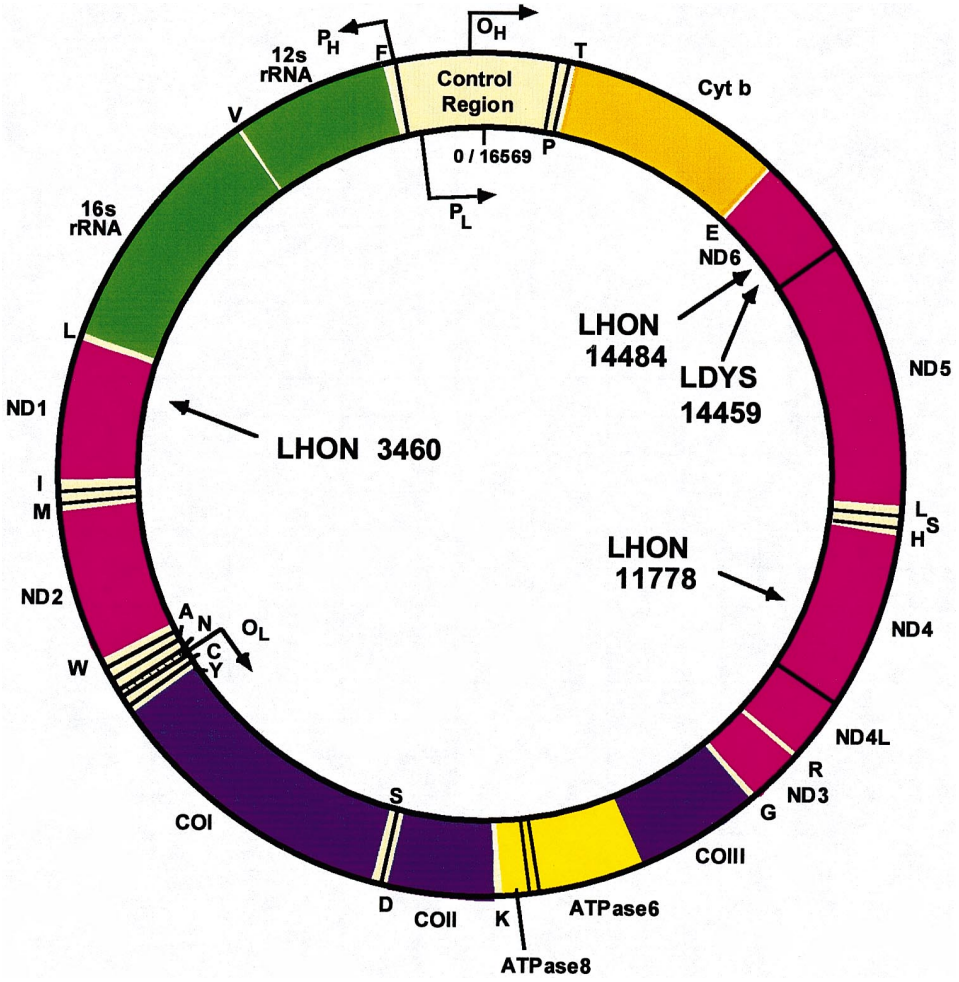

Fig. 1. The human mtDNA map, showing the location of the genes and the 'primary' Leber's hereditary optic neuropathy (LHON) mutations. The human mtDNA is a 16 569 base pair circular molecule that codes for seven (ND1, 2, 3, 4L, 4, 5, and 6) of the 43 subunits of complex I, shown in pink; one (cytochrome b, cyt b) of the 11 subunits of complex III, shown in orange; three (COI, II, and III) of the 13 subunits of complex IV, shown in purple; and two (ATPase 6 and 8) of 16 subunits of complex V, shown in yellow. It also codes for the small and large rRNAs, shown in green, and 22 tRNAs, shown in light yellow, with the adjacent letters indicating the cognate amino acids. The heavy (H)-strand origin of replication (O_H) and the H-strand and light (L)-strand promoters, P_H and P_L, are indicated in the control region.

As the percentage of deleterious mutant mtDNAs increases, cellular energy output declines. Since different tissues and organs rely on mitochondrial energy generation to different extents, this results in tissue-specific symptoms for systemic genetic defects. The differential organ sensitivities to mitochondrial defects, in decreasing order, are the central nervous system including the eye, the heart and skeletal muscle, renal system, endocrine system and liver (Wallace, 1987, 1995; Wallace et al., 1988a, 1994; Shoffner et al., 1990).

The mtDNA has a very high sequence evolution (Brown et al., 1979), in the order of 10 to 17 times faster than nuclear DNA genes of similar function (Neckelmann et al., 1987; Wallace et al., 1987). This has resulted in the accumulation of a broad spectrum of mtDNA sequence polymorphisms in human populations, but also may be a common source of mutations causing mitochondrial disease.

3. MtDNA variation in human populations

The mtDNA sequence evolution rate is in the same time frame as the origin and radiation of human continental populations. As a result, mtDNA polymorphisms have accumulated sequentially as women migrated west of Africa and into the various continents. Generally, for mtDNA variants to reach polymorphic frequencies, they need to be selectively neutral or near neutral to avoid being eliminated by selection and thus become prevalent through genetic drift. In fact, the rapid shift in mtDNA lineages that is observed between continent may have been influenced by selection as well as by drift.

3.1. World mtDNA phylogeny and the origin of women

Because the mtDNA is strictly maternally inherited, the mtDNA sequence has evolved by the sequential accumulation of base substitutions along radiating maternal lineages. Thus, as women migrated out of Africa into the different continents about 150 000 YBP they accumulated mtDNA mutations that today are seen as high frequency, continent-specific mtDNA sequence polymorphisms. These polymorphisms are associated with specific mtDNA haplotypes, and groups of related haplotypes (haplogroups) (Torroni and Wallace, 1994; Wallace, 1995).

The first clear evidence that mtDNA variation correlated with the ethnic and geographic origin of the individual came from our survey of *HpaI* RFLPs in African, Asian, and European–American mtDNAs. This revealed that, in Africans, 96% of Pygmies, 93% of San Bushman, and 71% of Bantus harbored an *HpaI* restriction site at np 3592 not seen in Asians or Europeans. By convention, all polymorphic restriction sites, including the *HpaI* np 3592 site, are defined by the 5′ end of the recognition sequence, but the polymorphic nucleotide may be different. For example, the *HpaI* np 3592 site change is caused by a C-to-T transition at np 3594. In contrast to African mtDNAs, about 13% of Asians lacked an *HpaI* restriction site at np 12 406 (G to A at np 12 406), which was present in all other mtDNAs (Denaro et al., 1981). A further survey of the mtDNA variation detected using six highly informative restriction enzymes (*HpaI*, *BamHI*, *HaeII*, *MspI*, and *AvaII*) and Southern blotting confirmed that mtDNA variation was high and correlated strongly with the geographic origin of the individual. It also showed that all mtDNAs were part of a single phylogenetic tree, that the greatest variation was in Africa, and that the tree was about 100 000 years old (Johnson et al., 1983). Extensive studies by our group, as well as by others, ultimately led to the characterization of 3065 mtDNAs from 62 geographic samples using these six enzymes. This revealed 149 haplotypes and 81 polymorphic sites. This analysis confirmed: (1) that the mtDNA polymorphisms within each mtDNA were virtually in total linkage disequilibrium, consistent with a low frequency of recombination; that mtDNA variation correlated highly with the ethnic and geographic origin of the individual; (2) that there was a single mtDNA tree; (3) that the greatest variation and deepest root of the tree was in Africa, consistent with an African origin of humans. The extent of mtDNA sequence differences between continental populations was estimated from this data by calculating the GST statistic. For the mtDNA, the GST was 0.35 ± 0.025, implying that about 35% of the mtDNA variation was continent-specific. By contrast, the comparable nDNA value was 0.12. Hence, the mtDNA encompasses much greater continent-specific sequence diversity than the nDNA (Merriwether et al., 1991).

A recent African origin of human mtDNAs was also demonstrated by the investigations of Cann et al. (1987). These investigators purified the individual mtDNAs from cells or tissues, digested the DNA with 12 restriction endonucleases (*Hpal*, *AvaII*, *FnuDII*, *HhaI*, *HpaII*, *Mbol*, *TaqI*, *Rsal*, *HinfI*, *HaeIII*, *AluI*, and *DdeI*), end-labeled the fragments, and resolved the fragments using polyacrylamide gels and autoradiography (Brown, 1980). A survey of 147 mtDNAs, including 34 Asians, 21 Australian aboriginals, 26 aboriginal New Guineans, 46 Caucasians, and 20 Africans (18 of whom were Black Americans), also revealed that there was a single mtDNA tree, that the deepest root occurred in Africa, and that Africa harbored the greatest sequence diversity. Hence, Africa is the origin of *Homo sapiens*. Using an estimated sequence evolution rate of 2–4% per million years (MYR), the human mtDNA tree was calculated to be about 200 000 years old (Cann et al., 1987).

This analysis was extended to include 62 Japanese (Horai and Matsunaga, 1986) and 119 Papua New

Guineans (Stoneking et al., 1990). The Papua New Guineans were sampled from 25 localities, and significant differences in mtDNA variation were found between the highland and coastal populations. Combining the Papua New Guinea data with the previous European, Asian, and African data permitted calculation of a global GST of 0.31 (Stoneking et al., 1990), a value similar to that found for the six-enzyme analysis discussed above.

The African origin of mtDNA variation was also supported through sequence analyses of the 1121 np non-coding control region of the mtDNA. This region has a three- to four-fold greater sequence diversity than the coding region. Analysis of the control-region sequences from 189 individuals, 121 of whom were native African, once again confirmed that the greatest sequence diversity was in Africans, that the deepest root was between Africans, and that the coalescence time of the mtDNA tree (phylogeny) was between 166 000 and 249 000 YBP (Vigilant et al., 1991). The African root of this phylogeny was subsequently challenged on the basis that multiple equally probable parsimony trees could be generated from the data (Templeton, 1992). However, other phylogenetic analysis procedures, such as neighbor-joining trees, have reaffirmed the cohesiveness of the deepest African associations and thus support the African origin of the mtDNA phylogeny (Hedges et al., 1991).

Analysis of the control-region sequence of 95 individuals, including 61 Japanese, confirmed that the greatest diversity and deepest root occurred in Africa and revealed that 'Mongoloid' mtDNAs were subdivided into two distinct groups (Horai and Hayasaka, 1990). Analysis of 117 Caucasian mtDNAs confirmed the distinctive nature of many European mtDNAs and revealed that the various mtDNA lineages were widely disseminated throughout Europe (Di Rienzo and Wilson, 1991).

Finally, comparison of the original European mtDNA sequence (Anderson et al., 1981) with that from an African, a Japanese, and four African apes (common and pygmy chimpanzees, gorilla, and orangutan) revealed that the European and Japanese mtDNAs were most similar, that the African mtDNA was more divergent, and that the nearest ape relatives, the chimpanzees, were 10 times more divergent from humans than Africans are from Asians and Europeans. Using the orangutan–African ape divergence time of 13 million YBP as reference, this study gave an age for human mtDNA radiation of $143\,000 \pm 18\,000$ YBP and a time for European and Japanese radiation of $70\,000 \pm 13\,000$ YBP (Horai et al., 1995).

All of these studies lead to the same conclusion. The human mtDNA tree appears to have originated in Africa about 150 000 YBP. As women migrated from Africa to colonize new lands, additional mtDNA mutations arose and became established by genetic drift, resulting in continent-specific mtDNA variation. Today, these population-specific polymorphisms constitute the background on which potentially pathogenic mtDNA mutations must be identified.

3.2. Cataloging continent-specific mtDNA variation

Although the above methods permitted elucidation of the general features of human mtDNA evolution, a more detailed analysis of mtDNA variation has been necessary for clinical studies and for addressing additional anthropological questions on the age and origin of Africans, Europeans, Asians, and Native Americans. To increase the sensitivity of our analyses, we developed a new mtDNA analysis procedure — high resolution RFLP analysis — in which the mtDNAs from a variety of human samples could be amplified by using PCR in nine overlapping fragments. Each fragment was then digested with 14 restriction endonucleases (*AluI*, *AvaII*, *BamHI*, *Ddel*, *HaeII*, *HhaI*, *HinfI*, *HincII*, *HpaI*, *Hpal*, *MspI*, *Mbol*, *RsaI*, and *TaqI*), and the fragments resolved on agarose gels and detected by ethidium bromide staining and UV fluorescence. This procedure surveys $>20\%$ of the mtDNA sequence, and the aggregate of the restriction-site polymorphisms for each mtDNA is used to define the mtDNA haplotype (Ballinger et al., 1992; Torroni et al., 1992). The regional PCR fragments can also be sequenced, permitting extension of the analysis to areas of interest such as the hypervariable control region (Torroni et al., 1993a,b). The sequence differences between mtDNAs can then be compared by using various phylogenetic procedures including parsimony, neighbor-joining, and unweighted pair-group analyses. These phylogenetic trees reveal the relatedness of the mtDNAs, with the more similar mtDNAs clustering together. The extent of sequence diversity within or between groups of related haplotypes (haplogroup) can also be calculated (Tateno et al., 1982; Nei and Tajima, 1983; Saitou and Nei, 1987; Swofford, 1993).

3.3. African mtDNA variation

To characterize better African mtDNA variation, we have surveyed, using high-resolution RFLP analysis, the mtDNAs from 214 Africans: 101 from Senegal (60 Mandenkalu, 20 Wolof, 8 Pular, 13 others from eight tribes); 22 Mbuti (Eastern) Pygmies from Zaire and 17 Biaka (Western) Pygmies from Central African Republic; 74 South Africans, including 43 Kung and 31 Khwe (Chen et al., 1995b, 1999). This survey revealed 105 haplotypes defined by greater than 157 polymorphic sites. Phylogenetic analysis revealed that 75 of the haplotypes formed a single, coherent, African-specific haplogroup designated 'L' (Fig. 2), which is defined by

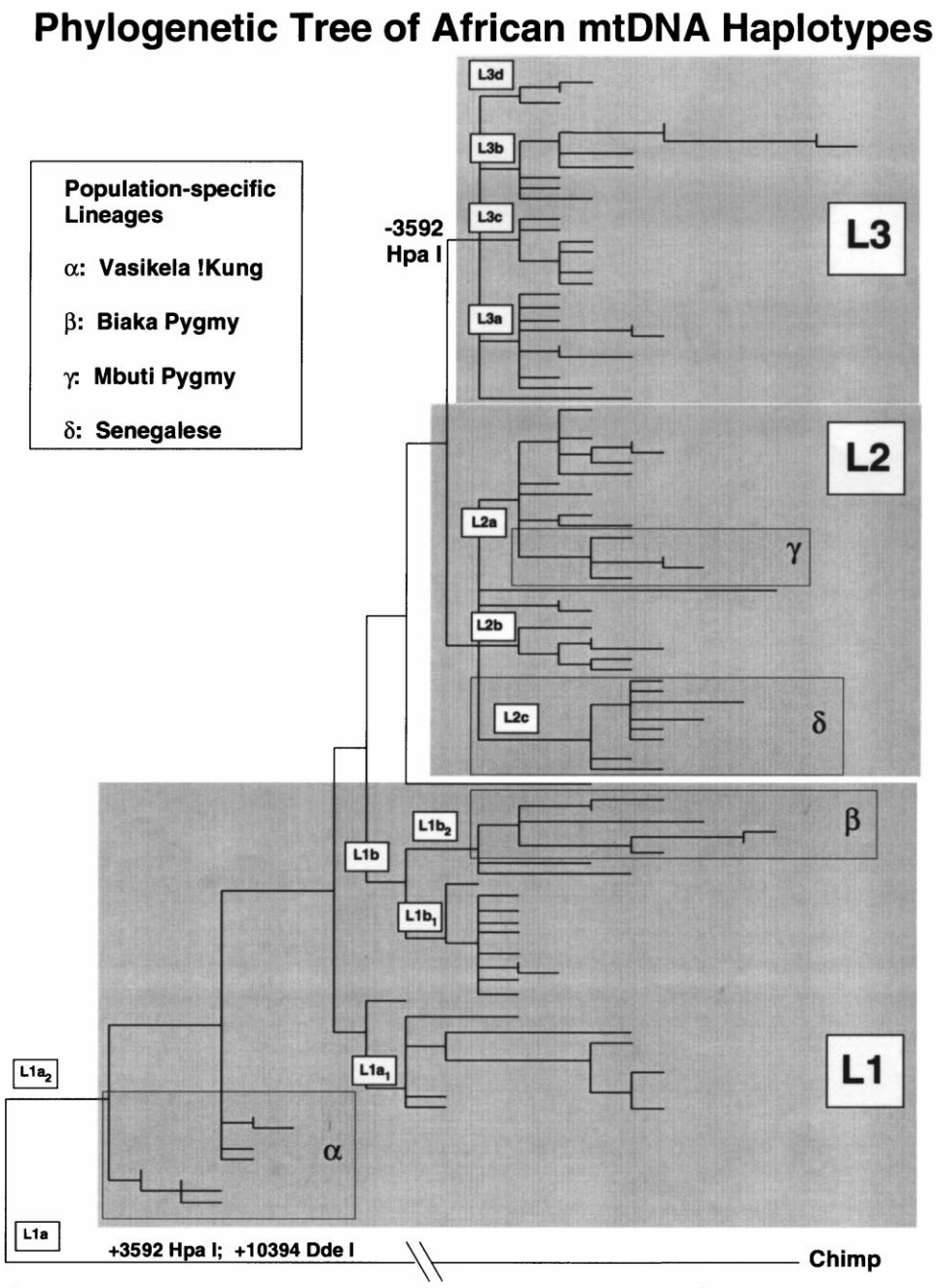

Fig. 2. Phylogeny of African mtDNA haplotypes. This tree summarizes the results from 214 Africans, including 101 Senegalese, 22 Mbuti Pygmies, 17 Biaka Pygmies, 43 Kung and 31 Khwe, encompassing 105 haplotypes. Each haplotype is positioned relative to the others based on number of genetic similarities or differences. The closer together two haplotypes are in the tree, the more similar is their sequence. MtDNA polymorphisms that arose recently are present in only one or a few mtDNAs and define the terminal twigs of the tree. By contrast, polymorphisms that arose early in human mtDNA evolution are shared by larger groups of related mtDNAs and define major branches of the tree. These groups of related mtDNA haplotypes are called haplogroups. To identify the oldest mtDNA haplogroups, the human mtDNA haplotypes are compared with that of a more distantly related mtDNA, the outgroup, in this case chimpanzee mtDNA. This permits the 'rooting' of the tree. For African mtDNAs, one of the oldest and most significant polymorphisms is the *HpaI* site at np 3592. This site is African-specific with the oldest 2/3 of the tree having the *HpaI* 3592 site (+3592 *HpaI*) and being defined as macro-haplogroup L, with haplogroup subdivisions L1 and L2. The absence of this site is defined as haplogroup L3. Haplogroup L3 forms the bridge between African mtDNAs and European and Asian mtDNAs. Within the L1, L2, and L3 divisions are multiple haplogroups and sub-haplogroups, some of which correlate highly with the population sampled. Thus a core block of related haplotypes in L1a, designated α, are specific for the Kung; a second block in $L1b_2$, labeled β, are specific for the Biaka Pygmies; a third block in L2a, designated γ, are specific for Mbuti Pygmies; a fourth block in L2c, labeled δ, are specific for the Senegalese. This implies that these core haplotypes arose with these populations. The fact that the α block of the Kung and the β block the Biaka Pygmies are closest to the root of the tree implies that these are among the oldest African populations. Although the deep branches of this tree are very robust, the most distal twigs are subject to greater interpretive fluctuation. Hence, this tree is one of the thousands that could be drawn in which the major branches are retained but the precise branching order of the individual haplotypes can vary. This is standard. This tree is reproduced from Chen et al. (1999) with permission.

the African-specific *HpaI* site at np 3592 together with the *DdeI* site at np 10 394 (A to G at np 10 398). This lineage is subdivided into two sublineages, L1 and L2. L1 encompasses 52% of the L haplotypes and 29% of all African mtDNAs and is defined by an additional *HinfI* site at np 10 806 (T to C at np 10 810). L2 encompasses 48% of the L haplotypes and 34% of the African mtDNAs and is defined by an additional combined *HinfI* site gain at np 16 389 and *AvaII* site loss at np 16 390 (G to A at np 16 390). All *HpaI* np 3592 positive haplotypes are of African origin, with the only exceptions occurring in populations known from historical evidence to have had Africa contact.

Several other features of haplogroup L are of interest. Two length mutations have been observed in L1: a nine np COII/tRNA^{Lys} deletion between nps 8272 and 8289 (Cann and Wilson, 1983; Wrischnik et al., 1987) found in two African haplotypes, AFR 60 (representing 27% of Mbuti Pygmies) and AFR 61 (representing 24% of Biaka Pygmies); and a 10–12 bp insertion of cytosines (Cs) between the tRNA^{Tyr} and COI gene (nps 5895–5899) in Biaka Pygmies with the ARF66 haplotype (Fig. 2).

The remaining African mtDNAs form a heterogeneous array of four lineages, designated haplogroup L3, each defined by specific restriction-site gains or losses (Watson et al., 1996). One of these lineages is defined by loss of the *DdeI* site at np 10 394. This lineage represents only a few percent of the African mtDNAs, yet it appears to be the progenitor of roughly half of all European, Asian, and Native American mtDNAs. Within this lineage are mtDNAs that also lack an *HinfI* site at np 12 308. This mtDNA haplotype is closely related to the European-specific haplogroup H.

Analysis of African mtDNA control region sequences reveals many of the same population subdivisions. However, in some cases the control region sequences subdivide haplotypes, and in others the haplotype markers subdivide control region sequence groups (Chen et al., 1999).

Analyses of the population distribution of the African haplotypes revealed that each of the four primary populations studied (the Senegalese of West Africa, the Mbuti Pygmies, the Biaka Pygmies and the Vasikela Kung) has a distinctive set of related core haplotypes that are specific for that population (Fig. 2). The core haplotypes of the Vasikela Kung of South Africa, designated α, occur in haplogroup L1 and are defined by *MspI* site losses at np 8112 and 8150, and *AvaII* site gain at np 8249, and an *HaeIII* site loss at np 8250. This cluster of Vasikela Kung haplotypes is at the deepest root of the African phylogeny, suggesting that the Kung are one of the oldest populations. The Biaka Pygmy core haplotypes, designated β, also reside in haplogroup L1 and are defined by an *AluI* site gain at np 10 319. The Mbuti Pygmies, designated γ, reside in haplogroup L2 and are delineated by a *DdeI* site loss at np 13 065, and an *RsaI* site gain at np 11 776. Finally, the Bantu-derived Senegalese core haplotypes also belong to haplogroup L2 and are defined by an *HaeIII* site loss at np 322, a *DdeI* site loss at np 679, and an *HaeIII* site loss at np 13 957. Calculation of the sequence divergence of the core haplotypes for each population reveals that the Vasekela Kung α lineage and Biaka Pygmy β lineages are the oldest, whereas the Mbuti Pygmy and Senegalese lineages are much younger. Hence, the Kung and Biaka Pygmies are more representative of the proto-Africans and the Biaka and Mbuti Pygmies may have had independent origins.

Calculation of the accumulated sequence diversity of the African-specific haplogroup L and its subhaplogroups L1 and L2 gave values of 0.356%, 0.328%, and 0.171% respectively. The total African mtDNA sequence diversity was 0.364%. This means that haplogroup L has the highest sequence diversity of any continent-specific haplogroup and that Africa encompasses the greatest diversity of any continent. Using our estimate of the mtDNA sequence evolution rate of 2.2–2.9%/MYR (Torroni et al., 1994a), the L haplogroup is between 123 000 and 162 000 years old, and the total African mtDNA lineage is between 126 000 and 166 000 years old (Fig. 2).

Calculation of the intragroup and intergroup sequence variations of the various African groups strengthen these observations. The Biaka Pygmies and the Vakekela Kung had the greatest intragroup sequence variation: 0.342% and 0.320% respectively, comparable to that of Africa as a whole. This further supports the conclusion that these are among the oldest populations. The Mbuti Pygmies and Senegalese have intragroup sequence diversities in the range of 0.241 to 0.277, confirming that these populations are younger. A neighbor-joining tree analysis clearly separates the Biaka Pygmies from the Mbuti Pygmies, and places the Mbuti Pygmies on the same side of the tree as the Bantu-derived populations. Hence, the Mbuti and Biaka pygmies probably are distinct populations (Chen et al., 1999).

Overall, the mtDNA data show that African mtDNAs are distinct, they are the oldest with the greatest diversity and deepest root, that the Vasikela Kung and Biaka Pygmies are among the original populations of Africa, and that the 'Pygmy' morphology arose two independent times in Africa.

3.4. European mtDNA variation

European mtDNA sequence variation has been defined by the analysis of 259 samples from individuals of European ancestry living in the United States, Canada, Finland, Italy, and Sweden (Torroni et al., 1994b, 1996a). Restriction analysis revealed 178 poly-

morphic sites that define in the order of 170 haplotypes (Torroni et al., 1994b, 1996a). Phylogenetic analysis showed that all European mtDNAs could be subdivided into two groups by the presence (1/4) or absence (3/4) of the *Ddel* site at np 10 394 (Fig. 3). Thus, Europeans exhibit a marked increase in the proportion of −10 394 *Dde*I mtDNAs over the 4% seen in Africans.

In addition to the macro-subdivision of European mtDNAs by the *Ddel* site at np 10 394, nine distinct European mtDNA haplogroups have also been observed. Those lacking the *Dde*I site at np 10 394 are haplogroups H, T, U, V, W, and X; those retaining the *Dde*I np 10 394 site are I, J, and K (Fig. 3).

Of the haplogroups that lack the *DdeI* np 10 394 site, haplogroup H also lacks an *AluI* site at np 7025 (C to T at np 7028). This haplogroup encompasses 40.5% of European mtDNAs. Haplogroup T is defined by the presence of a *BamHI* site at np 13 366 and an *AluI* site

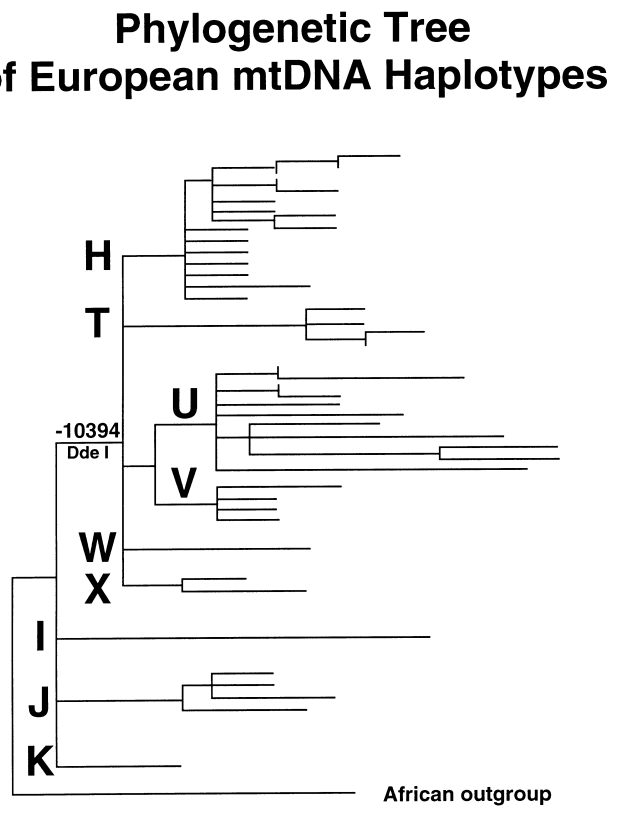

Fig. 3. Phylogeny of European mtDNA haplotypes. This representative tree encompasses the analysis of 86 samples from individuals of European ancestry. The tree is bifurcated by the presence or absence of an ancient polymorphism, the *DdeI* site at np 10 394. European mtDNAs lacking the 10 394 site fall into six distinct haplogroups: H, T, U, V, W, and X; those retaining the site fall into three major haplogroups: I, J, and K. Each haplogroup is defined by distinctive polymorphisms (see text). These nine European-specific haplogroups account for over 98% of all mtDNAs found in Europe (Torroni et al., 1994a,b). As for all parsimony trees, there are many alternative European mtDNA phylogenies, but the central branches are quite robust.

at np 15 606 and accounts for 15.2% of European mtDNAs. Haplogroup U is defined by the presence of an *HinfI* site at np 12 308 and accounts for 14.7% of European mtDNAs. Haplogroup V is delineated by the loss of an *NlaIII* site at np 4577 and is found in 4.8% of Europeans, whereas haplogroup X is, in part, defined by the loss of a *DdeI* site at np 1715 and is found in about 6.9% of European mtDNAs. Of the haplotypes that retain the *DdeI* site, haplogroup I is defined by the loss of the *DdeI* np 1715 site, and the gain of an *AvaII* site at 8249 and an *AluI* site at np 10 028 and represents 6.7% of European mtDNAs; haplogroup J is identified by the loss of a *BstNI* site at np 13 708 and is found in about 11.3% of European mtDNAs; and haplogroup K is delineated by the loss of an *HaeII* site at np 9052 and the gain of an *HinfI* site at np 12 308 and is found in 9.1% of Europeans (Fig. 3). The sequence diversity of haplogroup H is 0.065%, giving an age of this lineage of 22 000 to 30 000 YBP. However, the sequence divergence of haplogroup U, which is shared between Europeans and African Bantu, is 0.148%, giving an age of 51 000 to 67 000 YBP. Hence, haplogroup U may represent one of the founder lineages of Europe (Torroni et al., 1996a). The overall sequence divergence between the two major branches of the European phylogeny is 0.113%, giving an age for the colonization of Europe of between 39 000 and 51 000 YBP (Torroni et al., 1994b).

Analysis of the control regions of the European mtDNAs revealed additional continent-specific markers (Di Rienzo and Wilson, 1991). However, one control-region mutation in haplogroup I proved to be totally novel, with potential implications for the evolution of the human mtDNAs. All haplogroup I mtDNAs were found to have a homoplasmic insertion of two to six Cs within a cluster of Cs in the sequence ACCCCCC (Box 2), where the A is located at np 567. This germ-line mutation increases the homology between this sequence and the nearby control-region sequence ACGCCCCC-TCCCCCGCT (Box 1), where the A is located at np 302. Because of this homology, every individual who inherits the Box 2 germ-line insertion mutation becomes prone to undergo a somatic mutation during development. In this somatic mutation, the region between Boxes 1 and 2 is duplicated as a 270 np direct repeat, possibly through slipped misreplication (Torroni et al., 1994b). The somatic 270 bp duplication duplicates the H-strand promoter (nps 545–567), the L-strand promoter (nps 392–445), the two intervening mitochondrial transcription factor binding sites (nps 418–445 and nps 523–550), CSBIII (nps 346–363), part of CSBII (nps 299–315), and the putative replication primer processing site (nps 317–321) (Brockington et al., 1993; Torroni et al., 1994b). This raises the possibility that the duplicated molecules are transcribed twice as frequently and may be preferentially replicated, providing selective advantage for this mutation. This may explain why the

unstable Box 2 insertion has been maintained throughout the 34 000 year history of haplogroup I (Torroni et al., 1994b).

3.5. Asian mtDNA variation

To define further the Asian mtDNA sequence variation, we have analyzed the mtDNAs from 153 Central and Southeast Asians, including aboriginal Malays and Orang Ash, aboriginal Borneans, Han Chinese, Vietnamese, Koreans, and Malaysian Indians (Ballinger et al., 1992) as well as 54 Tibetans (Torroni et al., 1994c) and 758 Siberians from 11 aboriginal populations, including the Chukchi and Koryaks from northeastern-most Siberia (Torroni et al., 1993b; Starikovskaya et al., 1998; Schurr et al., 1999). A representative Asian phylogenetic tree encompassing 42 Tibetan haplotypes, 106 Asian haplotypes, and 34 Siberian haplotypes is presented in Fig. 4. This phylogeny shows that all Asian mtDNAs can be subdivided into two macro-haplogroups defined by the presence or absence of the polymorphic site at *DdeI* at np 10 394, which also bifurcates the European mtDNA lineages. Moreover, every Asian mtDNA that harbors the *Ddel* site at np 10 394 also has an adjacent *AluI* site at np 10 397 (C to T at np 10 400). The macro-haplogroup defined by the presence of the *DdeI* np 10 394 and the *AluI* np 10 397 sites has been designated macro-haplogroup M [or also as (+/+)]. The constant association of the *DdeI* np 10 394 and *AluI* np 10 394 in Asians, but not in Africans or Europeans, implies that the *AluI* np 10 397 mutation must have arisen on an mtDNA carrying the *DdeI* np 10 394 mutation as women migrated out of Africa and into Asia (Fig. 4).

In addition to this major bifurcation of Asian mtDNAs, there are a number of distinctive sublineages of relevance to Asian and Native American prehistory. Haplogroups A, B, C, and D have proved to be the progenitors of virtually all Native American mtDNAs. Haplogroups A and B lack both the *DdeI* site at np 10 394 and the *AluI* site at np 10 397, whereas haplogroups C and D have these sites. In addition, haplogroup A is defined by an *HaeIII* site at np 663 (A to G at np 663), haplogroup B by an independent occurrence of the 9 np deletion between the COII and $tRNA^{Lys}$ genes, haplogroup C by the simultaneous *HincII* site loss at np 13 259 and an *AluI* site gain at np 13 262 (A to G at np 13 262), and haplogroup D by the loss of an *AluI* site at np 5176 (C to A at np 5178). These haplogroups are further delineated in most Asians and Native Americans by specific control-region variants. For haplogroup A, these include variants at nps 16 362 (T to C), 16 319 (G to A), 16 290 (C to T), and 16 223 (C to T); for haplogroup B, variants at nps 16 217 (T to C) and 16 189 (T to C); for haplogroup C, variants at nps 16 327 (C to T), 16 298 (T to C), and 16 223 (C to T); and, for haplogroup D, variants at nps 16 362 (T to C) and 16 223 (C to T) (Torroni et al., 1993a).

Three other prominent Asia haplogroups are E, F, and G. Haplogroups E and G have the combined *DdeI* and *AluI* sites at nps 10 394 and 10 397, whereas haplogroup F lacks these sites. Haplogroup E is further defined by an *HpaI* site loss at np 7598, haplogroup G by the presence of an *HaeIII* site at np 4830 and an *HpaI* site at np 4831; haplogroup F is delineated by the combined *HpaI*/*HincII* site loss at np 12 406, the first Asian-specific polymorphism observed (Denaro et al., 1981; Blanc et al., 1983). All of these haplogroups show marked frequency variation throughout Asia. Haplogroup F is prominent in southern Asian populations, being found in 32% of Vietnamese mtDNAs and 21% of Malay mtDNAs. It is present in about 15% of Koreans and Tibetans, but is virtually absent in Siberia. By contrast, haplogroups A, C, D, E, and G are absent in southern Asian populations, including Vietnamese, Malays, Sabah, Malay aboriginals, and New Guineans, but these groups are found at significant frequencies in Tibetans, Koreans, and Han Chinese. This north–south distinction supports the dichotomization of Asians into the Sinodont (northern) and Sunodont (southern) Asian populations (Turner, 1983, 1987). Furthermore, haplogroups A, C, and D extend into the Siberia populations analyzed, reaching maximum frequencies of 68%, 84% and 28% respectively. Haplogroup A reaches its highest frequencies in the Chukchi and Koryaks, the northeastern-most populations of Siberia and likely progenitors of Native Americans. The haplogroup frequencies of the Koryaks are 5% A, 36% C, 1% D, and 42% G, 10% Y, and 6% other, whereas those of the Chukchi are 68% A, 11% C, 12% D, and 9% G (Torroni et al., 1993b; Starikovskaya et al., 1998; Schurr et al., 1999).

Haplogroup B, defined by the 9 np COII-$tRNA^{Lys}$ deletion, displays a markedly different distribution. It is common throughout central and southern Asia and is prominent in coastal Asian populations, approaching fixation (100%) in certain Pacific island populations (Hertzberg et al., 1989; Stoneking et al., 1990; Ballinger et al., 1992). It is virtually absent from all nine Siberian populations analyzed, yet it reappears throughout North, Central, and South American Native American populations (Schurr et al., 1990; Torroni et al., 1993a,b). The high frequency of this haplogroup among coastal Asian and Pacific island populations, and its striking absence in Siberians relative to Central Asians and Native Americans, raises the possibility that haplogroup B mtDNAs did not come to the Americas via a trans-Siberian migration, but rather may have crossed from Asia to the Americas by migration along the Siberian coast. This deduction has been questioned, however, based on mathematical analysis of control-region sequence diversity (Forster et al., 1996).

The co-occurrence of the *AluI* site at np 10 397 and

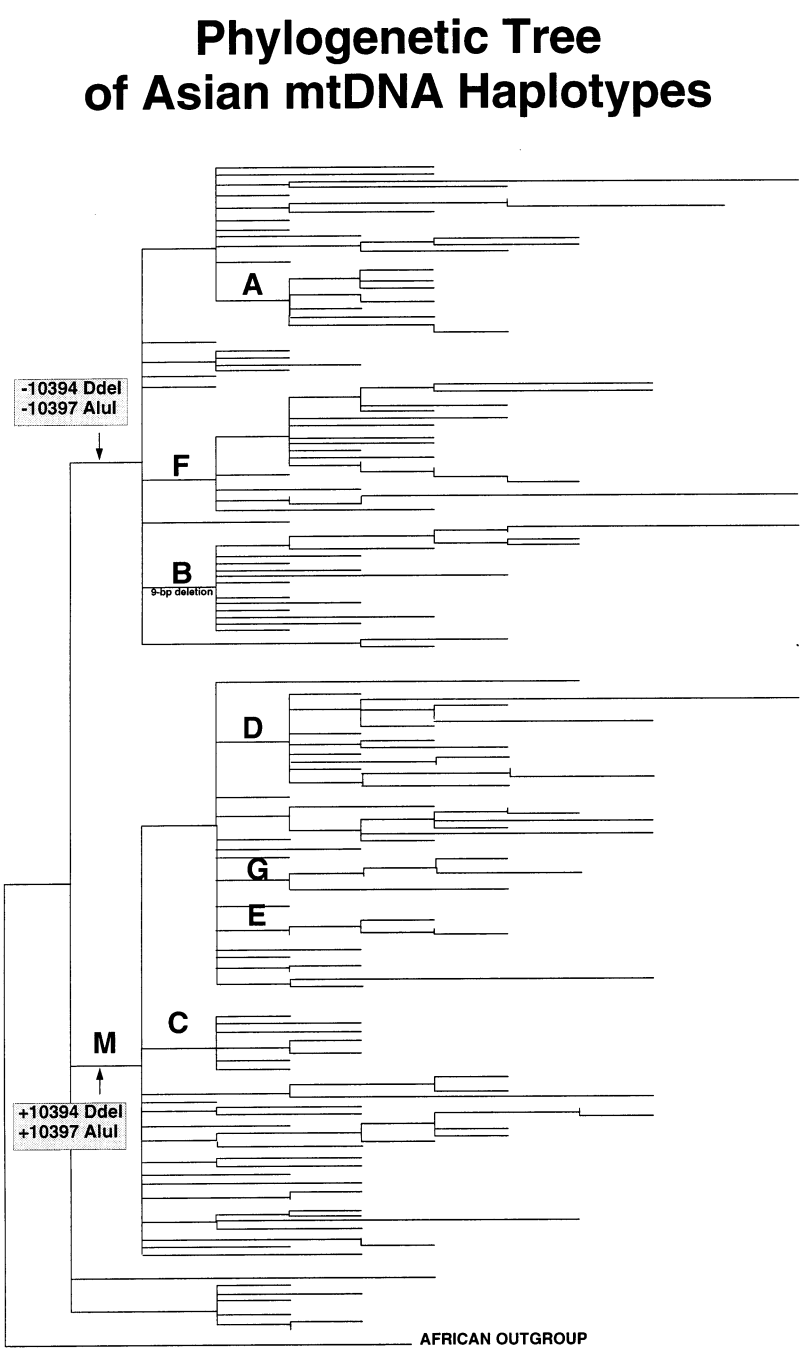

Fig. 4. Phylogeny of Asian mtDNA Haplotypes. This representative tree is based on data from 153 Central and Southeast Asians, 54 Tibetans, and 758 aboriginal Siberians. The Asian mtDNA phylogeny is bifurcated into two macro-haplogroups. One macro-haplogroup lacks the *DdeI* site at np 10 394, as well lacking an *AluI* site at np 10 397, and thus is designated (−/−), whereas the other macro-haplogroup has these restriction sites and is thus designated (+/+). The later group is called macro-haplogroup 'M'. Within these two major Asian mtDNA lineages are multiple important haplogroups. Haplogroup F is (−/−) and is found at high frequency in Southeast Asia, but declines toward Northeast Asia. Haplogroups A and B are −/−, whereas haplogroups C and D are (+/+). These four mtDNA lineages are at a low frequency in Southern Asia, but rise to high frequencies in Northeast Asia, where they participated in the peopling of the Americas. This tree is reprinted from Torroni et al. (1994c) with permission.

the *DdeI* site at np 10 394 in macro-haplogroup M mtDNAs throughout Asia indicates that the *AluI* site gain occurred at the beginning of Asian habitation. This hypothesis has been supported by the discovery of macro-haplogroup M mtDNAs in East African Ethiopians. Either the haplogroup M mtDNA entered East Africa through relatively recent migrations from Asia, or they originated in East Africa. If the latter is true, this would place the origin of M at close to the time of the migration of proto-Asians out of Africa and

into Asia (Passarino et al., 1996, 1998). Consequently, the sequence diversity that has accumulated in the *DdeI* np 10 394 + *AluI* np 10 397 lineage should be indicative of the age of the Asian population. The overall sequence diversity in this lineage is 0.161%. This gives an age for the Asian population of 56 000–73 000 YBP.

3.6. Native American mtDNA variation

To learn more about the origin of Native Americans, we analyzed 743 Native American mtDNAs. Multiple hypotheses have been put forward to explain the origin and radiation of Native Americans. One hypotheses is based on the classification of Native American languages by Greenberg et al. (1986). These authors divided all Native American languages into three major groups: Amerind, which encompasses the great diversity of languages spoken by the Paleo-Indian peoples occupying most of North America and all of Central and South America; Na-Déné, which is spoken by the Athapaskans of the northwestern United States, Canada, and Alaska, as well as by the Navajo and Apache, who migrated south through the great plains around 1000 AD; the Eskaleut languages, which are spoken by the Eskimos and Aleuts of the Arctic region. Greenberg et al. (1986) hypothesized that each of these language groups corresponded to a different migration, arising in a different geographic homeland. Using glottochronology-dating based on the divergence rate of languages, they estimated that these migrations occurred at about 11 000 YBP, 9000 YBP, and 5000 YBP respectively.

In our first studies on Native American mtDNA variation, we focused on the Pima and the Papago, Paleo-Indians of the southwestern United States. Using Southern blot analysis and our initial six informative restriction endonucleases, we discovered that about 40% of these Native American mtDNAs lacked the *HincII* site at np 13 259 (Wallace et al., 1985), whereas only 1.8% of central Asian mtDNAs lacked this site (Blanc et al., 1983). This led to the hypothesis that Native American mtDNAs were derived from a limited number of founding mtDNA haplotypes that crossed the Bering land bridge in distinct migrations (Wallace et al., 1985; Schurr et al., 1990; Wallace and Torroni, 1992). This hypothesis has subsequently been confirmed by our more extensive analysis encompassing 563 Paleo-Indians from 24 tribes, 130 Na-Déné representing five tribes, and 50 Eskimos (Torroni et al., 1992, 1993a, 1994a,d).

Our analysis of mtDNA variation in Paleo-Indians revealed a dramatic result. Virtually all of the mtDNAs fell into one of the four Asian haplogroups: A (*HaeIII* site at np 663), B (9 np deletion between COII/$tRNA^{Lys}$), C (*HincII* site loss at np 13 259 and *AluI* site gain at np 13 262), and D (*AluI* site loss at np 5176), with haplogroups C and D also belonging to macro-haplogroup M and thus harboring the *DdeI* np 10 394 and *AluI* np 10 397 site gains found in their Asian progenitors (Schurr et al., 1990; Torroni et al., 1992, 1994d) (Fig. 5). Each of the four primary Native American haplogroups traces back to a single nodal mtDNA haplotype that is shared by Asia and the Americas and that initiated the mtDNA radiation in the Americas. However, none of the derived haplotypes is shared by Asians and Americans, as demonstrated by analyzing haplogroup C and D mtDNAs from Siberians and Native Americans (Torroni et al., 1993b). Hence, it appears that, primarily, these four mtDNA haplotypes crossed from Siberia into the Americas. All four of these haplogroups are distributed throughout the Paleo-Indians of North, Central, and South America, though individual tribes may have lost one or more of the haplogroups through genetic drift. The broad distribution of all four mtDNA haplogroups suggests that they either came together or were subsequently thoroughly mixed.

The calculation of the mtDNA sequence diversity that has accumulated within each haplogroup revealed that haplogroups A, C, and D had relatively similar values: A = 0.075%, C = 0.096%, and D = 0.053%, with a mean value of 0.075%. By contrast, haplogroup B had a much lower value, 0.034%, suggesting that haplogroup B arrived in the Americas much more recently than A, C, and D (Schurr et al., 1999). This difference is consistent with the absence of haplogroup B in Siberia, even though haplogroups A, C, and D are prevalent. These two results imply that the Paleo-Indians of the Amerind linguistic group may have been derived from two migrations. The first migration moved up from central Asia through Siberia, during which it became progressively enriched for the founder haplotypes of haplogroups A, C, and D. Ultimately, only these haplotypes crossed the Bering land bridge to found the Paleo-Indians. The second migration came much later, bearing the founder haplotype of haplogroup B. This migration bypassed Siberia, possibly moving along the Siberian and Alaskan coasts, and entered the Americas, where it interspersed with the already present haplogroups A, C, and D (Schurr et al., 1990; Torroni et al., 1992, 1993a) (Fig. 5). An independent analysis of the number of Native American migration based on mtDNA control region sequence data concluded that all Paleo-Indians (Amerinds) were derived from a single migration. That is, B came with A, C, and D (Forster et al., 1996).

To investigate further the Paleo-Indian tribalization process, we examined the mtDNAs from an isolated groups of Aztec descendants in central Mexico: the Mixtec of Alta and Baja, the Zapotecs, and the adjacent Mixe. These tribes were compared with the Pima of Arizona, the Maya of Yucatan, the Chibchan-speakers of Panama, the Bella Coola and the Nuu-Chah-Nulth of North America, and the Yanomama and Wapishana of South America. In aggregate, the Aztec-derived Mixtec

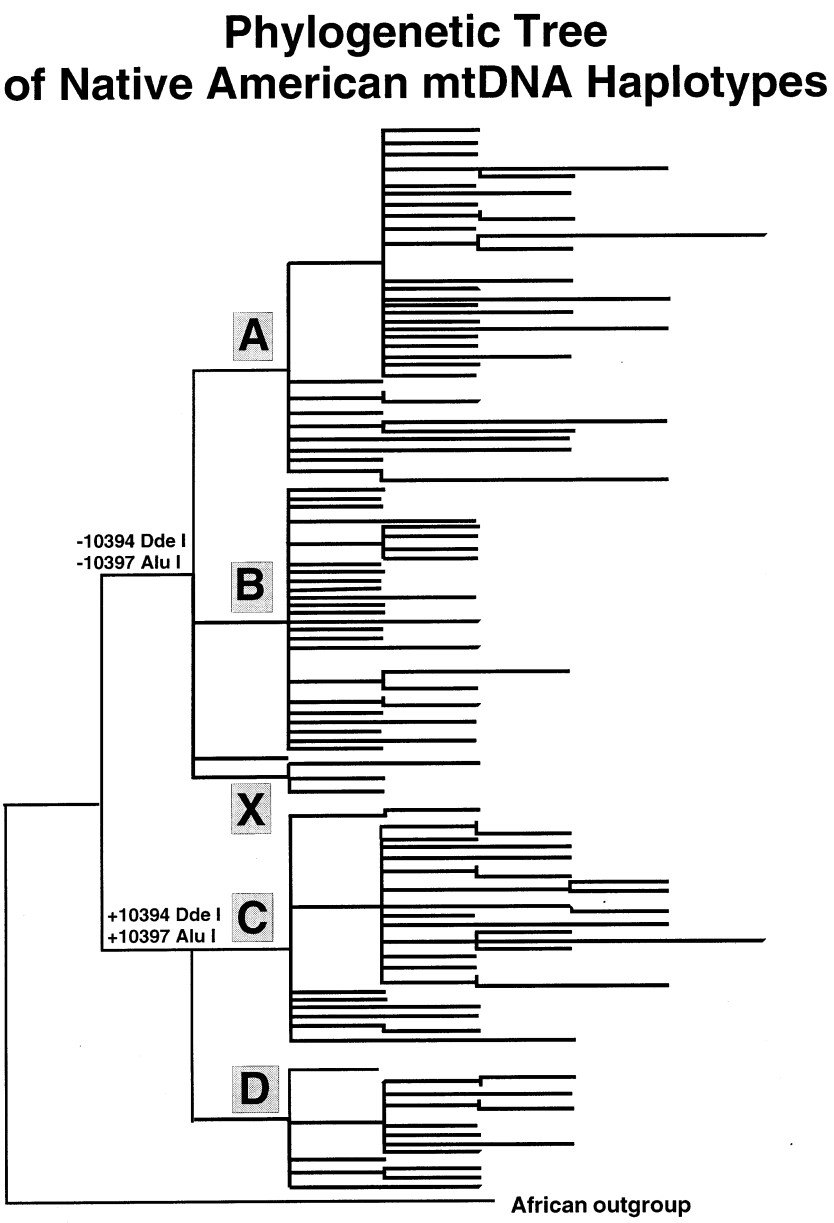

Fig. 5. Phylogeny of Native American mtDNAs. This representative tree was derived from the haplotype data obtained from approximately 563 Paleo-Indians, 130 Na-Dené, and 50 Eskimos, and reveals a dramatic point: all Native American mtDNAs were derived from only five founder mtDNAs. These founders originated from haplogroups A, B, C, and D, found in Asia, and from haplogroup X, found only in Europe. The Amerind-speaking Paleo-Indians of North, Central, and South America encompass haplogroups A, B, C, and D. The Paleo-Indians of North Central, North America also encompass haplogroup X. The Na-Dené of the Pacific Northwest and Southwest (Apache and Navajo) are primarily A with 1/3 having the Na-Dené-specific marker, an *RsaI* site loss at np 16 329. The Eskimos of the sub-Arctic are primarily haplogroups A and D. This phylogeny is derived from data reported by Wallace and coworkers (Wallace et al., 1985; Schurr et al., 1990, 1999; Torroni et al., 1992, 1993a, 1994a,d; Torroni and Wallace, 1995; Brown et al., 1998; Starikovskaya et al., 1998).

and Zapotecs harbored three haplogroups: 66% A, 18% B, and 16% C. The linguistically related Pima of Arizona harbored the same haplogroups, as did the Mixe, suggesting a common ancestry. The Maya were more similar to the North American Paleo-Indians, whereas the Chibchans and South American tribes differed from the more northern tribes and from each other. These results suggest that the Maya and Aztecs may have been derived from different populations (Torroni et al., 1994d).

The prevalence of haplogroup A, B, C, and D mtDNAs in Native American populations has now been confirmed by multiple investigators (Ward et al., 1991, 1993; Horai et al., 1993; Santos and Barrantes, 1994). These same four haplogroups have also been found in the Native American skeletons excavated from a pre-Columbian burial site in Central North America (Stone and Stoneking, 1993). Although some Native American mtDNAs have been found not to exhibit one of the four

primary mtDNA restriction site markers (Torroni et al., 1992, 1993a; Bailliet et al., 1994), most of these mtDNAs can be shown to result from either recent genetic admixture with European or African immigrants or the secondary gain or loss of informative restriction sites (Torroni and Wallace, 1995). The major exception to this is found in the Native Americans of Central North America. These populations have been discovered to harbor a fifth ancient founding mtDNA lineage, designated haplogroup X (Fig. 5). Analysis of the mtDNA of 42 Ojibwa of the Great Lakes region revealed 11 (26%) that were not from haplogroups A, B, C, or D. Similarly, of the Nuu-Chah-Nulth and the Bella Coola of the Pacific northwest, 13.3% and 4% respectively were non-A–D. Extensive analysis of these mtDNAs revealed that they had some features in common with the rare European haplogroup X. The shared markers included restriction site polymorphisms for an *AvaII* site at np 14 465, the absence of *DdeI* sites at np 1715 and at np 10 394, and the presence of the *HaeIII* site at np 16 517. Control-region sequence analysis further revealed that both the Native American and European haplogroup X mtDNAs shared additional nucleotide variants, located at 16 189C, 16 223T, 16 278T, 73G, 153G, 195G, 225A, and 263B. However, comparison of the European and Native American control-region sequences also revealed clear differences, indicating that the last common ancestor of the European and Native American haplogroup X mtDNAs lived long ago. Moreover, whereas the above-mentioned survey of the pre-Columbian burial site revealed mostly mtDNAs of haplogroups A, B, C, or D, two skeletons were found to be different and to have the same control region sequence as a subset of the Ojibwa haplogroup X mtDNAs (Brown et al., 1998). Hence, the Central North American haplogroup X mtDNAs are not the result of recent European admixture with Native Americans, but arrived in an ancient, pre-Columbian, migration (Brown et al., 1998). A survey of the distribution of haplogroup X throughout Asia, including 411 Siberians (Torroni et al., 1993b) and 207 Asians (Ballinger et al., 1992; Torroni et al., 1994c), failed to reveal a single haplogroup X mtDNA. This implies that the Native American haplogroup X may not be of Asian ancestry. This raises the possibility that haplogroup X came to the Americas in a separate migration originating in Europe. Analysis of the haplogroup X restriction and control region sequence data has permitted us to estimate the divergence time of the American and European haplogroup X mtDNAs, which proved to be in the range of 15 000 to 30 000 YBP. Thus, haplogroup X must have come to the Americas with the original Native American migrations and represents a novel founder lineage of possibly European origin (Brown et al., 1998).

Analysis of mtDNA variation of the Na-Déné has confirmed that they are distinct from the Paleo-Indians. Analysis of northern Na-Déné, including Dogrib, Tlingit, and Haida, indicates that these people harbor mtDNAs from only one of the founding mtDNA haplogroups, that of haplogroup A (Torroni et al., 1992, 1993a). This is substantiated by the southern Na-Déné, including the Apache and Navajo, who are $>60\%$ haplogroup A. The remaining Apache and Navajo mtDNAs are from haplogroups B, C, and D and probably represent recent admixture with adjacent Paleo-Indian tribes. The distinctive nature of the Na-Déné mtDNAs is confirmed by the fact that about a third of all Na-Déné haplogroup A mtDNAs carry a novel variant, an *RsaI* site loss at np 16 329 (A to G at 16 331). Interestingly, this variant is found in all Na-Déné except the Haida. This may suggest that the *RsaI* site loss occurred in the original Beringian populations that gave rise to the Na-Déné in the Americas.

Analyses of the sequence diversity of the Na-Déné haplogroup A mtDNAs gave a value of 0.021%. This is substantially lower than the diversity of the Paleo-Indian haplogroups A+C+D and also different from the Paleo-Indian haplogroup B. Hence, the Na-Déné do appear to have arrived as a single independent migration, which occurred more recently than the Paleo-Indians migration (Torroni et al., 1992, 1993a). This conclusion has been confirmed by analysis of control-region sequence data (Forster et al., 1996).

Analysis of the mtDNA variation in Eskimos has been more difficult, because of the limited availability of samples and their recent divergence from ancestral populations. However, a survey of 129 Siberian Yupik Eskimos, who represent Eskimo peoples inhabiting both sides of the Bering Strait, revealed only haplogroups A and D. Hence, the Eskimos may also be distinct from the Na-Déné and Paleo-Indians, as predicted from linguistic associations (Torroni et al., 1993b; Starikovskaya et al., 1998).

Times of the Native American migrations, as in the cases of the African, European, and Asian populations, were calculated using the haplogroup-specific mtDNA sequence diversity and the mtDNA sequence evolution rate of 2.2–2.9%/MYR. This rate was calculated using the Chibcha-speaking peoples of Central America, who are estimated from anthropological and nuclear genetic data to have originated about 8000–10 000 YBP. Analyzing the mtDNA variation of 110 Chibchans representing five tribes revealed that all but one of the Chibchan mtDNAs were from haplogroups A or B, the exception being a single haplogroup D. Moreover, 62% of the Chibchan haplogroup A mtDNAs showed a distinctive private polymorphism, the loss of an *MspI* site at np 104, the result of a small deletion in the 3′ end of the D-loop hypervariable region II (Santos and Barrantes, 1994; Torroni et al., 1994a). Averaging the sequence diversity of the haplogroup A and B mtDNAs and dividing by the putative age of the population gave

the sequence evolution rate of 2.2–2.9%/MYR (Torroni et al., 1994a).

Using this Native American sequence evolution rate, we calculated that the first Paleo-Indian migration carrying haplogroups A, C, and D (sequence diversity of 0.075%) arrived 26 000–34 000 YBP and that the second Paleo-Indian migration bringing haplogroup B (0.034%) arrived 12 000–15 000 YBP. The Na-Déné migration with a sequence diversity of 0.021% was estimated to have arrived 7200–9000 YBP, a value strikingly similar to the 9500 YBP value estimated by glottochronology (Torroni et al., 1994a) (Fig. 5).

Comparable estimates from control-region sequence data have suggested an initial major expansion out of northeastern Siberia into the Americas to generate the Paleo-Indians (Amerind) occurring about 20 000 to 25 000 YBP, followed by a second rapid expansion out of Beringia about 11 300 YBP giving rise to the Na-Déné and Eskimos (Forster et al., 1996).

3.7. Siberian antecedents to Native American migrations

With the Siberian origin of Native Americans established, it was of interest to determine which Siberian populations are the most likely progenitors of the various Native American migrations. To address this question, we have performed an extensive analysis of the mtDNAs of the Siberian Eskimos and Chukchi of the Chukotka peninsula, and of the Koryaks and Itel'men of the Kamchatka peninsula. The Chukotka peninsula is the region of northeastern Siberia closest to Alaska, and the Kamchatka peninsula extends south from the Chukotcha peninsula forming the eastern border of the Sea of Okhotsk. Analysis of 79 Siberian Eskimos revealed an mtDNA haplogroup distribution of 77.2% A, 2.5% C, and 20.3% D. Similarly, the haplotype distribution of the Chuckchi was 68.2% A, 10.6% C, 12.1% D, and 9.1% G. Control-region sequence analysis of the Chukotka haplogroup A mtDNAs, and comparison with those of Native American Na-Déné and Northwestern Amerinds, revealed a 16 111 C to T transition that appears to delineate an 'American' enclave of northeastern Siberian mtDNAs. Furthermore, derivatives of this sublineage were found to include some of the progenitors of the coastal Amerinds, some bearing an additional 16 129A variant; the Haida and Bella Coola bear an additional 16 355T variant, and some Eskimos bear an additional 16 265G variant. The 16 111T lineage of haplogroup A is further subdivided by the additional 16 192T variant. In addition to being found in Chuckotka, this 16 111T + 16 192T lineage is the progenitor of the Na-Déné lineage, which bears the variants 16 233G and 16 331G, the latter site generating the characteristic Na-Déné *RsaI* site loss at np 16 329 (16 331 A to G). Hence Chukotka seems to harbor some remnants of the progenitors of the Beringia expansion that gave rise to the Na-Déné and Eskimo populations (Starikovskaya et al., 1998).

Analysis of the Koryaks and Itel'men of the Kamchatka peninsula revealed a very different haplogroup distribution. The Koryaks proved to be 5.2% A; 36.1% C, 1.3% D, 41.9% G, 9.7% Y, and 5.8% Z. The Itel'men were 6.4% A, 14.9% C, 68.1% G, 4.3% Y, and 6.4% Z. Thus, even though the Chukotka and Kamchatka populations are proximate to each other and the Chukchi and Koryaks have related languages, the mtDNAs of these are strikingly distinct. Indeed, in a neighbor-joining tree analysis, the Eskimos and Chukchi were located at one end of the tree with the Amerinds and Na-Déné, whereas the Koryaks and Itel'men were at the other end of the tree with the Ainu, Japanese, Koreans, and other Siberians (Schurr et al., 1999).

The predominance of haplogroup C in the Paleo-Indians (Amerind) is most similar to the haplogroup distribution seen in the central Siberian, where the Evenks and Udeguys are 84.3% C and 17.8% C respectively. This raises the possibility that the progenitors of the Paleo-Indians may currently reside in central Siberia (Starikovskaya et al., 1998; Schurr et al., 1999).

4. Leber's hereditary optic neuropathy (LHON) and the interaction between mtDNA disease mutations and background mtDNA variation

Whereas some germline mtDNA mutations are neutral and become established in the human populations by genetic drift, many are deleterious and result in genetic disease. Deleterious mtDNA mutations vary widely in their severity, but the mildest mutations result in an acute onset blindness syndrome known as LHON. Extensive analysis of the mtDNAs of LHON patients has revealed that certain mtDNA haplotypes are prone to expression of LHON more than others. This indicates that different population-specific mtDNA may be functionally different and hence might have been influenced by selection.

LHON classically manifests as acute-onset, bilateral, central vision loss associated with the degeneration of the retinal ganglion cell layer and optic nerve. Typically, the onset and progression of blindness is relatively rapid, with both eyes developing vision loss within a year of each other and with blindness usually the only clinical sign. However, in some LHON cases other more sever neurological abnormalities are also observed, the most common of which is early-onset dystonia associated with basal ganglia degeneration.

LHON is strictly maternally inherited, though the penetrance is highly variable and there is an unexplained predilection for males to be preferentially affected. A typical pedigree for the most common mtDNA mutation, MTND4*LHON11778A, is shown in Fig. 6.

4.1. MtDNA mutations and LHON

A total of 23 mtDNA missense mutations have been associated with LHON patients (Table 1). However, detailed genetic analysis has revealed that only a few 'primary' mutations contribute in a major way to the development of blindness. The remaining 'secondary' mutations may contribute to LHON by increasing the probability of expressing the phenotype or may simply be linked to other clinically important variants in the same mtDNA haplotype.

The four 'primary' LHON mutations, listed in order of decreasing severity, are MTND6*LDYT14459A (Jun et al., 1994; Shoffner et al., 1995), MTND4*LHON 11778A (Wallace et al., 1988b), MTND1*LHON3460A (Howell et al., 1991; Huoponen et al., 1991, 1993), and MTND6*LHON14484C (Johns et al., 1992a; Mackey and Howell, 1992) (Table 1). The individual primary mutations represent strong risk factors for expression of maternally inherited LHON, they have been observed in a number of unrelated LHON families, they rarely co-occur with each other, and they have not been detected in a large number of control mtDNAs (Brown and Wallace, 1994a; Brown et al., 1995; Howell et al., 1995; Howell, 1997). In patients of European descent, the MTND4*LHON11778A mutation accounts for about 50% of cases, whereas the MTND1*LHON3460A and MTND6*LHON14484C mutations encompass roughly 15% each (Table 1). In Asia, the MTND4*LHON11778A mutation accounts for 95% of patients (Mashima et al., 1993). Such continental differences suggest that genetic backgrounds may affect the expression of the different LHON mutations.

The MTND6*LDYT14459A mutation, while relatively rare, is the most severe of the primary LHON mutations (Jun et al., 1994; Shoffner et al., 1995). This mutation can be manifested in two very different phenotypes, LHON and/or generalized dystonia. The MTND6*LDYT14459A mutation converts the highly conserved alanine at codon 72 in the ND6 polypeptide to a valine, has been found to be heteroplasmic in at least some family members in every pedigree studied (Table 1) (Jun et al., 1994; Shoffner et al., 1995). The optic atrophy associated with his mutation is similar to

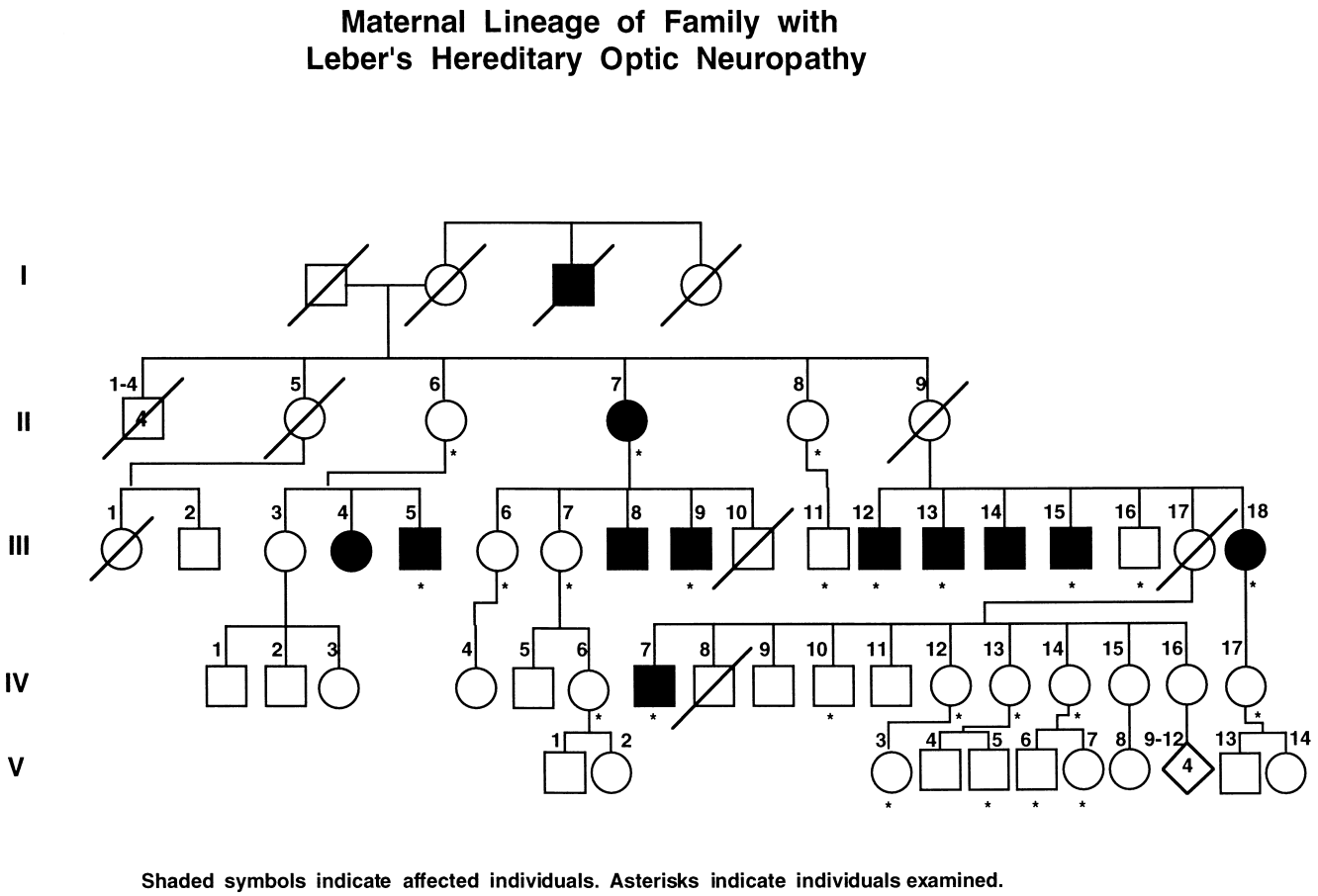

Fig. 6. Maternally inherited pedigree of LHON due to the MTND4*LHON11778A mutation. Affected individuals (filled symbols) experience acute onset optic atrophy and central vision loss, generally as young adults. Even though the mutation is essentially homoplasmic, penetrance among maternal relatives is highly variable. Moreover, males are about three to four times more likely to lose their vision than females (Wallace et al., 1988b; Newman et al., 1991).

Table 1
LHON disease mutations

No.[a]	Mutation	Class[b]	Other neurol. disease	Amino acids[c]		Approx. European patients (%)	Controls (%)	Heteroplasmy	Penetrance[d] (%)		Recovery[e] (%)
				Cons	Change				Relatives	Males	
1	MTND6*LDYT14459A	Primary	+/−	A72V	M	Rare	0	+	61	58	Low
2	MTND4*LHON11778A	Primary	+/−	R340H	H	50	0	+/−	33–60	82	4
3	MTND1*LHON3460A	Primary	+/−	A52T	M	15	0	+/−	14–75	40–80	22
4	MTND6*LHON14484C	Primary	−	M64V	L	15	0	+/−	27–80	68	37-50
5	MTND2*LHON5244A	Primary	−	G2595	H	Rare	0	+	NA	NA	UN
6	MTND5*LHON13730A	Primary	−	G465E	M	Rare	0	+	NA	NA	Yes
7	MTCO3*LHON9804A	Primary?	−	A200T	H	1.5	0	−	UN	UN	UN
8	MTND3*LHON10663C	Primary?	−	V65A	L	Rare	0	−	56	60	UN
9	MTATP6*LHON9101C	Primary?	−	I192T	L	Rare	0	−	NA	NA	UN
10	MTND4*LDYT11696G	Primary?	+/−	V312I	L	Rare	0	−	UN	UN	UN
11	MTND6*LHON14482G	Primary?	−	M64I	L	Rare	0	−	UN	89	UN
12	MTND6*LHON14498T	Primary?	−	Y59C	M	Rare	0	+	31	50	UN
13	MTND6*LHON14568T	Primary?	−	G36S	L	Rare	0	−	NA	NA	UN
14	MTND6*LDYT14596A	Primary?	+/−	I26M	M	Rare	0	+	UN	UN	UN
15	MTCYB*LHON15257A	Intermediate	−	D171N	H	9	0.4	−	NA	NA	NA
16	MTND5*LHON13708A	Secondary	−	A458T	M	30	6	−	NA	NA	NA
17	MTND1*LHON3394C	Secondary	−	Y30H	H	Rare	0.9	−	NA	NA	NA
18	MTND1*LHON4160C	Secondary	++	L285P	H	Rare	0	−	76	54	0
19	MTND1*LHON4216C	Secondary	−	Y304H	L	~40	13	−	NA	NA	NA
20	MTND2*LHON4917G	Secondary	−	D150N	H	3	3	−	NA	NA	NA
21	MTCO1*LHON7444A	Secondary	−	Ter→K	NA	5	1	−	NA	NA	UN
22	MTCO3*LHON9438A	Secondary	−	G78S	H	2.5	4.6	−	NA	UN	NA
23	MTCYB*LHON15812A	Secondary	−	V357M	M	4	0.1	−	NA	NA	NA

[a] The first 10 LHON-associated mtDNA mutations are listed in order of estimated severity (see text).
[b] A question mark in classification indicates transient assignment pending more data.
[c] H = high amino acid conservation; M = moderate; L = low; NA = not applicable; Ter = termination codon.
[d] NA = not applicable; UN = unknown; penetrance estimate for the 10 663 mutation is from a single LHON family that does not harbor a common primary LHON mutation by complete mtDNA sequence analysis.
[e] Low = anecdotal low degree of vision recovery; UN = unknown; NA = not applicable.

that seen for other LHON mutations. By contrast, the dystonia caused by this mutation has a mean age of onset of 4 years, with a range of 1.5 to 9 years. In these dystonia patients, the motor system was primarily involved, resulting in gait disturbance and rigidity of the lower extremities, which advances with age to include the upper extremities. Patients also developed pseudobulbar syndrome (swallowing and speech problems), impaired intelligence, short stature and myopathic features. These symptoms are often associated with bilateral striatal necrosis (Novotny et al., 1986), the loss of cells in the striatum, putamen and caudate. The penetrance of the MTND6*LDYT14459A mutation among maternal relatives is about 61%, with 58% of affected individuals being male. There is no record of affected individuals recovering (Table 1).

The MTND4*LHON11778A mutation is the most common cause of LHON, and the second most severe mutation. It converts the highly conserved ND4 codon 340 from an arginine to a histidine (Wallace et al., 1988b). Among MTND4*LHON11778A families, about 33–60% of maternal relatives are affected, with 82% of the affected individuals being male and 18% female. About 14% of cases are heteroplasmic (Newman et al., 1991; Brown and Wallace, 1994b). Furthermore, only about 4% of affected individuals experience visual recovery. Hence, the MTND4*LHON11778A mutation has a slightly lower penetrance than the MTND6*LDYT14459A mutation, and a much greater bias towards males being affected.

Of the affected individuals, the mean age of onset is 27.6 years, with a range of onset from 8 to 60 years. About 58% of patients show additional ophthalmological features, including peripapillary telangiectasias, microangiopathy, disk pseudo-edema, and vascular tortuosity. A total of 55% of patients have a simultaneous onset of vision loss in both eyes, and once vision loss begins it can progress rapidly or slowly, with a mean length of progression for 3.7 months and a range of 0 to 24 months. In about 98% of cases, the final visual acuity is 20/200 or worse, and only 2% are better than 20/200 (Newman et al., 1991). Occasional individuals also show additional neurological symptoms, reminiscent of those of the MTND6*LDYT14459A mutation including ataxia, spastic paraparesis, extrapyramidal signs, dystonic rigidity associated with bilateral basal ganglia, and cerebellar and pontine atrophy, etc. (Larsson et al., 1991; Funakawa et al., 1995; Vergani

et al., 1995). In LHON pedigrees with the MTND4*LHON11778A mutation, about 1–2% of female patients also manifest a multiple-sclerosis-like demyelination disease (Newman et al., 1991; Harding et al., 1992; Flanigan and Johns, 1993; Hanefeld et al., 1994; Kellar-Wood et al., 1994; Olsen et al., 1995). Interestingly, in the Japanese there is no such association between the MTND4*LHON11778A mutation and multiple sclerosis (Nishimura et al., 1995), again suggesting that genetic background may be important in disease expression.

The MTND1*LHON3460A mutation is the next most severe. It changes a moderately conserved alanine at codon 52 in the ND1 gene to a threonine and has been observed to be heteroplasmic in a number of families (Howell et al., 1991; Huoponen et al., 1991, 1993; Brown and Wallace, 1994b). Generally, the clinical manifestations of this mutation are confined to LHON, having features quite similar to MTND4*LHON11778A. Only occasionally is the MTND1*LHON3460A mutation associated with other neurological signs (Table 1), though patients harboring this mutation can manifest multiple sclerosis (Kellar-Wood et al., 1994; Nikoskelainen et al., 1995). The number of affected maternal relatives has varied widely in different studies, but can approach 75% (Johns et al., 1992b; Harding et al., 1995), with between 40% and 80% of all affected individuals being male. In contrast to the MTND4*LHON11778A mutation, however, approximately 22% of affected individuals with the MDND1*LHON3460A have been reported to experience visual recovery (Johns et al., 1993).

The MTND6*LHON14484C mutation is the mildest of the common primary mutations. This mutation changes the weakly conserved methionine at codon 64 in the ND6 protein to a valine. The pathogenicity of the MTND6*LHON14484C mutation is nicely confirmed by one of the rare LHON mutations: MTND6*LHON14482G. This mutation alters the same codon as the MTND6*LHON14484C mutation, but converts the methione to an isoleucine, yet both mutations are associated with LHON (Howell, 1998). The MTND6*LHON14482C mutation has only very rarely been reported to be heteroplasmic (Table 1) (Johns et al., 1992a; Mackey and Howell, 1992; Biousse et al., 1997), suggesting that the great majority of mtDNA patients must be mutant to cause a sufficient biochemical defect to result in optic atrophy. Clinically, this mutation is almost exclusively confined to an LHON phenotype. The penetrance of LHON in MTND6*LHON14484C mutation maternal relatives is about 27–80%, and the male to female ratio of affected individuals is similarly high compared with that for the MTND4*LHON11778A and MTND1*LHON3460A mutations (Johns et al., 1993; Harding et al., 1995; Riordan-Eva et al., 1995).

The MTND6*LHON14484C mutation is additionally noteworthy in that fully 37–50% of patients report visual improvement (Table 1) (Johns et al., 1993; Riordan-Eva et al., 1995). Thus, the MTND6*LHON14484C mutation has the lowest pathogenicity of the primary LHON mutations.

Analysis of the background mtDNA haplotypes in LHON families harboring the various primary mutations has shown that most families are new mutations. However, striking background haplotype associations have been found in patients with the milder LHON mutations, suggesting an interaction between the recent mild pathogenic mutation and the ancient background genotype.

The MTND6*LDYT14459A mutation was first identified in a large Hispanic family, which proved to harbor a Native American mtDNA from haplogroup D (Jun et al., 1994, fig. 20). A second case was found in an African–American family harboring an African haplogroup L mtDNA, and involved a mother and daughter with LHON in which the daughter also had unilateral striatal degeneration. The third case involved a European child with generalized dystonia, who was found to harbor European haplogroup I mtDNA (Jun et al., 1994; Shoffner et al., 1995). Since mtDNA haplotypes of these three families encompass most of the world's mtDNA sequence diversity, the MTND6*LDYT14459A mutations must have arisen recently and independently in each family (Jun et al., 1994; Shoffner et al., 1995). Whether the phenotype variation is in any way related to variation in haplotype could not be determined.

A similar haplotype analysis has been performed on 47 LHON patients of European descent harboring the three common 'primary' mutations: MTND4*LHON11778A, MTND1*LHON3460A, and MTND6*LHON14484C (Fig. 7) (Brown et al., 1995). This revealed that patients harboring the MTND4*LHON11778A and MTND1*LHON3460A mutations had a variety of different mtDNA background haplotypes dispersed throughout the European mtDNA haplogroups. Thus, most, if not all, LHON families with these mutations are due to independent mutational events. By contrast, most patients harboring the MTND6*LHON14484C mutation clustered together on a single mtDNA lineage, haplogroup J (Fig. 7). However, the MTND6*LHON14484C patients were still dispersed among the haplogroup J controls, indicating that most, if not all, MTND6*LHON14484C pedigrees are new mutations (Brown et al., 1995). Still, the association between the MTND6*LHON14484C mutation and haplogroup J is striking. Haplogroup J is present in only about 9% of the general European population, but it was found in 37% of LHON patients harboring the MTND4*LHON11778A mutation and 80% of LHON patients harboring the MTND6*

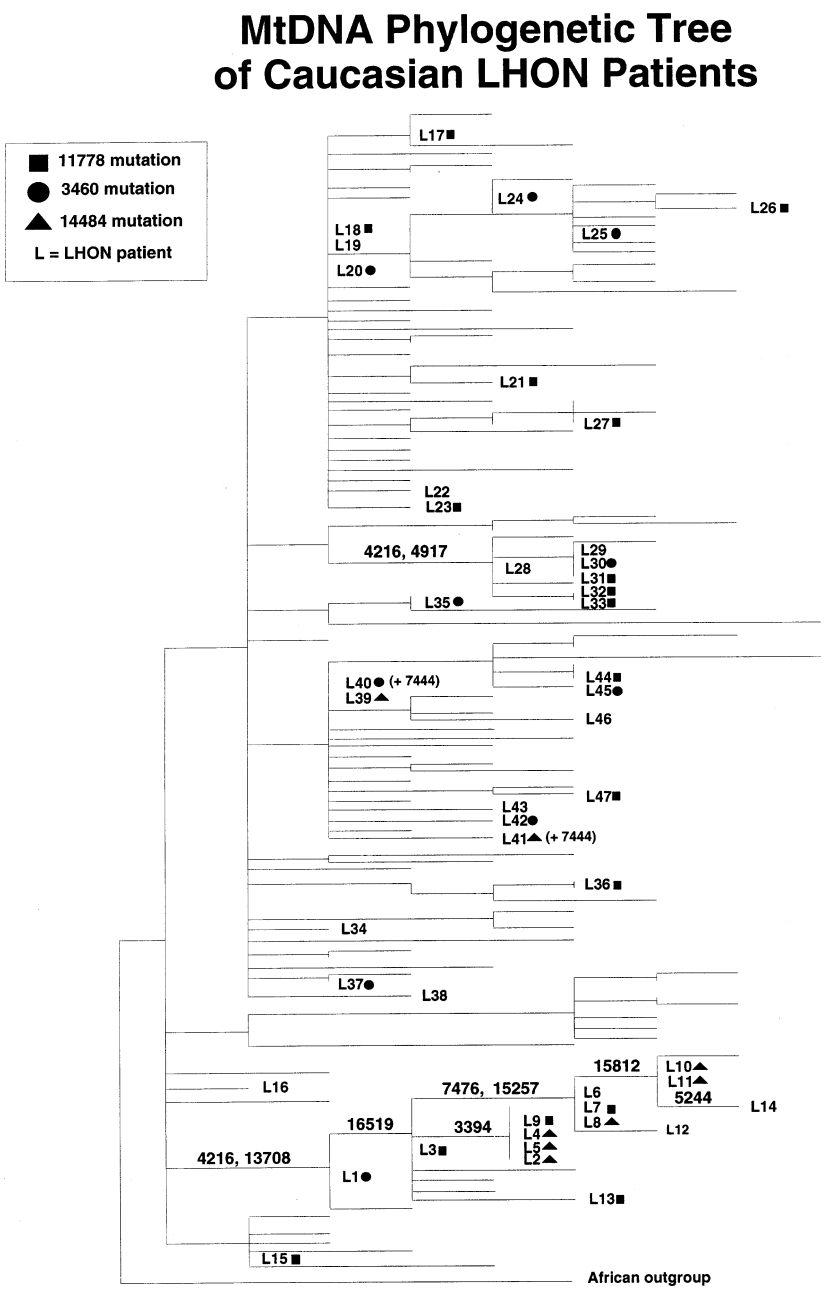

Fig. 7. Phylogenetic tree of Caucasian mtDNAs from 47 LHON patients (L1–L47, bold) and 175 Caucasian controls obtained from the United States and Canada. Haplotypes were determined by RFLP analysis using the entire mtDNA. For the patients, all mtDNA haplotypes are shown and the presence of a common primary mutation is indicated (see key). Additional mutations of interest are noted in parentheses. Multiple occurrences of the three common primary LHON mutations are indicated by their presence in independent mtDNA lineages, and, when occurring within the same large mtDNA lineage, by the presence of intermediate haplotypes lacking LHON mutations. The numbers in large, bold characters on branches indicate mutations that define specific groups of mtDNAs, but they are not necessarily the mutational steps used to create the phylogeny. The horizontal branch lengths are proportional to the number of mutational events that separate haplotypes. From Brown et al. (1995) with permission.

LHON14484C mutation (Brown et al., 1997). Indeed, only rarely have European patients with the MTND6*LHON14484C mutation been found to have a non-haplogroup J mtDNA (Biousse et al., 1997), and in only one case has the MTND6*LHON14484C mutation been reported to be associated with the African haplogroup L mtDNA (Torroni et al., 1996b).

Why might haplogroup J augment the expression of the MTND6*LHON14489C mutation? Haplogroup J mtDNAs might harbor a number of important amino acid substitution polymorphisms that augment the expression of the MTND6*LHON14484C mutation. The two variants that define the main haplogroup J lineage are MTND5*LHON13708A and MTND1*

LHON4216C (Johns et al., 1992a, 1993; Obermaier-Kusser et al., 1994; Oostra et al., 1994; Brown et al., 1995; Howell et al., 1995; Riordan-Eva et al., 1995). The MTND5*LHON13708A mutation converts a moderately conserved alanine at codon 458 in the ND5 protein to a threonine (Johns and Neufeld, 1991; Brown et al., 1992a), whereas the MTND1*LHON4216C mutation converts a weakly conserved tyrosine at codon 304 in the ND1 protein to histidine (Johns and Neufeld, 1991; Brown et al., 1992a). The sequencing of multiple haplogroup J mtDNAs has failed to reveal any other obvious sequence variants that could account for this mtDNA lineage's predilection toward increased expression of LHON (Brown et al., 1992b).

One major sub-branch of haplogroup J is delineated by the MTCYB*LHON15257A variant. This variant converts the highly conserved aspartate at codon 171 in the cytochrome b gene to an asparagine (Brown et al., 1992a), and is even more frequently associated with LHON. Hence the MTCYB*LHON15257A mutation might also contribute to the mitochondrial defect (Fig. 7, Table 1). A second sub-branch of haplogroup J is defined by the MTND1*LHON3394C variant. This changes the highly conserved tyrosine at codon 30 in the ND1 protein to histidine. The MTND1*LHON3394C variant is prevalent in French Canadians, but is also unlikely to contribute to the disease (Fig. 7, Table 1).

All haplogroup J mtDNAs harbor the MTND1*LHON4216C variant along with the MTND5*LHON 13708A variant. However, MTND1*LHON4216C can also be associated with MTND2*LHON4917G. The MTND2*LHON4917G variant changes a highly conserved aspartate at codon 150 in the ND2 protein to an asparagine (Brown et al., 1995). The sharing of the MTND1*LHON4216C between these two lineages raises the possibility that the two lineages may be sub-branches of a common single larger haplogroup that predisposes individuals to LHON (Howell et al., 1995). If so, then the mild MTND1*LHON4216C variant might be relevant for diseases. Alternatively, the predisposition to LHON might be caused by the MTND5*LHON13708A variant on one lineage and the MTND2*LHON4917 variant on the other.

4.2. Adaptation and the mtDNA

Regardless of the actual polymorphisms that contribute to the expressivity of the MTND6*LHON14484C mutation, it is clear that haplogroup J mtDNAs are biochemically different from those of other population-specific mtDNA lineages. This clearly demonstrates that at least some of the population-specific polymorphisms in the mtDNA can have adaptive significance.

If haplogroup J is more deleterious, why has it reached polymorphic frequencies in the European population? One possibility is that it was selected during the ice ages.

It has been proposed that a major selective pressure that acted on early European populations was the repeated periods of severe cold, associated with the advance and retreat of the glaciers. In this context, mitochondrial OXPHOS has two major functions: providing energy for the cell in the form of ATP and generating heat of thermal regulation. However, the generation of heat is accomplished at the expense of ATP-generating capacity, since both processes compete for the potential energy stored in the electrochemical gradient ($\Delta\psi$) maintained across the mitochondrial inner membrane. Therefore, during the ice ages in Europe there may have been strong selective advantage for mtDNA mutations that partially uncoupled OXPHOS, biasing the use of calories toward increased heat production and better thermal regulation, at the expense of efficient ATP production. Although such mutations would be relatively neutral in conditions of caloric excess, and hence be maintained in the population at polymorphic frequencies, they would have an increased tendency for disease expression when combined with a very mild pathogenic mutation such as MTND6*LHON14484C. Hence, the presence of this mtDNA would increase the expressivity of subsequent pathogenic mtDNA mutations.

If this analysis is correct, then the population distributions of the various population-specific mtDNA lineages might not simply reflect genetic drift. Further, selection at various times in human history might have acted to increase the prevalence of certain lineages over others.

References

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G., 1981. Nature 290, 457–465.

Bailliet, G., Rothhammer, F., Carnese, F.R., Bravi, C.M., Bianchi, N.O., 1994. American Journal of Human Genetics 55, 27–33.

Ballinger, S.W., Schurr, T.G., Torroni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.H., Wallace, D.C., 1992. Genetics 130, 139–152.

Biousse, V., Brown, M.D., Newman, N.J., Allen, J.C., Rosenfeld, J., Meola, G., Wallace, D.C., 1997. Neurology 49, 1136–1138.

Blanc, H., Chen, K.H., D'Amore, M.A., Wallace, D.C., 1983. American Journal of Human Genetics 35, 167–176.

Brockington, M., Sweeney, M.G., Hammans, S.R., Morgan-Hughes, J.A., Harding, A.E., 1993. Nature Genetics 4, 67–71.

Brown, M.D., Wallace, D.C., 1994a. Clinical Neuroscience 2, 138–145.

Brown, M.D., Wallace, D.C., 1994b. Journal of Bioenergetics and Biomembranes 26, 273–289.

Brown, M.D., Voljavec, A.S., Lott, M.T., Torroni, A., Yang, C.C., Wallace, D.C., 1992a. Genetics 130, 163–173.

Brown, M.D., Voljavec, A.S., Lott, M.T., MacDonald, I., Wallace, D.C., 1992b. FASEB Journal 6, 2791–2799.

Brown, M.D., Torroni, A., Reckord, C.L., Wallace, D.C., 1995. Human Mutation 6, 311–325.
Brown, M.D., Sun, F., Wallace, D.C., 1997. American Journal of Human Genetics 60, 381–387.
Brown, M.D., Hosseini, S.H., Torroni, A., Bandelt, H.J., Allen, J.C., Schurr, T.G., Scozzari, R., Cruciani, F., Wallace, D.C., 1998. American Journal of Human Genetics 63, 1852–1861.
Brown, W.M., 1980. Proceedings of the National Academy of Sciences of the United States of America 77, 3605–3609.
Brown, W.M., George, M., Wilson, A.C., 1979. Proceedings of the National Academy of Sciences of the United States of America 76, 1967–1971.
Cann, R.L., Wilson, A.C., 1983. Genetics 104, 699–711.
Cann, R.L., Stoneking, M., Wilson, A.C., 1987. Nature 325, 31–36.
Case, J.T., Wallace, D.C., 1981. Somatic Cell Genetics 7, 103–108.
Chen, X., Prosser, R., Simonetti, S., Sadlock, J., Jagiello, G., Schon, E.A., 1995a. American Journal of Human Genetics 57, 239–247.
Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S., Wallace, D.C., 1995b. American Journal of Human Genetics 57, 133–149.
Chen, Y.S., Schurr, T.G., Olckers, A., Kogelnik, A., Huoponen, K., Wallace, D.C., 1999.
Denaro, M., Blanc, H., Johnson, M.J., Chen, K.H., Wilmsen, E., Cavalli Sforza, L.L., Wallace, D.C., 1981. Proceedings of the National Academy of Sciences of the United States of America 78, 5768–5772.
Di Rienzo, A., Wilson, A.C., 1991. Proceedings of the National Academy of Sciences of the United States of America 88, 1597–1601.
Flanigan, K.M., Johns, D.R., 1993. Neurology 43, 2720–2722.
Forster, P., Harding, R., Torroni, A., Bandelt, H.J., 1996. American Journal of Human Genetics 59, 935–945.
Funakawa, I., Kato, H., Terao, A., Ichihashi, K., Kawashima, S., Hayashi, T., Mitani, K., Miyazaki, S., 1995. Journal of Neurology 242, 75–77.
Giles, R.E., Blanc, H., Cann, H.M., Wallace, D.C., 1980. Proceedings of the National Academy of Sciences of the United States of America 77, 6715–6719.
Greenberg, J.H., Turner II, C.G., Zegura, S.L., 1986. Current Anthropology 27, 477–497.
Gyllensten, U., Wharton, D., Josefsson, A., Wilson, A.C., 1991. Nature 352, 255–257.
Hanefeld, F.A., Ernst, B.P., Wilichowski, E., Christen, H.J., 1994. Neuropediatrics 25, 331.
Harding, A.E., Sweeney, M.G., Miller, D.H., Mumford, C.J., Kellar-Wood, H., Menard, D., McDonald, W.I., Compston, D.A.S., 1992. Brain 115, 979–989.
Harding, A.E., Sweeney, M.G., Govan, G.G., Riordan-Eva, P., 1995. American Journal of Human Genetics 57, 77–86.
Hedges, S.B., Kumar, S., Tamura, K., 1991. Science 255, 737–739.
Hertzberg, M., Mickleson, K.N.P., Serjeantson, S.W., Prior, J.F., Trent, R.J., 1989. American Journal of Human Genetics 44, 504–510.
Hopkin, K., 1999. Scientific American 280, 21.
Horai, S., Matsunaga, E., 1986. Human Genetics 72, 105–117.
Horai, S., Hayasaka, K., 1990. American Journal of Human Genetics 46, 828–842.
Horai, S., Kondo, R., Nakagawa-Hattori, Y., Hayashi, S., Sonoda, S., Tajima, K., 1993. Molecular Biology and Evolution 10, 23–47.
Horai, S., Hayasaka, K., Kondo, R., Tsugane, K., Takahata, N., 1995. Proceedings of the National Academy of Sciences of the United States of America 92, 532–536.
Howell, N., 1997. Journal of Bioenergetics and Biomembranes 29, 165–173.
Howell, N., 1998. Vision Research 38, 1495–1504.
Howell, N., Bindoff, L.A., McCullough, D.A., Kubacka, I., Poulton, J., Mackey, D., Taylor, L., Turnbull, D.M., 1991. American Journal of Human Genetics 49, 939–950.
Howell, N., Kubacka, I., Halvorson, S., Howell, B., McCullough, D.A., Mackey, D., 1995. Genetics 140, 285–302.
Huoponen, K., Vilkki, J., Aula, P., Nikoskelainen, E.K., Savontaus, M.L., 1991. American Journal of Human Genetics 48, 1147–1153.
Huoponen, K., Lamminen, T., Juvonen, V., Aula, P., Nikoskelainen, E., Savontaus, J.L., 1993. Human Genetics 92, 379–384.
Johns, D.R., Neufeld, M.J., 1991. Biochemical and Biophysical Research Communications 181, 1358–1364.
Johns, D.R., Neufeld, M.J., Park, R.D., 1992a. Biochemical and Biophysical Research Communications 187, 1551–1557.
Johns, D.R., Smith, K.H., Miller, N.R., 1992b. Archives of Ophthalmology 110, 1577–1581.
Johns, D.R., Heher, K.L., Miller, N.R., Smith, K.H., 1993. Archives of Ophthalmology 111, 495–498.
Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C., Cavalli-Sforza, L.L., 1983. Journal of Molecular Evolution 19, 255–271.
Jun, A.S., Brown, M.D., Wallace, D.C., 1994. Proceedings of the National Academy of Sciences of the United States of America 91, 6206–6210.
Kaneda, H., Hayashi, J., Takahama, S., Taya, C., Lindahl, K.F., Yonekawa, H., 1995. Proceedings of the National Academy of Sciences of the United States of America 92, 4542–4546.
Kellar-Wood, H., Robertson, N., Govan, G.G., Compston, D.A., Harding, A.E., 1994. Annals of Neurology 36, 109–112.
Larsson, N.G., Andersen, O., Holme, E., Oldfors, A., Wahlstrom, J., 1991. Annals of Neurology 30, 701–708.
Mackey, D., Howell, N., 1992. American Journal of Human Genetics 51, 1218–1228.
Mashima, Y., Hiida, Y., Oguchi, Y., Kudoh, J., Shimizu, N., 1993. Human Genetics 92, 101–102.
Merriwether, D.A., Clark, A.G., Ballinger, S.W., Schurr, T.G., Soodyall, H., Jenkins, T., Sherry, S.T., Wallace, D.C., 1991. Journal of Molecular Evolution 33, 543–555.
Michaels, G.S., Hauswirth, W.W., Laipis, P.J., 1982. Developmental Biology 94, 246–251.
Neckelmann, N., Li, K., Wade, R.P., Shuster, R., Wallace, D.C., 1987. Proceedings of the National Academy of Sciences of the United States of America 84, 7580–7584.
Nei, M., Tajima, F., 1983. Genetics of South America 105, 207–217.
Newman, N.J., Lott, M.T., Wallace, D.C., 1991. American Journal of Ophthalmology 111, 750–762.
Nikoskelainen, E.K., Marttila, R.J., Huoponen, K., Juvonen, V., Lamminen, T., Sonninen, P., Savontaus, M.L., 1995. Journal of Neurology, Neurosurgery and Psychiatry 59, 160–164.
Nishimura, M., Obayashi, H., Ohta, M., Uchiyama, T., Hao, Q., Saida, T., 1995. Neurology 45, 1333–1334.
Novotny, E.J., Singh, G., Wallace, D.C., Dorfman, L.J., Louis, A., Sogg, R.L., Steinman, L., 1986. Neurology 36, 1053–1060.
Obermaier-Kusser, B., Lorenz, B., Schubring, S., Paprotta, A., Zerres, K., Meitinger, T., Meire, F., Cochaux, P., Blankenagel, A., Kommerell, G., Jaksch, M., Gerbitz, K.D., 1994. American Journal of Human Genetics 55, 1063–1066.
Olsen, N.K., Hansen, A.W., Norby, S., Edal, A.L., Jorgensen, J.R., Rosenberg, T., 1995. Acta Neurologica Scandinavica 91, 326–329.
Oostra, R.J., Bolhuis, P.A., Wijburg, F.A., Zorn-Ende, G., Bleeker-Wagemakers, E.M., 1994. Journal of Medical Genetics 31, 280–286.
Passarino, G., Semino, O., Bernini, L.F., Santachiara-Benerecetti, A.S., 1996. American Journal of Human Genetics 59, 927–934.
Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M., Santachiara-Benerecetti, A.S., 1998. American Journal of Human Genetics 62, 420–434.
Riordan-Eva, P., Sanders, M.D., Govan, G.G., Sweeney, M.G., DaCosta, J., Harding, A.E., 1995. Brain 118, 319–337.
Saitou, N., Nei, M., 1987. Molecular Biology and Evolution 4, 406–425.
Santos, M., Barrantes, R., 1994. American Journal of Human Genetics 55, 413–414.

Schurr, T.G., Ballinger, S.W., Gan, Y.Y., Hodge, J.A., Merriwether, D.A., Lawrence, D.N., Knowler, W.C., Weiss, K.M., Wallace, D.C., 1990. American Journal of Human Genetics 46, 613–623.

Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B., Wallace, D.C., 1999. American Journal of Physical Anthropology 108, 1–39.

Shoffner, J.M., Lott, M.T., Lezza, A.M., Seibel, P., Ballinger, S.W., Wallace, D.C., 1990. Cell 61, 931–937.

Shoffner, J.M., Brown, M.D., Stugard, C., Jun, A.S., Pollok, S., Haas, R.H., Kaufman, A., Koontz, D., Kim, Y., Graham, J., Smith, E., Dixon, J., Wallace, D.C., 1995. Annals of Neurology 38, 163–169.

Starikovskaya, E.B., Sukernik, R.I., Schurr, T.G., Kogelnik, A.M., Wallace, D.C., 1998. American Journal of Human Genetics 63, 1473–1491.

Stone, A.C., Stoneking, M., 1993. American Journal of Physical Anthropology 92, 463–471.

Stoneking, M., Jorde, L.B., Bhatia, K., Wilson, A.C., 1990. Genetics 124, 717–733.

Swofford, D.L., 1993. University of Illinois, Champaign.

Tateno, Y., Nei, M., Tajima, F., 1982. Journal of Molecular Evolution 18, 387–404.

Templeton, A.R., 1992. Science 255, 737

Torroni, A., Wallace, D.C., 1994. Journal of Bioenergetics and Biomembranes 26, 261–271.

Torroni, A., Wallace, D.C., 1995. American Journal of Human Genetics 56, 1234–1236.

Torroni, A., Schurr, T.G., Yang, C.C., Szathmary, E.J., Williams, R.C., Schanfield, M.S., Troup, G.A., Knowler, W.C., Lawrence, D.N., Weiss, K.M., 1992. Genetics 130, 153–162.

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M., Wallace, D.C., 1993a. American Journal of Human Genetics 53, 563–590.

Torroni, A., Sukernik, R.I., Schurr, T.G., Starikovskaya, Y.B., Cabell, M.F., Crawford, M.H., Comuzzie, A.G., Wallace, D.C., 1993b. American Journal of Human Genetics 53, 591–608.

Torroni, A., Neel, J.V., Barrantes, R., Schurr, T.G., Wallace, D.C., 1994a. Proceedings of the National Academy of Sciences of the United States of America 91, 1158–1162.

Torroni, A., Lott, M.T., Cabell, M.F., Chen, Y., Laverge, L., Wallace, D.C., 1994b. American Journal of Human Genetics 55, 760–776.

Torroni, A., Miller, J.A., Moore, L.G., Zamudio, S., Zhuang, J., Droma, R., Wallace, D.C., 1994c. American Journal of Physical Anthropology 93, 189–199.

Torroni, A., Chen, Y., Semino, O., Santachiara-Beneceretti, A.S., Scott, C.R., Lott, M.T., Winter, M., Wallace, D.C., 1994d. American Journal of Human Genetics 54, 303–318.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L., Wallace, D.C., 1996a. Genetics 144, 1835–1850.

Torroni, A., Carelli, V., Petrozzi, M., Terracina, M., Barboni, P., Malpassi, P., Wallace, D.C., Scozzari, R., 1996b. American Journal of Human Genetics 59, 248–252.

Turner II, C.G., 1983. In: Shutler Jr, R. (Ed.), Early Man in the New World. Sage, Beverly Hills, CA, pp. 147–157.

Turner II, C.G., 1987. American Journal of Physical Anthropology 73, 305–321.

Vergani, L., Martinuzzi, A., Carelli, V., Cortelli, P., Montagna, P., Schievano, G., Carrozzo, R., Angelini, C., Lugaresi, E., 1995. Biochemical and Biophysical Research Communications 210, 880–888.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., Wilson, A.C., 1991. Science 253, 1503–1507.

Wallace, D.C., 1986. Somatic Cell and Molecular Genetics 12, 41–49.

Wallace, D.C., 1987. In: McKusick, V.A., Roderick, T.H., Mori, J., Paul, M.W. (Eds.), Medical and Experimental Mammalian Genetics: A Perspective vol. 23. A.R. Liss, New York, pp. 137–190.

Wallace, D.C., 1995. American Journal of Human Genetics 57, 201–223.

Wallace, D.C., Torroni, A., 1992. Human Biology 64, 403–416.

Wallace, D.C., Garrison, K., Knowler, W.C., 1985. American Journal of Physical Anthropology 68, 149–155.

Wallace, D.C., Ye, J.H., Neckelmann, S.N., Singh, G., Webster, K.A., Greenberg, B.D., 1987. Current Genetics 12, 81–90.

Wallace, D.C., Zheng, X., Lott, M.T., Shoffner, J.M., Hodge, J.A., Kelley, R.I., Epstein, C.M., Hopkins, L.C., 1988a. Cell 55, 601–610.

Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., Elsas, L.J., Nikoskelainen, E.K., 1988b. Science 242, 1427–1430.

Wallace, D.C., Lott, M.T., Shoffner, J.M., Ballinger, S., 1994. Epilepsia 35, S43–S50.

Ward, R.H., Frazier, B.L., Dew-Jager, K., Paabo, S., 1991. Proceedings of the National Academy of Sciences of the United States of America 88, 8720–8724.

Ward, R.H., Redd, A., Valencia, D., Frazier, B., Paabo, S., 1993. Proceedings of the National Academy of Sciences of the United States of America 90, 10 663–10 667.

Watson, E., Bauer, K., Aman, R., Weiss, G., von Haeseler, A., Paabo, S., 1996. American Journal of Human Genetics 59, 437–444.

Wrischnik, L.A., Higuchi, R.G., Stoneking, M., Erlich, H.A., Arnheim, N., Wilson, A.C., 1987. Nucleic Acids Research 15, 529–542.