Journal of Bioenergetics and Biomembranes, Vol. 26, No. 3, 1994

Mitochondrial DNA Variation in Human Populations and Implications for Detection of Mitochondrial DNA Mutations of Pathological Significance

Antonio Torroni[1] and Douglas C. Wallace[1]

Received January 10, 1994; accepted January 20, 1994

Haplotype and phylogenetic analyses of "normal" mitochondrial DNAs (mtDNAs) have allowed a clarification of several controversial issues concerning the origin of humans, the time and colonization pattern of the various regions of the world, and the genetic relationships of modern human populations. More recently, the same type of analyses has also been applied to mtDNA disease studies. A review of these studies indicates that exhaustive screenings of "normal" mtDNA variation in all human populations associated with haplotype and phylogenetic analyses are essential if we are to understand the etiology of mitochondrial pathologies.

The human mitochondrial (mtDNA) has a very high sequence evolution rate, more than ten times that of nuclear genes (Miyata *et al.*, 1982; Wallace *et al.*, 1987). As a consequence of these random mutations, human populations now harbor a wide variety of ethnic-specific polymorphisms which have permitted investigation of our human origins and ancient migration patterns. Since mutations are random, many new mutations alter functionally important gene products. Such deleterious mutations are acted on by selection and eliminated in the form of genetic disease. Since the mtDNA is the most compact chromosome in the human genome, many mtDNA mutations are deleterious and consequently mtDNA diseases are common. Severely deleterious mutations cause multisystem disorders which are rapidly eliminated. Hence, all severe mitochondrial DNA diseases are due to relatively recent mutations. By contrast, mildly deleterious mutations may not be expressed until late in life. Since such mutations do not reduce reproductive fitness, they may have arisen thousands of years ago and became established in the populations as low-frequency polymorphisms. Thus, mtDNA variation provides a continuous array of phenotypic effects from neutral, though mildly deleterious, to severely deleterious. Therefore, to understand the role of mtDNA mutations in human health, we must understand both the nature and origin of new deleterious mutations as well as the characteristics and extent of existing population variation.

The high sequence evolution rate of the mtDNA is a consequence of mitochondrial biology and the unique genetics of the mtDNA. The unusual genetics of the mtDNA derives from its cytoplasmic location and high copy number.

The human mtDNA is a 16,569 nucleotide pair (np), closed circular molecule which codes for a small (12S) and large (16S) ribosomal RNA (rRNA), 22 transfer RNAs (tRNAs), and 13 polypeptides, all of which are components of the mitochondrial energy generating process oxidative phosphorylation (OXPHOS). The mtDNA is particularly prone to mutation since the mtDNA lacks protective proteins like histones, has a low-efficiency repair system, and is attached to the mitochondrial inner membrane, the site of OXPHOS were most of the mutagenic oxygen radicals are generated (Bandy and Davison, 1990).

[1] Department of Genetics and Molecular Medicine, Emory University School of Medicine, Atlanta, Georgia 30322.

0145-479X/94/0600-0261$07.00/0

Mitochondrial inheritance is dictated by four rules. First, the mtDNA is inherited through the oocyte cytoplasm and thus shows maternal inheritance (Giles *et al.*, 1980). Second, since the mtDNA of different humans do not mix at fertilization, mtDNAs cannot recombine. Hence all mtDNA variation is the product of the sequential accumulation of mutations along radiating maternal lineages. Third, when new mutations arise in a cell, they create an intracellular mixture of mutant and normal mtDNAs known as heteroplasmy. As these cells replicated during mitosis or meiosis, the normal and mutant mtDNAs are randomly distributed to the daughter cells. This creates an intracellular genetic drift called replicative segregation which ultimately results in the generation of cells or family lineages that contain pure mutant or normal mtDNAs, called homoplasmy. Fourth, different cells and tissues rely on mitochondrial energy to different extents, in descending order of the central nervous system, heart, skeletal muscle, endocrine system, kidney, and liver. Thus, as the severity of the OXPHOS defect increases, the nature and extent of the clinical symptoms also increases.

The special features of mtDNA genetics together with the nature and extent of mtDNA diversity has required new approaches for identifying and understanding deleterious mtDNA mutations. This is being accomplished by adapting the principles and methods of mitochondrial population genetics to the analysis of the common human degenerative diseases.

This report provides an overview of the temporal development of mtDNA analysis in human populations and demonstrates how haplotype and phylogenetic analyses have helped to finally clarify longly debated anthropological issues concerning the origin of Native Americans and Caucasians. In addition, those mtDNA disease studies which have applied the same approaches in their analyses are discussed. These studies confirm that if we are to understand the molecular bases of human mtDNA diseases, we must first define the "normal" variation in human populations through population and phylogenetic studies.

THE ORIGIN AND ASIAN AFFINITIES OF NATIVE AMERICANS

Native Americans belong to one of the few extant human groups whose ancestors entered a vast uninhabited area and then apparently remained isolated from other human contacts for thousands of years. They are descendants of Asian peoples who, after their expansion to the Americas, underwent through a rapid process of tribal radiation and isolation, localized genetic and cultural differentiation, and adaptation to a variety of different environmental conditions. According to Greenberg (1987), Native Americans can be subdivided into three major linguistic groups, the reasonably well-defined Eskimo-Aleut and Na-Dene of northern North America and the highly heterogeneous Amerind of North, Central, and South America. Analysis of mtDNA variation in Native Americans began in 1985 with analysis of one Amerind population, the Pima-Papago of southern Arizona and was carried out by using six restriction enzymes and Southern blot technique (Wallace *et al.*, 1985). It was observed that about 40% of the Pima-Papago mtDNAs were defined by the loss of an HincII site at np 13259, a polymorphism previously reported at very low frequency in some east Asian populations (Blanc *et al.*, 1983). This preliminary observation raised the possibility that the colonization of the Americas could have been associated with a dramatic reduction of mtDNA variation. In 1990, a second study of Native American mtDNA variation was carried out by using the recently developed PCR technology. The Pima-Papago, the Maya from the Yucatan Peninsula, and the Ticuna from Brazil were screened for a number of mtDNA mutations previously reported in Asians (Schurr *et al.*, 1990). Similarly to the Pima, the Maya and the Ticuna showed high frequencies of the np 13259 HincII site loss, indicating that the hypothesized bottleneck was not limited to the Pima, but preceded the tribal radiation of the Amerinds. In the same study, two Asian polymorphisms, a HaeIII site gain at np 663, and a 9-bp deletion in the intergenic region between the COII and $tRNA^{Lys}$ genes, were found to define two other common sets of Native American mtDNAs. Together with the lineage defined by the HincII site loss, these two newly defined mtDNA groups comprised about 84% of the mtDNAs in the three tribes. Therefore, it was proposed that the remaining 16% of Amerind mtDNAs could have been all members of a fourth group of mtDNAs, whose specific mutation were still unidentified, and that only four Asian mtDNA lineages (haplogroups) colonized the Americas.

In the last few years, we have refined the PCR techniques applied to our population studies. We

amplify the entire mtDNA of each individual in nine PCR fragments of about 2–3 kb. Each fragment is then digested with a standard set of 14 restriction enzymes permitting the screening of about 20% of the mtDNA sequence. With this approach we have analyzed mtDNA variation in 130 Na-Dene from five tribes and almost 500 Amerinds from 24 tribes ranging from Canada to Argentina, and detected 135 polymorphic sites (Torroni *et al.*, 1992, 1993a, 1994a, b). One of the first outcomes obtained by this approach was the identification of an AluI site loss at np 5176 which characterized all the mtDNAs belonging to fourth haplogroup previously lacking specific markers. In addition, it was found that a DdeI site gain at np 10394 and an AluI site gain at np 10397 were always associated with the AluI site np 5176 and the HincII 13259 site losses. At that moment it became clear that all Native American mtDNAs were defined by one of only four sets of specific mutations and clustered in four well-defined haplogroups named A, B, C, and D. Haplogroup A mtDNAs is defined by the HaeIII site at np 663, haplogroup B mtDNAs by the COII/tRNA^{Lys} 9-bp deletion, haplogroup C mtDNAs by the HincII site loss at np 13259, and haplogroup D mtDNAs by the AluI site loss at np 5176. The presence of only four mtDNA groups was independently confirmed by sequencing about 360 np of the D-loop, the major noncoding mtDNA region which evolves faster than the mtDNA coding regions. Additional D-loop polymorphisms were found to specifically characterize each of the four haplogroups (Fig. 1). Since the mutations detected by both restriction analysis and D-loop sequencing are associated, they create complex haplotypes which remain recognizable even when reversions and/or parallel mutations occur (Torroni *et al.*, 1993a).

Comparison of Native American mtDNA variation showed that North, Central, and South American Amerinds harbor all four haplogroups, indicating that the Amerind tribes have a common ancestral origin. By contrast, the Na-Dene were found to harbor only haplogroup A mtDNAs with about one-third of their mtDNAs carrying a Na-Dene specific RsaI site loss at np 16329. Moreover, the amount of genetic variation accumulated within Na-Dene haplogroup A was 0.021, more than fourfold lower than that within Amerind haplogroup A (0.091%), and consistently much lower than that observed in haplogroup A of each individual Amerind

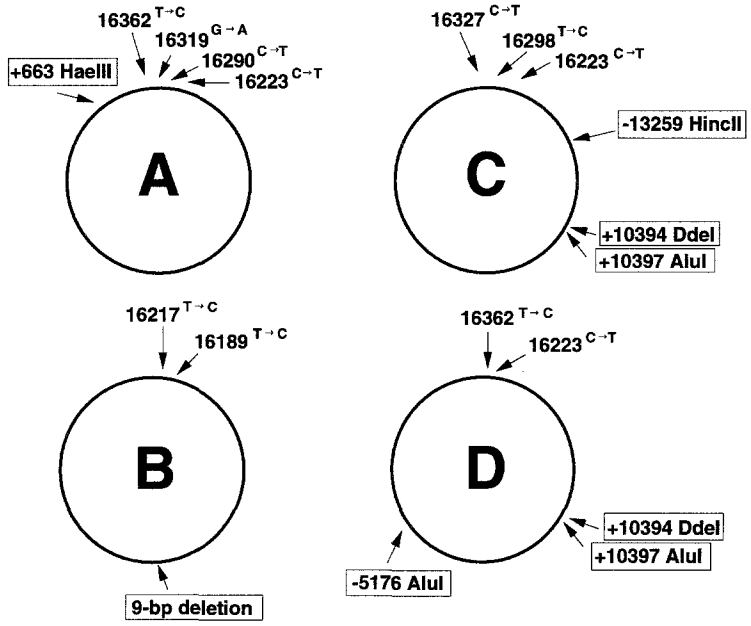

Fig. 1. Polymorphisms of the four Native American haplogroups. Circles represent typical haplogroup A, B, C, and D mtDNAs. Polymorphisms detected by haplotype analysis are boxed, while the other mutations are detected by sequencing and are located in the D-loop region. Because of the large number of haplogroup-specific polymorphisms, Native American mtDNAs can still be easily affiliated to one of the haplogroups even when reversions or parallel mutations occur.

tribe. These findings indicate that the Na-Dene were founded by a Siberian migration which carried only group A mtDNAs and acquired the RsaI np 16329 mutation in the process. This migration was independent and much more recent than the migration of the ancestral Amerinds (Torroni *et al.*, 1992).

The frequency distribution of the four haplogroups A, B, C, and D in Old World population is presented in Table I. None of the four haplogroups is found in Caucasians and Africans, while all four were found in Asians as expected since Native Americans are people of Asian ancestry. The complete absence of these haplogroups in non-Mongoloid populations allows an immediate identification of exogenous mtDNAs in Native American populations since virtually all of the gene flow from Old World populations is derived from Caucasians and Africans. It also permits an easy detection of those subjects who, although classified as "Caucasians" or "Afro-Americans," have Native American female ancestors on their maternal lineage.

The distribution of the four haplogroups in Asian populations provided additional information about the origin of Native Americans (Table I). Haplogroups A, C, and D show a marked cline of frequencies. They are completely absent in Southeast Asia and reach their highest frequencies in northeastern Siberian populations like the Yukagirs, the Evenks, and the Siberian Eskimos in which they represent virtually all of the mtDNAs. This observation not only confirms that Siberians derived from Central–East Asian populations and that, in turn, Native Americans derived from Siberians, but it also indicates that the expansion of humans, first into Siberia and then into the Americas, was associated with a progressive impoverishment of mtDNA variation. This extends the conclusion of early mtDNA data (Wallace *et al.*, 1985), indicating that population constrictions began already during the colonization of Siberia.

A striking anomaly in the distribution of one of the four haplogroups, haplogroup B, can be observed

Table I. Percent Frequencies of MtDNA Haplogroups A, B, C, and D in Human Populations

Populations	N	Haplogroups				
		A	B	C	D	Other
Siberia[1]						
Siberian Eskimos	50	80.0	—	—	20.0	—
Chukchi	24	37.5	—	16.7	16.7	29.2
Koryaks	46	23.9	—	21.7	8.7	45.6
Yukagirs	27	—	—	59.3	33.3	7.4
Evens	43	—	—	58.1	7.0	34.9
Nivkhs	57	—	—	—	28.1	71.9
Udegeys	46	—	—	19.6	—	80.4
Evenks	51	3.9	—	84.3	9.8	2.0
Nganasans	49	2.0	—	38.8	36.7	22.4
Sel'kups	20	—	—	35.0	—	65.0
East and Southeast Asia[2]						
Koreans	13	7.7	15.4	—	23.1	53.8
Han (Taiwan)	20	10.0	35.0	5.0	5.0	45.0
Tibetans	54	11.1	5.6	3.7	16.7	62.9
Han (Malaysia)	14	7.1	—	—	14.3	78.6
Vietnamese	28	—	14.3	—	—	85.7
Malays	14	—	14.3	—	—	85.7
Sabah Aborigines	32	—	18.8	—	—	81.2
Malay Aborigines	32	—	3.1	—	—	96.9
Oceania[3]						
New Guineans	119	—	19.3	—	0.8	79.9
Polynesians	150	ND	93.0	ND	ND	7.0
Europe[4]	175	—	—	—	—	100.0
Africa[5]	140	—	—	—	—	100.0

N is the number of subjects analyzed. "Other" indicates haplotypes not belonging to haplogroups A–D. (1) Torroni *et al.* (1993b); (2) Ballinger *et al.* (1992) and Torroni *et al.* (1994c); (3) Stoneking *et al.* (1990) and Hertzberg *et al.* (1989); (4) Torroni *et al.* (1994d); (5) Chen, Torroni, and Wallace, unpublished data. ND, not determined.

in Table I. Group B mtDNAs are widely distributed in East Asians (Ballinger *et al.*, 1992), and occur at very high frequencies in Melanesian (Stoneking *et al.*, 1990) and Polynesian (Hertzberg *et al.*, 1989) populations. They are also widely dispersed throughout the Amerinds of North, Central, and South America. Yet, all populations (Siberians, Eskimos, and Na-Dene) which live in the geographic area which separates the Central–East Asian and the Amerind populations completely lack haplogroup B mtDNAs. This intriguing data could indicate that the founding haplotypes of Amerind group A, B, C, and D mtDNAs were not carried to the Americas by a single founding population but could be the product of two separate migrations. The first migration carried those haplogroups (A, C, and D) which are commonly found in eastern Siberian populations. A second migration could have carried haplogroup B. The most likely route for such a second migration would be from East Asia along the coast of Siberia. By this route the group B migration could have avoided contacts with populations which already inhabited the tundra of eastern Siberia. After their arrival and expansion, the people that carried group B mtDNAs to the Americas would have then admixed with the descendants of those who carried group A, C, and D mtDNAs and generated the Amerinds. In such a scenario the extent of variation accumulated within Amerind haplogroup B should be lower than that observed in the other three Amerind haplogroups since a shorter time elapsed by the time of its arrival. Indeed the sequence divergence accumulated within mtDNA group B (0.034%) is much lower than that observed within mtDNA groups A, C, and D (0.075%). While a group B migration could have followed the first human migration to the Americas, however, it would have preceded the migration from Siberia which gave rise to the modern Na-Dene. Indeed, similarly to all Siberian populations, the Na-Dene lack B mtDNAs.

Having established that the Amerinds and the Na-Dene are the results of independent migrations from Siberia and that the Amerinds could have been originated by two independent migrations both preceding the arrival of the Na-Dene, it remained to define when these migrations occurred. Recently we have estimated that the average rate of evolution of human mtDNAs is between 2.2% and 2.9% per million years. This estimate has been obtained by using divergence times of the various Chibcha-speaking tribes of lower Central America developed from archaeological, linguistic, and genetic data by Barrantes *et al.* (1990), and the mtDNA variation accumulated in 110 subjects from seven Chibcha populations (Torroni *et al.*, 1994b). When this rate of evolution was applied to the sequence divergence accumulated in the Amerind haplogroups A, C, and D, it provided an arrival date for the first human migration to the Americas of between 26,000 and 34,000 years before present (YBP), for the group B migration of 12,000–15,000 YBP, and for the Na-Dene migration of 7,000–10,000 YBP (Fig. 2).

Analysis of Native American mtDNA variation not only allowed a reconstruction of human migrations to the Americas, but also permits a verification of the degree of overlap between genetics and culture. In his drastic linguistic reclassification of Native American languages, Greenberg (1987) proposed that all Amerind languages could be subdivided in only 11 major linguistic stocks. This classification has generated an acrimonious debate among linguists and anthropologists since many scholars do believe that the linguistic techniques used to aggregate the more than 100 "Amerind" distinct linguistic families in only 11 stocks are inadequate (Campbell and Mithun, 1979; Nichols, 1990). Some of the linguistic aggregations proposed by Greenberg were tested by comparing mtDNA variation observed in a large number of tribes from Middle America (Torroni *et al.*, 1994a). As may be expected, this analysis suggested that the truth is somewhere in between. For instance, mtDNA data support the aggregation of the Otomangueans (Mixtecs and Zapotecs of Southern Mexico) and Uto-Aztecans (Pima of Arizona) in a single stock named Central Amerind; however, they do not support the aggregation of the Mixe-Zoquean and the Mayan in the Penutian stock or the inclusion of the Yanomama in the Chibchan-Paezan stock.

These results indicate that mtDNA offers a new powerful tool for investigating complex physical and cultural anthropology questions. However, what is the correlation between these population data and disease studies, and how should these haplotype and phylogenetic data be utilized for mtDNA disease studies? An interesting approach to correlate population data and mtDNA disease studies has been used by Jun *et al.* (1994) to identify a mtDNA mutation associated with Leber's hereditary optic neuropathy (LHON) and dystonia in a five-generation Hispanic family from California.

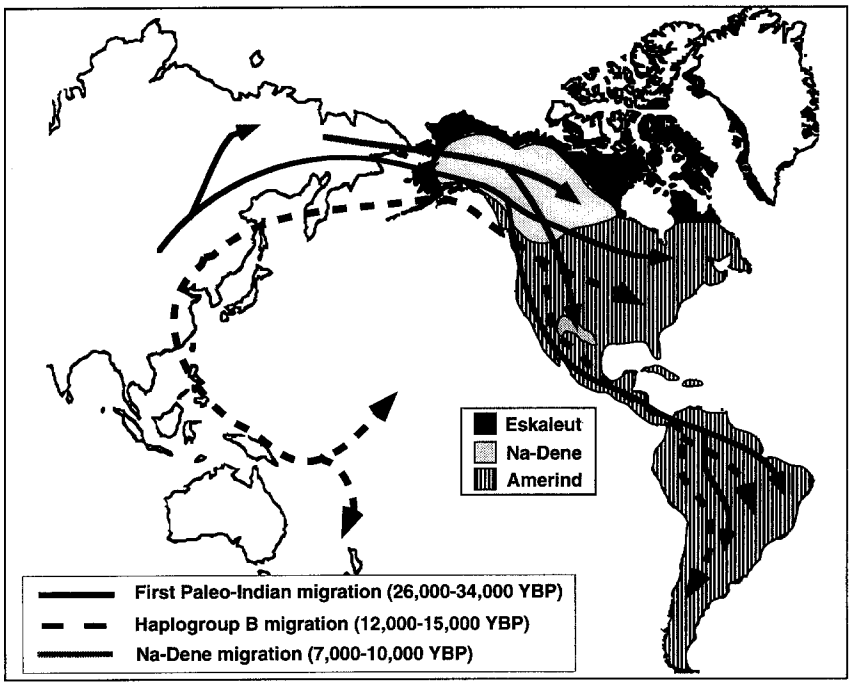

Fig. 2. Migrations from Asia which gave rise to Amerinds and Na-Dene. The first migration occurred between 26,000 and 34,000 years before present (YBP) and carried haplogroup A, C, and D mtDNAs to the Americas. A second migration may have occurred between 12,000 and 15,000 YBP and carried haplogroup B mtDNAs. This people mixed with the pre-existing populations derived from the first migration, and originated the Amerinds. A third more recent migration from Siberia carried only haplogroup A mtDNAs and gave rise to the Na-Dene. The Navajo and the Apache derive from Na-Dene who moved from Canada to the Southwest only about 1,000 YBP.

Sequencing of almost the entire proband's mtDNA revealed 40 nucleotide differences from the reference sequence (Anderson *et al.*, 1981). Among these, the mutations generating the np 5176 AluI site loss, the np 10394 DdeI site gain, and the np 10397 AluI site gain were found. Since these mutations characterize Native American haplogroup D mtDNAs (Fig. 1), they revealed that the mtDNA in question was Native American, a finding supported by phylogenetic analysis (Fig. 3). Such an identification is of great importance especially when the general population is constituted by numerous subpopulations and the ethnic background of single individuals is unclear. This is because most of the mtDNA variation is population-specific and common polymorphisms in one population can be completely absent in other populations. The assignment of the proband's mtDNA to a specific haplogroup identified the set of mtDNAs which are evolutionarily the most closely related and thus are the best population controls to be screened for the presence or absence of candidate mutations. Among the numerous mutations found in the proband's mtDNA, two best illustrate the importance of this approach. A C-to-T transition at np 2092 in the 16S rRNA gene, previously never described in human populations, was found in 35 out of 37 Amerind haplogroup D mtDNAs. A C-to-T transition at np 8414 which results in a leucine-to-phenylalanine substitution and was never described before, was observed in all haplogroup D mtDNAs tested. Obviously, these mutations represent private polymorphisms of haplogroup D. However, without the screening of the group D mtDNAs, it would have been difficult to rule out the possibility that these mutations had a pathological significance. On the contrary, the mtDNA mutation, to which the LHON plus dystonia phenotype in that family was finally attributed, was not only absent in the general population but was also absent in all group D control mtDNAs.

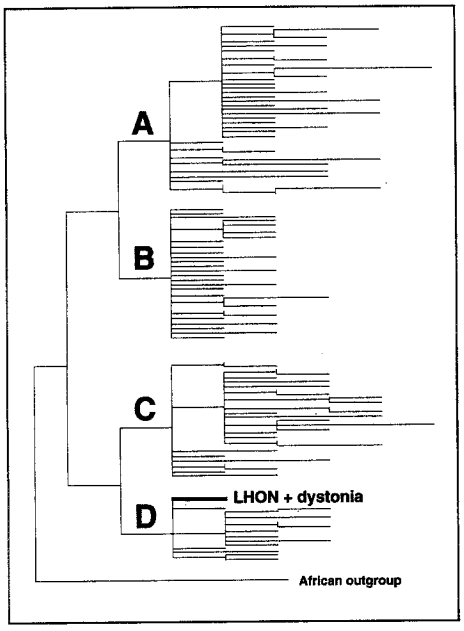

Fig. 3. Phylogenetic tree of Native American mtDNA haplotypes including the LHON + dystonia mtDNA from a Hispanic family. One hundred-twenty distinct haplotype from 508 Native Americans are aggregated in this tree (Torroni *et al.*, 1992, 1993a, 1994a, b). The haplotypes cluster in the four haplogroups A, B, C, and D, and the mtDNA associated with LHON + dystonia is a member of haplogroup D. The tree has been generated by PAUP analysis (Swofford, 1992), and the horizontal branch lengths are proportional to the number of mutational events that separate the haplotypes.

EUROPEAN mtDNA VARIATION

It is known that Europe was already populated by modern *Homo sapiens* populations around 40,000 YBP (Rigaud and Simek, 1990) and that these people probably derived from the morphologically modern humans (Proto-Cro-Magnons) who had been living in the Middle east since 80,000–100,000 YBP (Valladas *et al.*, 1988). A second population expansion from the Middle East into Europe started about 10,000 YBP. These people were Neolithic farmers and admixed with the local hunter–gatherers populations, generating a gradient of nuclear gene frequencies still observed in modern Europeans (Ammerman and Cavalli-Sforza, 1984; Cavalli-Sforza *et al.*, 1988; Sokal *et al.*, 1991; Piazza, 1993). One debated anthropological issue about the origin of modern Europeans is whether modern *Homo sapiens* populations genetically admixed with the Neanderthal populations which were coexisting at the same time in the Middle East and Europe (Bar-Yosef, 1992) and, therefore, whether a portion of the Neanderthal gene pool is present in modern Caucasian populations.

Analysis of mtDNA variation in Caucasians represent the ideal tool to clarify this issue. Unfortunately, despite the fact that thousands of Caucasians have been analyzed for mtDNA variation (Wallace *et al.*, 1992), till recently, very few Caucasian-specific mtDNA markers had been identified and all of them had very low frequencies. This is because most studies of mtDNA variation have employed a battery of six rare cutter restriction endonucleases and no high-resolution analyses have been carried out on large population samples.

Recently we have analyzed 175 mtDNAs from Caucasians living in the US and Canada (Torroni *et al.*, 1994d). This analysis led to the identification of four mtDNA haplogroups, named H, I, J, and K, each defined by sets of specific mutations which are absent in Africans and Mongoloids and represent more than 60% of Caucasian mtDNAs. Haplogroup H is defined by the lack of the AluI site at np 7025. Haplogroup I is defined by a large number of specific mutations including a DdeI site at np 1715, an AluI site at np 10028, an AvaII site at np 8249, a BamHI/MboI site at np 16389, and the absence of a HaeII site at np 4529. Haplogroup J is characterized by BstNI np 13704 and RsaI np 16065 site losses. Haplogroup K is defined by a concomitant HaeII np 9052/HhaI np 9053 site loss (Fig. 4). These haplogroups are the most divergent within the Caucasian phylogeny and, being characterized by mutations which are Caucasian-specific, they had to have originated in the ancestral Caucasians not long after their genetic separation from the ancestors of modern Asians and Africans. The most prevalent of the Caucasian haplogroups, haplogroup H, showed the highest intragroup divergence with a value of 0.102%; haplogroup I was next at 0.073%, and haplogroups J and K followed at about 0.040%. Using the 2.2–2.9%/myr mtDNA evolution rate, we obtained times of 35,000–46,000 YBP for haplogroup H, 25,000–33,000 YBP for I, 15,000–20,000 YBP for J, and 13,000–18,000 YBP for K. Estimates of the inter-haplogroup sequence divergences provided concordant divergence times to those obtained from the intra-group divergences. Since Neanderthal skeletal remains date back in Europe to at least 150,000 YBP (Bar-Yosef and Vandermeersch, 1993), these ages clearly indicated that none of four Caucasian-specific haplogroups originated in Neanderthal

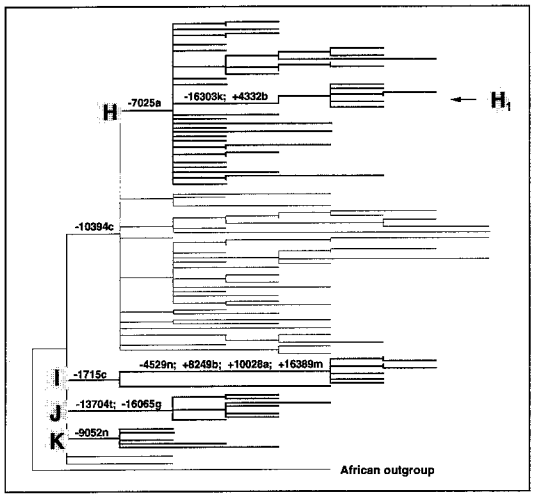

Fig. 4. Phylogenetic tree of Caucasian mtDNAs. This tree includes 117 distinct haplotypes observed in 175 Caucasians from the US and Canada. The capital letters H–K in shaded boxes indicate haplogroups defined by Caucasian specific mutations. The numbers in lower case letters indicate restriction sites defining the haplogroups. The restriction enzymes correspond to the following code: a, AluI; b, AvaII; c, DdeI; g, HinfI; k, RsaI; m, BamHI; n, HaeII; t, BstNI. An AvaII site gain at np 4332 has been associated with Alzheimer and Parkinson diseases (Shoffner *et al.*, 1993). That mutation defines a subset (H_1) of the mtDNAs within haplogroup H. All mtDNAs within haplogroup I share by descent a small homoplasmic insertion in the D-loop region which causes a recurrent 270 bp heteroplasmic duplication within the D-loop (Torroni *et al.*, 1994d). The np 15257 LHON mutation is usually found only on mtDNAs belonging to haplogroup J (Brown *et al.*, 1992).

populations. Therefore, the extent of mtDNA variation in modern Caucasians appears to refute the possibility that modern *Homo sapiens* admixed with Neanderthals while expanding in the Middle East and Europe. Moreover, since the radiation of *Homo erectus* in Europe and Asia was substantially earlier than the split of *Homo sapiens sapiens* and *Homo sapiens neanderthaliensis*, mtDNA data does not support the multiregional model for human racial origins (Wolpoff, 1989; Thorne and Wolpoff, 1992) but is consistent with a monogenesis and a recent origin of modern human populations (Johnson *et al.*, 1983; Cann *et al.*, 1987).

The definition of Caucasian-specific haplogroups has important implications for the understanding of the relationships between mtDNA mutations and mitochondrial pathologies since many of the disease mtDNA studies have been carried out on Caucasian subjects. One example is represented by Leber's hereditary optic neuropathy (LHON), the first human disease found to be caused by a mtDNA point mutation (Wallace *et al.*, 1988). The genetics of this disease has proven to be complex and heterogeneous as at least twelve mtDNA point mutations have now been associated with LHON (Brown *et al.*, 1994). Based on criteria such as nature of the amino acid change, penetrance and expressivity in LHON pedigrees, and presence in the general population, these mutations have been classified as high, moderate, or low-risk LHON mutations. The use of these criteria is often based on personal knowledge and experience and, therefore, is very subjective. Phylogenetic analysis provides a more objective way to classify mtDNA mutations. High-risk LHON mutations (nps 3460, 11778), i.e., mutations which can cause blindness by itself and have high degree of penetrance and expressivity, are under a relatively strong selective pressure. Therefore, they are likely to be lost together with the mtDNAs on which they occurred within few generations from their occurrence. However, these mutations are continuously generated in the general populations and continuously generate new LHON pedigrees. Since the new mutational events occur randomly, these high-risk mutations are expected to be found on heterogeneous mtDNA backgrounds and in all human populations. On the contrary, low-risk mutations which have low penetrance and expressivity and are under a milder selective pressure, are also found in the general population. Consequently, they are likely to be transmitted for numerous generations and, in particular demographic conditions, even to reach relatively high frequencies. Transmission through numerous generations makes it possible for a mutation generated by a single mutational event to be transmitted by descent to numerous pedigrees, which would appear to be unrelated but in fact share a common female ancestor. Since most of the pedigrees affected are indeed related, these mutations usually appear as population-specific. Phylogenetic analysis of mtDNA haplotypes of LHON patients proves that this is the case. Indeed, the high-risk LHON mutations at nps 3460 and 11778 are found associated with all mtDNA haplogroups within Caucasian, African, and Asian phylogenies, while the low-risk LHON mutation at np 15257 is observed only in Caucasians and is almost always found on mtDNAs which are members of the Caucasian-specific haplogroup J (Brown *et al.*, 1992) (Fig. 4). Being related, haplogroup J mtDNAs carrying the np 15257 mutation in

Leber's patients carry also numerous additional mutations by descent. It is then plausible that some of these associated mutations (nps 13708, 4216, 15812, 14484) represent an additional factor in modulating the penetrance and expressivity of the LHON phenotype in different pedigrees with the 15257 mutation as proposed by Brown *et al.* (1992).

MtDNA mutations have also been associated with late-onset degenerative diseases (Wallace and Lott, 1993). For instance, a A-to-G transition at np 4336 which generate an AvaII site gain at np 4332 has been found to be seven times more common in Alzheimer (AD) and Parkinson (PD) disease patients than in normal controls, and on the basis of several criteria has been hypothesized to be associated with AD and PD (Shoffner *et al.*, 1993). The late onset of AD and PD suggest that these diseases and the possibly associated mtDNA mutations are under limited, if any, selective pressure. Consequently, it would be expected that most of the subjects with the np 4332 mutation should carry it by descent. In addition, since these subjects are maternally related, the mutation should be limited to a single ethnic group. Indeed, population screenings have revealed that this mutation is limited to Caucasians, and phylogenetic analysis showed that all mtDNAs carrying this mutation are related, forming a specific sublineage (H_1) with the Caucasian-specific haplogroup H (Fig. 4).

An additional case stressing the importance of the haplotype and phylogenetic analyses to assess the pathological role of mtDNA mutation is exemplified by a postulated association between a 270 bp heteroplasmic duplication of the mtDNA D-loop region and deletions in mitochondrial myopathies (Brockington *et al.*, 1993). Haplotype analysis of the 175 Caucasian samples included in the phylogeny of Fig. 4 showed that all haplogroup I mtDNAs had a small homoplasmic insertion in the D-loop region (Torroni *et al.*, 1994d). This insertion adds 2–6 additional Cs to a stretch of 6 Cs at nps 568–573 and occurs within one of two identical repeats at np 302–308 (Box 1) and 567–573 (Box 2) which are separated by about 270 nps. The inserted Cs increase the homology between Boxes 1 and 2 and their surrounding regions. The screening for the 270 bp heteroplasmic duplication of the 175 Caucasians included in the phylogeny showed that all haplogroup I mtDNAs had the heteroplasmic duplication. A screening of chronic progressive external ophthalmoplegia (CPEO) and Kearns Sayre syndrome (KSS) patients found the duplication only in patients with haplogroup I mtDNAs. These findings lead to the conclusion that the duplication is caused by the extended homology between Boxes 1 and 2 in haplogroup I mtDNAs and is generated by slip replication. They also indicate that there is little if any correlation between these mutations and mitochondrial myopathy. This conclusion is further supported by the 25,000–33,000 year old age of haplogroup I. This haplogroup would be unlikely to have survived that long if the insertion of the Cs and the consequent heteroplasmic duplication predisposed individuals to KSS and hence was subject to strong negative selective pressure.

In conclusion, it is clear that haplotype and phylogenetic analysis of mtDNA variation have provided an essential contribution to define the origin and pattern of migration of human populations. However, it is now also clear that this type of analyses would provide new perspectives on our search for the etiology of mitochondrial pathologies.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM 46915 (DCW).

REFERENCES

Ammerman, A. J., and Cavalli-Sforza, L. L. (1984). *The Neolithic Transition and the Genetics of Populations in Europe*, Princeton University Press, Princeton, New Jersey.

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

Ballinger, S. W., Schurr, T. G., Torroni, A., Gan, Y. Y., Hodge, J. A., Hassan, K., Chen, K. H. *et al.* (1992). "Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations", *Genetics* **130,** 139–152.

Bandy, B., and Davison, A. J. (1990). "Mitochondrial mutations may increase oxidative stress: implications for carcinogenesis and aging? *Free Radical Biol. Med.* **8,** 523–539.

Barrantes, R., Smouse, P. E., Mohrenweiser, H. W., Gershowitz, H., Azoffifa, J., Arias, T. D., and Neel, J. V. (1990). "Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity", *Am. J. Hum. Genet.* **46,** 63–84.

Bar-Yosef, O. (1992). In *Continuity or Replacement: Controversies*

in Homo sapiens evolution (Brauer, G., and Smith, F., eds.), A. A. Balkema, Netherlands, pp. 261–272.

Bar-Yosef, O., and Vandermeersch, B. (1993). "Modern humans in the Levant", *Sci. Am.* **268(4),** 94–100.

Blanc, H., Chen, K. H., D'Amore, M. A., and Wallace, D. C. (1983). "Amino acid change associated with the major polymorphic HincII site of Oriental and Caucasian DNAs", *Am. J. Hum. Genet.* **35,** 167–176.

Brockington, M., Sweeney, M. G., Hammans, S. R., Morgan-Hughes, J. A., and Harding, A. (1993). "A tandem duplication in the D-loop of human mitochondrial DNAs is associated with deletions in mitochondrial myopathies", *Nature Genet.* **4,** 67–71.

Brown, M. D., Voljavec, A. S., Lott, M. T., Torroni, A., Yang, C-C., and Wallace, D. C. (1992). "Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy", *Genetics* **130,** 163–173.

Brown, M. D., Lott, M. T., and Wallace, D. C. (1994). "Mitochondrial DNA mutations and the eye", in *Molecular Genetics of Inherited Eye Disorders* (Jay, B., and Wright, A., eds.), Harwood Academic, Switzerland.

Campbell, L., and Mithun, M. (1979). *The languages of Native America*, University of Texas Press, Austin, Texas.

Cann, R. L., Stoneking, M., and Wilson, A. C. (1987). "Mitochondrial DNA and human evolution", *Nature (London)* **325,** 31–36.

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., and Mountain, J. L. (1988). "Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data", *Proc. Natl. Acad. Sci. USA* **85,** 6002–6006.

Giles, R. E., Blanc, H., Cann, H. M., and Wallace, D. C. (1980). "Maternal inheritance of human mitochondrial DNA", *Proc. Natl. Acad. Sci. USA* **77,** 6715–6719.

Greenberg, J. (1987). *Language in the Americas*, Stanford University Press, Stanford, California.

Hertzberg, M., Mickleson, K. N. P., Serjeantson, S. W., Prior, J. F., and Trent, R. J. (1989). "An Asian-specific 9-base pair deletion of mitochondrial DNA is frequently found in Polynesians", *Am. J. Hum. Genet.* **44,** 504–510.

Johnson, M. J., Wallace, D. C., Ferris, S. D., Rattazzi, M. C., and Cavalli-Sforza, L. L. (1983). "Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns", *J. Mol. Evol.* **19,** 255–271.

Jun, A., Brown, M. D., and Wallace, D. C. (1994). "A mitochondrial DNA mutation at np 14459 of the ND6 gene associated with maternally inherited Leber's hereditary optic neuropathy and dystonia", *Proc. Natl. Acad. Sci. USA*, in press.

Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M., and Koike, K. (1982). "Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes", *J. Mol. Evol.* **19,** 28–35.

Nichols, J. (1990). "Linguistic diversity and the first settlement of the New World", *Language* **66,** 475–521.

Piazza, A. (1993). "Who are the Europeans?" *Science* **260,** 1767–1769.

Rigaud, J. P., and Simek, J. F. (1990). The last pleniglacial in the south of France, in *The World at 18,000 BP* (Soffer, O., and Gamble, C., ed.), Unwin Hyman, London.

Schurr, T. G., Ballinger, S. W., Gan, Y. Y., Hodge, J. A., Merriwether, D. A., Lawrence, D. N., Knowler, W. C. *et al.* (1990). "Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages", *Am. J. Hum. Genet.* **46,** 613–623.

Shoffner, J. M., Brown, M. D., Torroni, A., Lott, M. T., Cabell, M. F., Mirra, S. S., Beal, M. F. *et al.* (1993). Mitochondrial DNA variants observed in Alzheimer disease and Parkinson disease patients", *Genomics* **17,** 171–184.

Sokal, R. R., Oden, N. L., and Wilson, C. (1991). "Genetic evidence for the spread of agriculture in Europe by demic diffusion", *Nature (London)* **351,** 143–145.

Stoneking, M., Jorde, L. B., Bhatia, K., and Wilson, A. C. (1990). "Geographic variation in human mitochondrial DNA from Papua New Guinea", *Genetics* **124,** 717–733.

Swofford, D. (1992). Phylogenetic Analysis Using Parsimony (PAUP), Version 3.0s. Illinois Natural History Survey, Champaign, Illinois.

Thorne, A. G., and Wolpoff, M. H. (1992). "The multiregional evolution of humans", Sci. Am. **266(4),** 76–83.

Torroni, A., Schurr, T. G., Yang, C-C., Szathmary, E. J. E., Williams, R. C., Schanfield, M. S., Troup, G. A. *et al.* (1992). "Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations", *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., and Wallace, D. C. (1993a). "Asian affinities and continental radiation of the four founding Native American mitochondrial DNAs", *Am. J. Hum. Genet.* **53,** 563–590.

Torroni, A., Sukernik, R. I., Schurr, T. G., Starikovskaya, Y. B., Cabell, M. F., Crawford, M. H., Comuzzie, A. G. *et al.* (1993b). "Mitochondrial DNA variation of Aboriginal Siberians reveals distinct genetic affinities with Native Americans", *Am. J. Hum. Genet.*, **53,** 591–608.

Torroni, A., Chen, Y-S., Semino, O., Santachiara-Benerecetti, A. S., Scott, R. C., Lott, M. T., Winter, M., and Wallace, D. C. (1994a). "Mitochondrial DNA and Y-chromosome polymorphisms in four Native American populations from Southern Mexico", *Am. J. Hum. Genet.*, **54,** 303–318.

Torroni, A., Neel, J. V., Barrantes, R., Schurr, T. G., and Wallace, D. C. (1994b). "A mitochondrial DNA 'Clock' for the Amerinds and its implications for timing their entry into North America", *Proc. Natl. Acad. Sci. USA*, **91,** 1158–1162.

Torroni, A., Miller, J. A., Moore, L. G., Zamudio, S., Zhuang, J., Droma, T., Wallace, D. C. (1994c). "Mitochondrial DNA analysis in Tibet. Implications for the origin of the Tibetan population and its adaptation to high altitude", *Am. J. Phys. Anthropol.*, **93,** 189–199.

Torroni, A., Lott, M. T., Cabell, M. F., Chen, Y-S., Lavergne, L., and Wallace, D. C. (1994d). "Mitochondrial DNA and the origin of Caucasians. Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region", *Am. J. Hum. Genet.*, submitted.

Valladas, H., Reyss, J. L., Joron, J. L., Valladas, G., Bar-Josef, O., and Vandermeersch, B. (1988). "Thermoluminescence dating of Mousterian 'Proto-Cro-Magnon' remains from Israel and the origin of modern man", *Nature (London)* **331,** 614–616.

Wallace, D. C., and Lott, M. T. (1993). "Maternally inherited diseases", In *Mitochondrial DNA in Human Pathologies* (Di Mauro, S. and Wallace, D. C., eds.), Raven Press, New York, pp. 63–83.

Wallace, D. C., Garrison, K., and Knowler, W. C. (1985). "Dramatic founder effects in Amerindian mitochondrial DNAs", *Am. J. Phys. Anthropol.* **68,** 149–155.

Wallace, D. C., Ye, J. H., Neckelmann, S. N., Singh, G., Webster, K. A., and Greenberg, B. D. (1987). "Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes substain seventeen times more mutations", *Curr. Genet* **12,** 81–90.

Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M., Elsas, L. J., and Nikoskelainen, E. K. (1988). "Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy", *Science* **242,** 1427–1430.

Wallace, D. C., Lott, M. T., Torroni, A., and Brown, M. D. (1993). *Report of the committee on human mitochondrial DNA. Chromosome Coordinating Meeting 1992* (Cuticchia, A. J., Pearson, P. L., Klinger, H. P., ed). Genome Priority Reports, Vol. 1, pp. 727–757. Basel, Karger.

Wolpoff, M. H. (1989). "Multiregional evolution. The fossil alternative to Eden", In *The Human Revolution—Behavioural and Biological Perspectives on the Origins of Modern Humans* (Mellars, P., and Stringer, C., eds.), Edinburgh University Press, Edinburgh, England, pp. 62–108.