Am. J. Hum. Genet. 57:133–149, 1995

Analysis of mtDNA Variation in African Populations Reveals the Most Ancient of All Human Continent-Specific Haplogroups

Yu-Sheng Chen,[1] Antonio Torroni,[1]* Laurent Excoffier,[2] A. Silvana Santachiara-Benerecetti,[3] and Douglas C. Wallace[1]

[1]Department of Genetics and Molecular Medicine, Emory University School of Medicine, Atlanta; [2]Department of Anthropology and Ecology, University of Geneva, Geneva; and [3]Department of Genetics and Microbiology, University of Pavia, Pavia

Summary

mtDNA sequence variation was examined in 140 Africans, including Pygmies from Zaire and Central African Republic (C.A.R.) and Mandenkalu, Wolof, and Pular from Senegal. More than 76% of the African mtDNAs (100% of the Pygmies and 67.3% of the Senegalese) formed one major mtDNA cluster (haplogroup L) defined by an African-specific *Hpa*I site gain at nucleotide pair (np) 3592. Additional mutations subdivided haplogroup L into two subhaplogroups, each encompassing both Pygmy and Senegalese mtDNAs. A novel 12-bp homoplasmic insertion in the intergenic region between $tRNA^{Tyr}$ and cytochrome oxidase I (COI) genes was also observed in 17.6% of the Pygmies from C.A.R. This insertion is one of the largest observed in human mtDNAs. Another 25% of the Pygmy mtDNAs harbored a 9-bp deletion between the cytochrome oxidase II (COII) and $tRNA^{Lys}$ genes, a length polymorphism previously reported in non-African populations. In addition to haplogroup L, other haplogroups were observed in the Senegalese. These haplogroups were more similar to those observed in Europeans and Asians than to haplogroup L mtDNAs, suggesting that the African mtDNAs without the *Hpa*I np 3592 site could be the ancestral types from which European and Asian mtDNAs were derived. Comparison of the intrapopulation sequence divergence in African and non-African populations confirms that African populations exhibit the largest extent of mtDNA variation, a result that further supports the hypothesis that Africans represent the most ancient human group and that all modern humans have a common and recent African origin. The age of the total African variation was estimated to be 101,000–133,000 years before present (YBP), while the age of haplogroup L was estimated at 98,000–130,000 YBP. These values substantially exceed the ages of all Asian- and European-specific mtDNA haplogroups.

Received February 8, 1995; accepted for publication March 31, 1995.

Address for correspondence and reprints: Dr. Douglas C. Wallace, Department of Genetics and Molecular Medicine, Emory University School of Medicine, Atlanta, GA 30322.

*Current address: Dipartimento di Genetica e Biologia Molecolare, Universita' "La Sapienza," Rome.

0002-9297/95/5701-0017$02.00

Introduction

African populations were among the first human groups to be analyzed for mtDNA variation by restriction analysis (Denaro et al. 1981). Since that time, considerable data have been collected on mtDNA restriction endonuclease site variation in several African populations (Scozzari et al. 1988, 1994; Soodyall and Jenkins 1992, 1993). However, with the exception of Cann et al. (1987), who analyzed 20 Afro-Americans, all mtDNA analyses of Africans using RFLP methods have employed a set of six rare cutter endonucleases and the Southern blot technique. Although this procedure allowed the screening of only 2%–3% of the mtDNA sequence variation, it revealed that between 60% and 100% of the sub-Saharan African mtDNAs were characterized by a *Hpa*I site gain at nucleotide pair (np) 3592. This marker is found at very low frequencies outside Africa, mostly in those populations (Arabs and southern Italians) that are historically known to have admixed with Africans (Bonne-Tamir et al. 1986; De Benedictis 1989; Semino et al. 1989; Ritte et al. 1993). Furthermore, the mtDNAs defined by the *Hpa*I site at np 3592 form a group of mtDNA haplotypes (haplogroup) that is the most divergent in the world mtDNA phylogeny. This has contributed to the hypothesis of an African origin of the human mtDNAs (Johnson et al. 1983; Scozzari et al. 1988, 1994; Soodyll and Jenkins 1992, 1993), though other interpretations of this data have been put forward (Excoffier and Langaney 1989; Templeton 1992).

Recently, a high-resolution methodology for analyzing mtDNA restriction-site variation from PCR fragments has been developed. This method allows the screening of ~15%–20% of mtDNA sequences and has been applied to screen extensive numbers of mtDNAs from Asia, Europe, and the Americas (Schurr et al. 1990; Ballinger et al. 1992; Torroni et al. 1992, 1993*a*, 1993*b*, 1994*a*, 1994*b*, 1994*c*, 1994*d*). These studies have revealed that the large majority of European, Asian, and Native American mtDNAs are defined by one or more continent-specific polymorphisms. The high frequency of these continent-specific mutations within one major continental group and their specificity to either Europeans, Asians, or Native Americans make them powerful

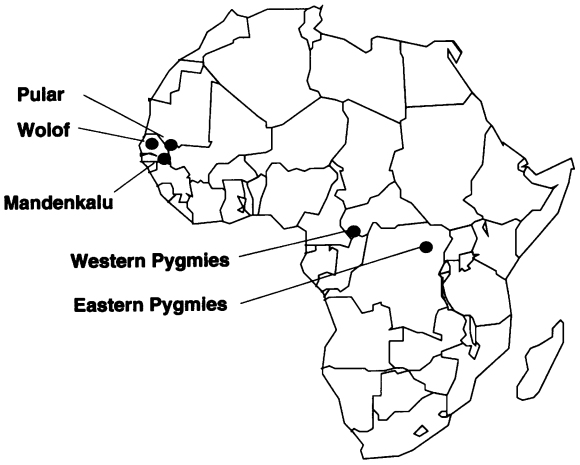


Figure 1 Map locations of the African populations analyzed for mtDNA variation. For Senegalese only, the more-represented ethnic groups examined are indicated.

genetic markers for inferring the ethnic and geographic origin of modern and ancient humans (Torroni et al., in press). Their geographic and ethnic specificity also indicates that these mutations arose after the genetic separation of the ancestral human populations that gave rise to the major modern human ethnic groups.

Phylogenetic analyses of the European-, Asian-, and Native American–specific polymorphisms define clusters of mtDNAs haplotypes (haplogroups). The largest of the haplogroups of each continent is usually the most divergent and, thus, the oldest within each major continental group (Torroni et al. 1993*a*, 1994*b*, 1994*c*; Wallace et al. 1994). Therefore, these haplogroups reflect the early stages in the radiation of the major human continental groups.

The only major human continental group that has not been analyzed by the high-resolution restriction methodology is the sub-Saharan Africans. Consequently, even though an African-specific mtDNA haplogroup has been defined by low-resolution restriction analysis, comparative analyses between this haplogroup and those observed in Asia, Europe, and the Americas are not currently possible. To overcome this limitation, we carried out a high-resolution restriction analysis of mtDNA variation in a large number of Africans.

Subjects and Methods

Subjects

Of the 140 unrelated Africans analyzed, 39 were Pygmies (22 Eastern Pygmies from Zaire and 17 Western Pygmies from Central African Republic [C.A.R.]), and the remaining 101 samples were collected in Senegal. Among the Senegalese samples there were 60 Niokolo Mandenkalu, 20 Wolof, 8 Pular, and 13 subjects belonging to other ethnic groups (5 Tukulor, 3 Maure, 2 Serere, 1 Bambara, 1 Sosse, and 1 from the Gambia) (fig. 1). A detailed description of the two Pygmy populations can be found in the work of Cavalli-Sforza (1986*b*). The Mandenkalu analyzed are a subset of the larger sample analyzed by Graven et al. (in press). All other Senegalese samples were chosen randomly from among the 186 previously analyzed by Scozzari et al. (1988). DNAs were extracted from either buffy coats (Wolof, Pular, and other Senegalese) or lymphoblast cell lines (Mandenkalu and Pygmies).

mtDNA Molecular Analysis

The entire mtDNA of each sample was amplified in nine overlapping fragments by PCR using the primer pairs and amplification conditions described in Torroni et al. (1992; 1993*a*). Each PCR segment was digested with 14 restriction endonucleases (*Alu*I, *Ava*II, *Bam*HI, *Dde*I, *Hae*II, *Hae*III, *Hha*I, *Hin*cII, *Hin*fI, *Hpa*I, *Msp*I, *Mbo*I, *Rsa*I, and *Taq*I). The resulting fragments were resolved through electrophoresis in NuSieve plus SeaKem agarose (FMC BioProducts) gels and visualized by UV-induced fluorescence. This restriction analysis screen of ~15%–20% of the mtDNA sequence for variation and permitted the definition of the mtDNA haplotypes for each individual (appendix).

The nature of some length polymorphisms detected by restriction analysis were further investigated by direct sequencing from PCR fragments, using Ampli*Taq* DNA polymerase with fluorescent terminator chemistry (Prism Ready Reaction Cycle sequencing kit; Applied Biosystem) and a 373A automated sequencer (Applied Biosystem). Some restriction-site losses found to occur multiple times on different branches of phylogenetic trees were also sequenced to determine whether they were generated by the same or by different nucleotide substitutions.

Parsimony Analysis

The phylogenetic relationships between haplotypes were inferred using parsimony (PAUP 3.1; Swofford 1993). The dendrograms were rooted in three different ways: by midpoint rooting, by using a Malay aborigine mtDNA haplotype characterized by the Asian-specific *Alu*I mutation at np 10397 (AS24; Ballinger et al. 1992; Torroni et al. 1994*c*) as an Asian outgroup reference, and by using a European mtDNA haplotype of the European-specific haplogroup I (CA85; Torroni et al. 1994*b*) as a European outgroup reference. Maximum-parsimony (MP) trees were generated through random addition of sequences using the tree bisection and reconnection (TBR) algorithm and saving no more than 10 MP trees for each replication. Because of the large number of terminal taxa, a large number of MP trees could be obtained. When using midpoint rooting, we terminated the search at 3,000 MP trees after 25,667 replications. When using the Asian haplotype as outgroup, we obtained 1,000 MP trees after 24,012 replications. When

using the European haplotype as outgroup, we obtained 3,000 MP trees after 3,461 replications (trees not shown). Although shorter trees could exist for each of these searches, none were observed in our analyses. Strict-consensus trees encompassing all MP trees were also obtained for each outgroup. Most of the relationships between haplotypes observed in the MP trees were also retained in the strict-consensus trees.

Sequence Divergence Estimations

Intra- and intersequence divergence estimations were calculated from restriction-analysis data, using the maximum-likelihood procedure of Nei and Tajima (1983). When calculating the divergence times, we used the range of 2.2%–2.9%/million years (Myr), estimated from the sequence divergence accumulated in the Chibchan-speaking populations of Central America (Torroni et al. 1994*d*).

Results

African Haplotypes

Seventy-nine haplotypes (AF1–AF79), defined by 119 restriction-site polymorphisms and 2 length polymorphisms, were observed among the 140 samples analyzed (appendix). Only 22 of these haplotypes were observed in more than one subject, and none was shared between either the Pygmy and the Senegalese populations or between the Eastern and Western Pygmy populations (table 1). The length polymorphisms were the COII/$tRNA^{Lys}$ 9-bp intergenic deletion, previously reported in other non-African human populations (Schurr et al. 1990; Stoneking et al. 1990; Ballinger et al. 1992; Torroni et al. 1992; Passarino et al. 1993); and a novel 10–12-bp poly-C insertion between the COI and $tRNA^{Tyr}$ genes, located in the same region as an insertion reported in Australian aborigine (Cann and Wilson 1983). A *Dde*I site at np 10394 was found in all but five haplotypes (AF1–AF3, AF62, and AF63) (table 1). This polymorphism appears to be very ancient, as it was observed in all world populations and tends to segregate human mtDNAs into two major groupings in continent-specific and global phylogenetic analyses (Torroni and Wallace 1994). Fifty-five of the 79 haplotypes (AF25–AF79) were defined by the African-specific *Hpa*I site gain at np 3592. Two additional mutations were found that subdivided all mtDNAs carrying the np 3592 *Hpa*I mutation: a *Hin*fI site gain at np 10806 due to a T-to-C transition at np 10810 and a combined *Hin*fI site gain at np 16389 and an *Ava*II site loss at np 16390 due to a G-to-A transition at np 16390. The first site was found in 21 haplotypes (AF25 and AF59–AF79) and 34.3% of the African samples, the latter site was present in 33 haplotypes (AF26–AF58) and 42.1% of the African samples.

Phylogenetic Analyses

Phylogenetic analyses indicated that all haplotypes carrying the np 3592 *Hpa*I site gain cluster into one major haplogroup (figs. 2 and 3) designated haplogroup L to conform to the nomenclature of Torroni et al. (1994*c*). The *Hin*fI site gains at nps 10806 and 16389 subdivide haplogroup L into two subhaplogroups, L1 and L2, each encompassing mtDNAs from the Senegalese and Western and Eastern Pygmies (fig. 1). Haplogroup L and subhaplogroups L1 and L2 were retained in all strict-consensus trees (figs. 2 and 3), indicating the robustness of this cluster and its monophyletic origin. One-hundred percent of the Pygmy mtDNAs cluster in haplogroup L, while 67.3% of the Senegalese mtDNAs clustered in haplogroup L. The remaining third of the Senegalese mtDNAs formed four additional small haplogroups defined by characteristic mutations (figs. 2 and 3).

Length Polymorphisms

A deletion of one of the 9-bp repeats located between the COII and $tRNA^{Lys}$ genes was found to be associated with two haplotypes, AF60 and AF61. Both of these belonged to subhaplogroup L1 and were limited to the Pygmies. AF60 was present in 27.3% of the Eastern Pygmies, while AF61 is present in 23.5% of the Western Pygmies. The similarity between AF60 and AF61 indicates that the 9-bp deletion arose as a single mutational event prior to the radiation of the two haplotypes.

This 9-bp deletion has previously been reported in several non-African populations (Wrischnik et al. 1987; Hertzberg et al. 1989; Stoneking and Wilson 1989; Schurr et al. 1990; Ballinger et al. 1992; Shields et al. 1992; Torroni et al. 1992). However, its association with the African-specific haplogroup L indicates that the 9-bp deletion observed in the Pygmies occurred independently from that observed in non-African populations. This finding confirms our previous evidence that this deletion has occurred independently several times during human mtDNA radiation (Ballinger et al. 1992; Torroni et al. 1993*a*, 1993*b*, 1994*a*) and indicates that the 9-bp deletion should be used only as an anthropological marker when viewed in the context of supporting haplotype information.

A novel poly-C insertion in the intergenic region between $tRNA^{Tyr}$ and COI was found in three Western Pygmies (17.6%) and defines haplotype AF66. This insertion was mapped at nps 5895–5899 and is constituted by a stretch of 10–12 Cs (fig. 5), which expands a stretch of 5 Cs that are normally present at this location. Because it creates a homogeneous sequence of 15–17 Cs, which affects in vitro DNA polymerase amplification and sequencing, the exact length of the insertion could not be unambiguously determined by either manual or automated sequencing procedures. To date, this insertion is the largest poly-C insertion observed in human

Table 1

mtDNA Haplotypes in African Populations

Haplotype	3592 *Hpa*I[a]	10394 *Dde*I[b]	Mandenkalu	Wolof	Pular	Other Senegalese[c]	Eastern Pygmies (Zaire)	Western Pygmies (C.A.R.)	No. Total
AF1	−	−	1	−	−	−	−	−	1
AF2	−	−	−	−	−	3	−	−	3
AF3	−	−	−	−	1	−	−	−	1
AF4	−	+	−	1	−	1	−	−	2
AF5	−	+	3	−	−	−	−	−	3
AF6	−	+	1	−	−	−	−	−	1
AF7	−	+	−	−	−	1	−	−	1
AF8	−	+	−	1	−	−	−	−	1
AF9	−	+	1	−	−	−	−	−	1
AF10	−	+	−	−	1	−	−	−	1
AF11	−	+	2	−	−	−	−	−	2
AF12	−	+	1	−	−	−	−	−	1
AF13	−	+	−	1	−	−	−	−	1
AF14	−	+	2	−	−	1	−	−	3
AF15	−	+	−	−	1	−	−	−	1
AF16	−	+	−	1	−	−	−	−	1
AF17	−	+	−	1	−	−	−	−	1
AF18	−	+	1	−	−	−	−	−	1
AF19	−	+	−	1	−	−	−	−	1
AF20	−	+	1	−	−	−	−	−	1
AF21	−	+	−	−	−	1	−	−	1
AF22	−	+	2	−	−	−	−	−	2
AF23	−	+	1	−	−	−	−	−	1
AF24	−	+	−	−	−	1	−	−	1
AF25	+	+	−	−	−	−	2	−	2
AF26	+	+	1	−	−	−	−	−	1
AF27	+	+	−	1	−	−	−	−	1
AF28	+	+	−	1	−	−	−	−	1
AF29	+	+	1	−	−	−	−	−	1
AF30	+	+	1	−	−	−	−	−	1
AF31	+	+	1	−	−	−	−	−	1
AF32	+	+	2	2	−	−	−	−	4
AF33	+	+	−	−	1	−	−	−	1
AF34	+	+	1	−	−	−	−	−	1
AF35	+	+	1	1	−	−	−	−	2
AF36	+	+	1	−	−	−	−	−	1
AF37	+	+	−	−	−	−	4	−	4
AF38	+	+	−	−	−	−	6	−	6
AF39	+	+	−	−	−	−	1	−	1
AF40	+	+	−	−	−	−	1	−	1
AF41	+	+	−	−	−	−	1	−	1
AF42	+	+	−	−	−	−	1	−	1
AF43	+	+	−	−	−	−	−	1	1
AF44	+	+	−	−	−	−	−	1	1
AF45	+	+	−	−	−	1	−	−	1
AF46	+	+	−	2	1	−	−	−	3
AF47	+	+	−	1	−	−	−	−	1
AF48	+	+	1	−	−	−	−	−	1
AF49	+	+	11	−	−	−	−	−	11
AF50	+	+	1	−	−	−	−	−	1
AF51	+	+	−	−	−	1	−	−	1
AF52	+	+	−	1	−	−	−	−	1
AF53	+	+	−	1	−	−	−	−	1
AF54	+	+	2	−	−	−	−	−	2
AF55	+	+	1	−	−	−	−	−	1
AF56	+	+	−	−	1	−	−	−	1

(*continued*)

Table 1 (continued)

Haplotype	3592 *HpaI*[a]	10394 *DdeI*[b]	Mandenkalu	Wolof	Pular	Other Senegalese[c]	Eastern Pygmies (Zaire)	Western Pygmies (C.A.R.)	No. Total
AF57	+	+	–	–	–	1	–	–	1
AF58	+	+	2	–	–	–	–	–	2
AF59	+	+	1	–	–	–	–	–	1
AF60	+	+	–	–	–	–	6	–	6
AF61	+	+	–	–	–	–	–	4	4
AF62	+	–	–	–	–	–	–	4	4
AF63	+	–	–	–	–	–	–	1	1
AF64	+	+	–	–	–	1	–	–	1
AF65	+	+	3	–	–	–	–	–	3
AF66	+	+	–	–	–	–	–	3	3
AF67	+	+	–	–	–	–	–	1	1
AF68	+	+	–	–	–	–	–	1	1
AF69	+	+	–	–	–	–	–	1	1
AF70	+	+	–	1	–	–	–	–	1
AF71	+	+	8	1	1	–	–	–	10
AF72	+	+	–	1	–	–	–	–	1
AF73	+	+	–	–	1	–	–	–	1
AF74	+	+	–	1	–	–	–	–	1
AF75	+	+	1	–	–	–	–	–	1
AF76	+	+	–	–	–	1	–	–	1
AF77	+	+	1	–	–	–	–	–	1
AF78	+	+	1	–	–	–	–	–	1
AF79	+	+	2	–	–	–	–	–	2
Total			60	20	8	13	22	17	140

[a] Indicates the presence or absence of the African-specific *HpaI* polymorphism at np 3592.

[b] Indicates the presence or absence of the *DdeI* state at np 10394.

[c] Among the other Senegalese, AF2 was found in three Tukulor, AF4 in one Serere, AF7 in one Maure, AF14 in one individual from the Gambia, AF21 in one Serere, AF24 in one Maure, AF45 in one Tukulor, AF51 in one Tukulor, AF57 in one Bambara, AF64 in one Maure, and AF76 in one Sosse.

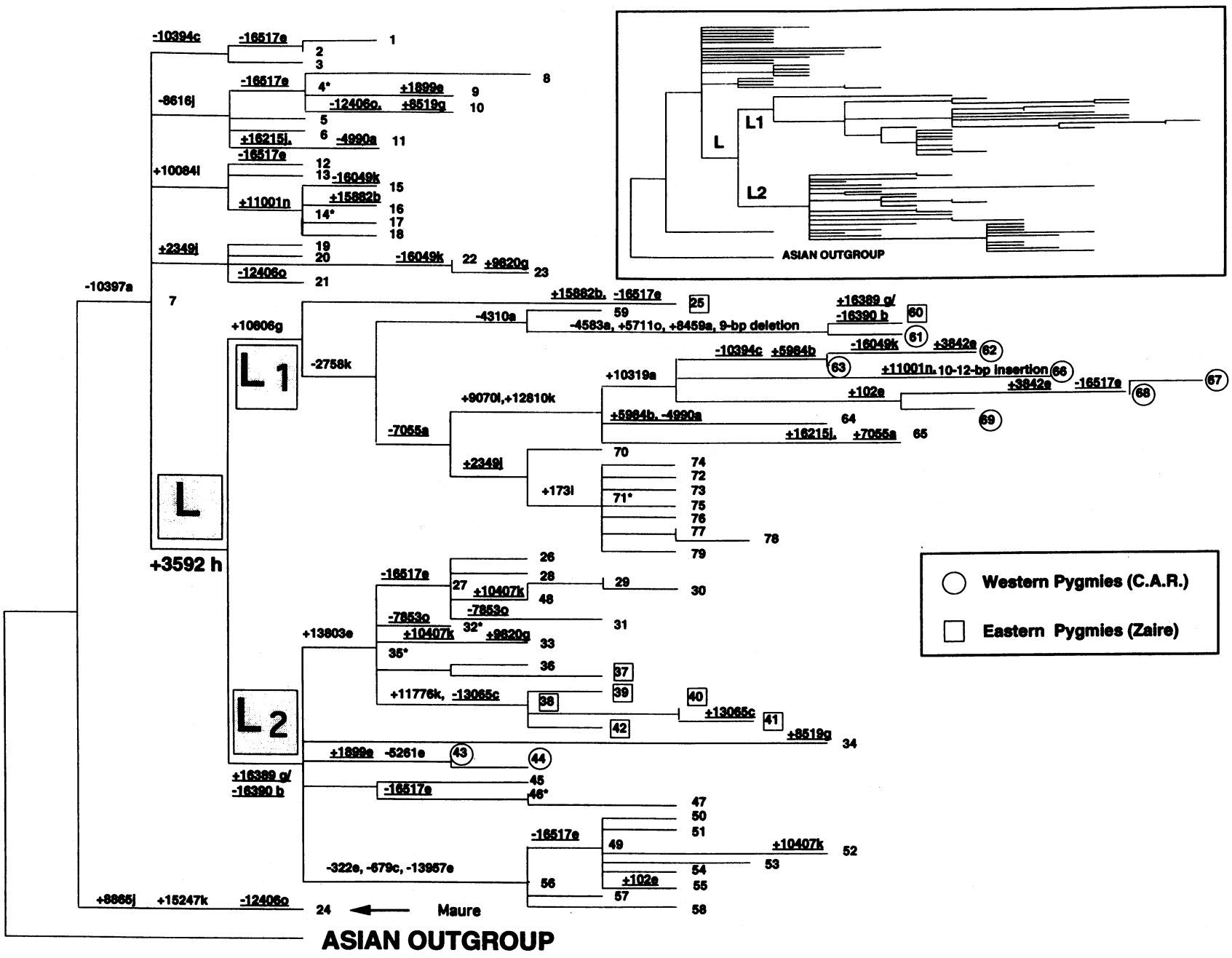
mtDNA, and its association with only one Western Pygmy haplotype suggests that it occurred recently.

Pygmy mtDNAs

Although all Pygmy mtDNAs were members of haplogroup L, the nature of mtDNA variation observed in Western and Eastern Pygmies clearly distinguishes the two populations. Almost 65% of the Western Pygmy mtDNAs (AF62, AF63, and AF66–AF69), but no Eastern Pygmies, form a mtDNA haplotype cluster defined by a *TaqI* site gain at np 9070 and a *RsaI* site gain at np 12810 (fig. 1). This haplogroup includes two haplotypes (AF64 and AF65) found in the Senegalese and shares numerous ancient mutations with other members of subhaplogroup L1. However, all Western Pygmy mtDNA in this small cluster are also defined by an additional *AluI* site at np 10319, which distinguishes them from the Senegalese and could represent a Western Pygmy private polymorphism, although this possibility needs to be confirmed by the analysis of other African populations.

Two other haplotypes, AF43 and AF44, were also observed in the Western Pygmies (table 1). These are the only Western Pygmy mtDNAs belonging to subhaplogroup L2, and they are closely related to each other, both being defined by a *HaeIII* site at np 1899 and a *HaeIII* site loss at np 5261. With the exception of the combined +16389 *HinfI*/−16390 *AvaII* site, which defines subhaplogroup L2, the Western Pygmies do not share any of the other polymorphic sites found in L2 with either Senegalese or Eastern Pygmy mtDNAs. The only Western Pygmy haplotype showing a close relationship with a haplotype from a different population is the above-mentioned AF61, which is very similar to the Eastern Pygmy haplotype AF60. In addition to the 9-bp deletion, AF60 and AF61 share an *AluI* site loss at np 4583, a *HincII* site at np 5711, and an *AluI* site at np 8459.

Among the Eastern Pygmies, ~54% of the mtDNAs (AF38–AF42) cluster within subhaplogroup L2 and are defined by a *DdeI* site loss at np 13065 and a *RsaI* site at np 11776. A more ancient *HaeIII* site at np 13803 predates the separation of this group of haplotypes from


L
L1
L2
ASIAN OUTGROUP
-10394c
-16517e
-8616j
+10084l
+11001n
+2349j
-10397a
+10806g
-2758k
-7055a
+9070l,+12810k
+10319a
+173l
+3592 h
+13803e
-322e, -679c, -13957e
+8865j
+15247k
-12406o
Maure
ASIAN OUTGROUP
Western Pygmies (C.A.R.)
Eastern Pygmies (Zaire)

those observed in other African populations (fig. 2). Consequently, it appears that this group of mtDNAs is specific to the Eastern Pygmies. Two additional haplotypes (AF25 and AF37) were observed in the Eastern Pygmies. Haplotype AF25 was observed in two individuals and is located in a peripheral position within L1. Haplotype AF37, found in four individuals, is a member of L2 that shares a *Hae*III site loss at np 9438 (Brown et al. 1994; Newman et al. 1994) with one Mandenkalu mtDNA (AF36).

Senegalese mtDNAs

Of the Senegalese mtDNAs, ~31% and 43% are found in the two African-specific subhaplogroups L1 and L2, respectively. Most of the remaining Senegalese mtDNAs form four non-L haplogroups, each defined by specific mtDNA mutations (fig. 2). The only exceptions are two Maure haplotypes, AF7 and AF24, each found in one individual. AF24 is a haplotype that in all phylogenies stands by itself and shows very little similarities to other African mtDNAs (figs. 2 and 3). It is characterized by a *Dde*I at np 10394 and by an *Alu*I site at np 10397. These two mutations are present in >60% of Asian mtDNAs (Torroni et. al. 1994*c*) but have not been reported previously in Europeans (Torroni et al. 1994*b*) or Afro-Americans (Cann et al. 1987). The Maure in Senegal come from Mauritania, which had some degree of interaction with Mediterranean populations. Therefore, it is possible that the presence of AF24 in the Maure is due to limited gene flow from Asians, which moved along the Mediterranean Sea. The aberrant haplotype, AF7, tends to occupy the most internal nodal position in the rooted phylogenies (figs. 2 and 3). This makes it central to the radiation of the mtDNAs belonging to the four small non-L haplogroups in the midpoint rooting phylogeny (fig. 3). Haplotype AF7 is one of the two haplotypes previously described in non-African populations: AF7 and AF2 are identical to haplotypes CA94 and CA36 observed in 1.1% and 3.4% of Europeans, respectively (Torroni et al. 1994*b*).

In contrast to haplogroup L, the four small haplogroups that encompass most of the Senegalese mtDNAs lacking the 3592 *Hpa*I site show limited differentiation from both the European reference sequence (Anderson et al. 1981) and certain haplotypes observed in Europeans (Torroni et al. 1994*b*) and Asians (Ballinger et al. 1992; Torroni et al. 1993*b*). This finding suggests that this set of Senegalese mtDNAs are relatively close to those observed in non-Africans.

In contrast to the two Pygmy populations that did not share haplotypes, six haplotypes were shared between the various Senegalese populations. One Serere and one Wolof shared AF4; two Mandenkalu and one Maure shared AF14; eight Mandenkalu, one Wolof, and one Tukolor shared AF71; two Mandenkalu and two Wolof shared AF32; one Mandenkalu and one Wolof shared AF35; and one Peul and two Wolof shared AF46. In addition, the haplotypes found in the various Senegalese groups did not form population-specific groups when within or outside the L group. For example, the haplogroup (AF12–AF18), which lacks the np 3592 and is defined by a *Taq*I site at np 10084 (fig. 2), is shared by Mandenkalu, Wolof, Pular, and Maure. Similarly, the mtDNA group within L1 defined by a *Taq*I site at np 173 (AF71–AF79) is shared between Mandenkalu, Wolof, Pular, and Sosse. These findings can be explained by gene flow that has occurred between the various Senegalese groups and/or by a common origin of these Senegalese populations.

Parallel Mutations in African mtDNAs Identified by Parsimony Analysis

Parsimony analysis revealed that, of the 119 polymorphic restriction sites (character states) observed in the 79 African haplotypes, 18 restriction sites have been the target of more than one mutation (fig. 2; table 2). In addition, two restriction sites (7055 *Alu*I and 13065 *Dde*I) have mutated and subsequently reverted along individual mtDNA lineages. Most of the parallel mutational events have occurred only twice. However, triple occurrences were observed for a *Rsa*I site at np 10407 in the $tRNA^{Arg}$ gene, a *Hin*cII site loss at np 12406 in the ND5 gene, and a *Rsa*I site at np 16049 in the D-loop region; and eight independent events were found for the *Hae*III site at np 16517 in

Figure 2 MP tree that includes the 79 haplotypes (AF1–AF79) observed in 140 Africans. The tree was rooted by using the Asian mtDNA AS24 (Ballinger et al. 1992; Torroni et al. 1994*c*), is 156 steps in length, has consistency and retention indices (CI and RI) of 0.643 and 0.900, respectively, and is one of 1,000 MP trees generated with the TBR branch-swapping algorithm. The capital letter "L" in a shaded box indicates the African-specific haplogroup defined by the 3592 *Hpa*I site. L1 and L2 indicate subhaplogroups within haplogroup L. The circles (○) represent Western Pygmies (C.A.R.), while the squares (□) represent Eastern Pygmies (Zaire). The numbers at the nodes or at the end of each branch indicate different mtDNA haplotypes. The numbers associated with the lower-case letters indicate restriction sites or length polymorphisms defining specific haplotype groupings, and, when underlined, they indicate sites that have occurred multiple times or have undergone reversions (table 2). The lower-case letters correspond to the following restriction enzymes: a = *Alu*I; b = *Ava*II; c = *Dde*I; e = *Hae*III; f = *Hha*I; g = *Hin*fI; h = *Hpa*I; i = *Hpa*II; j = *Mbo*I; k = *Rsa*I; l = *Taq*I; m = *Bam*HI; n = *Hae*II; and o = *Hin*cII. Asterisks (*) indicate the six haplotypes shared among Senegalese populations. The length of the horizontal branches is proportional to the number of mutational events that separate the haplotypes. The inset illustrates the strict-consensus tree of the 1,000 MP trees. Its length is 188 steps, with CI and RI of 0.466 and 0.793, respectively. The strict-consensus tree retains most of the haplotype relationships observed in the MP trees.

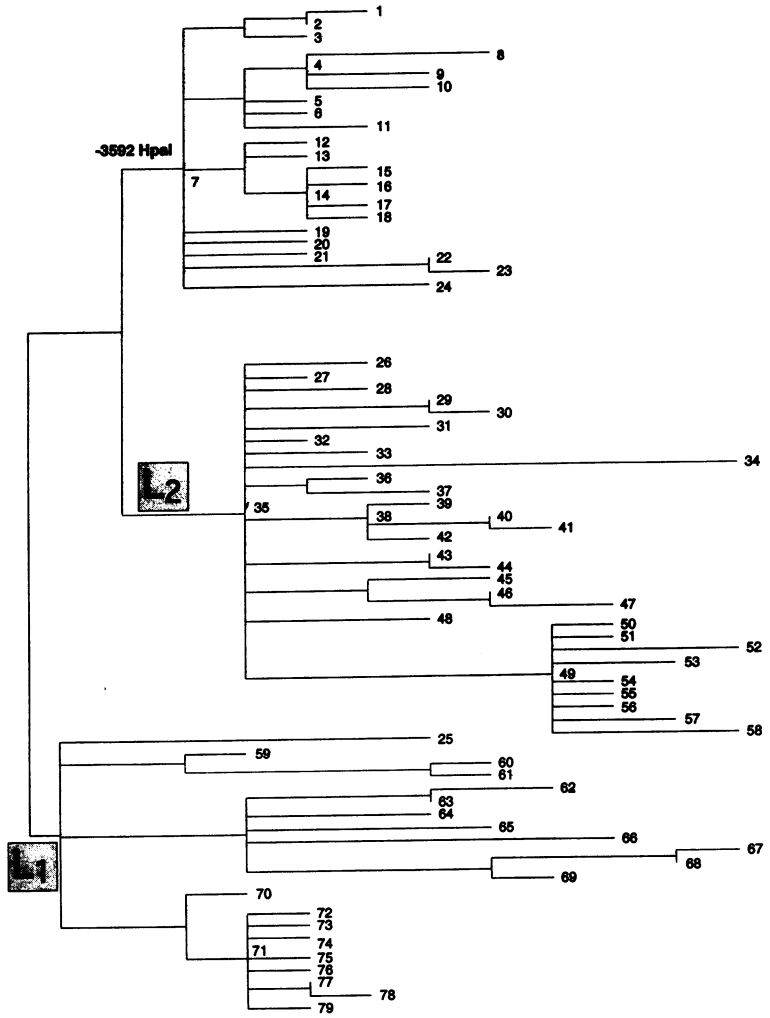


Figure 3 Strict-consensus tree of 3,000 MP trees obtained by midrooting point. This tree is 171 steps in length and has CI and RI of 0.515 and 0.832, respectively.

the D-loop region (fig. 2). This later finding is consistent with previously published reports from non-African populations (Torroni et al. 1993*a*) that the np 16517 *Hae*III site is extremely prone to multiple forward and reverse mutations. Some of the parallel site losses were sequenced to determine whether they were due to identical nucleotide changes. These included the np 4990 *Alu*I site, the np 7853 *Hin*cII site, the np 12406 *Hin*cII site, and the np 16049 *Rsa*I site. This analysis revealed that only the two occurrences of the *Hin*cII site loss at np 7853 were caused by different nucleotide changes (table 2).

The relatively high frequency of parallel mutations and reversions confirms that detailed haplotype analyses are necessary to determine accurately the relationships between mtDNAs. Caution should be used in extrapolating the results obtained by the screening of few specific mtDNA mutations to general issues concerning the origin of human populations (Torroni and Wallace 1995).

Discussion

The African-Specific Haplogroup L

Analysis of African mtDNA variation has revealed that 100% of Eastern and Western Pygmy mtDNAs and ~67% of Senegalese mtDNAs are defined by a *Hpa*I site gain at np 3592. In addition, all mtDNAs associated with this ancient polymorphism are monophyletic and cluster in haplogroup L. Haplogroup L is subdivided by additional polymorphic sites into the two subhaplogroups, L1 and L2. Subhaplogroup L1 is defined by a *Hin*fI site gain at np 10806, while subhaplogroup L2 is defined by a *Hin*fI site gain at np 16389. The ancient origin of haplogroup L and its two major subgroupings is suggested by their prevalence in Senegalese and Pygmy populations as well as in all other sub-Saharan populations (Johnson et al. 1983; Scozzari et al. 1988, 1994; Soodyall and Jenkins 1992, 1993).

Using the intra-group mtDNA sequence divergence, we estimated the ages of all African mtDNAs, of the

Table 2

Parallel Site Gains and Site Losses Identified by Parsimony Analysis

Restriction Site[a]	No. of Mutational Events[b]	Nucleotide Change	Region/Gene	Amino Acid Substitution[c]
102 *Hae*III (−→+)	2	102 A to G	D-loop	...
1899 *Hae*III (−→+)	2	1900 A to G	16S rRNA	...
2349 *Mbo*I (−→+)	2	2352 T to C	16S rRNA	...
3842 *Hae*III (−→+)	2	3843 A to G	ND 1	Syn
4990 *Alu* (+→−)	2	4991 G to A	ND 2	Syn
5984 *Ava*II (−→+)	2	5984 A to G	COI	Syn
7055 *Alu*I (+→−→+)	2	7055 A to G to A	COI	Syn
7853 *Hin*cII (+→−)	1	7853 G to A	COII	Val→Ile
	1	7858 C to T	COII	Syn
8519 *Hin*fI (−→+)	2	8522 C to T	ATPase 8	Pro→Ser
9820 *Hin*fI (−→+)	2	9824 T to C	COIII	Syn
10394 *Dde*I (+→−)	2	nd[d]	ND 3	nd[d]
10407 *Rsa*I (−→+)	3	10410 T to C	$tRNA^{Arg}$	...
11001 *Hae*II (−→+)	2	11002 A to G	ND 4	Syn
12406 *Hin*cII (+→−)	3	12406 G to A	ND 5	Val→Ile
13065 *Dde*I (−→+→−)	2	13068 A to G to A	ND 5	Syn
15882 *Ava*II (−→+)	2	15884 G to A	Cyt b	Ala→Thr
16049 *Rsa*I (+→−)	2	16051 A to G	D-loop	...
16215 *Mbo*I (−→+)	2	16215 A to G	D-loop	...
16389 *Hin*fI (−→+)	2	16390 G to A	D-loop	...
16517 *Hae*III (+→−)	8	nd[d]	D-loop	...

[a] A plus (+) indicates presence of the site, and a minus (−) indicates absence of the site.
[b] The location of the mutational events is shown in figure 2.
[c] Syn = synonomous substitution.
[d] nd = not determined.

common African mtDNA lineage, haplogroup L, and of the sub-haplogroups L1 and L2, on the assumption that nucleotide substitutions accumulate at a constant rate. The sequence divergence of all African mtDNAs was 0.292%. Using the mtDNA evolution rate of 2.2%–2.9%/Myr, we obtained a maximum age for African mtDNAs of 101,000–133,000 years before present (YBP). Similarly, haplogroup L showed an overall sequence divergence of 0.285%, while the subhaplogroups L1 and L2 had divergence values of 0.249% and 0.172%, respectively (table 3). This gives ages of 98,000–130,000 YBP for haplogroup L, 86,000–113,000 YBP for subhaplogroup L1, and 59,000–78,000 for subhaplogroup L2 (table 3).

Comparison of the sequence divergence of haplogroup L with the those estimated for the most divergent Asian-, European- and Native American-specific haplogroups (Torroni et al. 1994*b*, 1994*c*, 1994*d*) revealed that the African haplogroup is the most divergent among all continent-specific haplogroups (table 4). Its 0.285% sequence divergence is much higher than the 0.161% estimated for the Asian-specific

Table 3

Sequence Divergence and Divergence Time of African mtDNA Haplogroups

Haplogroup	No. of Haplotypes	No. of Subjects	Sequence Divergence (%)	Divergence[a] Time (years)
L	55	107	.285	98,000–130,000
L1	22	48	.249	86,000–113,000
L2	33	59	.172	59,000–78,000
African	79	140	.292	101,000–133,000

[a] Estimated using an mtDNA evolution rate of 2.2%–2.9%/Myr. Each divergence time estimate has been rounded to the nearest 1,000 years.

haplogroup M (Torroni et al. 1994*c*); the 0.039%–0.090% estimated for the European-specific haplogroups H, I, J, and K (Torroni et al. 1994*b*); and the 0.034%–0.096% estimated for the Native American haplogroups A, B, C, and D (Torroni et al. 1994*d*). Using the mtDNA evolution rate of 2.2%–2.9%/Myr, we estimated the age of the Asian haplogroup M to be 56,000–73,000 YBP, the age of the most divergent European haplogroup H to be 31,000–41,000 YBP, and the age of the most divergent Native American haplogroup C to be 33,000–44,000 YBP.

The ages estimated for each of the continent-specific haplogroups are congruent with the hypothesis that all modern human populations have a common and recent origin from an ancestral *Homo sapiens sapiens* population from Africa (Johnson et al. 1983; Cann et al. 1987) and agree with the estimated times of dispersal and colonization of each continent by modern-looking human populations (Cavalli-Sforza et al. 1993). In addition, if our estimate of the age of haplogroup L is correct, the origin of haplogroup L mtDNAs could have predated the expansion from Africa of *Homo sapiens sapiens,* which appears to have occurred ~100,000 YBP (McDermott et al. 1993). While haplogroup L is the most prevalent and divergent in modern African populations, it does not appear to have been carried from Africa by the migrating *Homo sapiens sapiens* populations that expanded into the Middle East and Asia and ultimately gave raise to modern Asian and Europeans populations. The only instances in which haplogroup L mtDNAs have been found outside Africa are in southern Italians and Arabs, which is probably the product of the more recent limited gene flow that is known to have taken place between sub-Saharan Africa and Europe.

Other mtDNA Haplogroups Observed in Africans

About one third of Senegalese mtDNAs lack the np 3592 *Hpa*I site and tend to form four small haplogroups in parsimony analyses. These haplotypes are relatively similar to some haplotypes observed in Europeans and Asians, and two of them, AF2 and AF7, have been observed in Europeans. Because Senegalese populations had extensive cultural and economic interactions with Saharan and North African populations, the presence of these haplotypes could be partially or completely attributed to genetic exchange with European populations. However, although we did not observe haplotypes lacking the 3592 *Hpa*I site in the Pygmies, haplotypes lacking the 3592 *Hpa*I site are not limited to Senegalese populations. These haplotypes have been described in 36% of the Bamileke from Cameroon (Scozzari et al.

Table 4

Sequence Divergence and Divergence Time of Ethnic-Specific mtDNA Haplogroups

Ethnic-Specific Haplogroup	Primary Defining Polymorphism	No. of Haplotypes (%)	No. of Subjects (%)	Sequence Divergence (%)	Divergence Time[a] (years)
African:					
L	+3592 *Hpa*I	55 (69.6)	107 (76.4)	.285	98,310–129,591
Asian:[b]					
M	+10397 *Alu*I	85 (48.9)	193 (53.6)	.161	55,517–73,181
Caucasian:[c]					
H	−7025 *Alu*I	45 (38.5)	70 (40.0)	.090	31,034–40,909
I	−1715 *Dde*I −4529 *Hae*II +10028 *Alu*I +8249 *Ava*II +16389 *Bam*HI	9 (7.7)	13 (7.4)	.074	25,517–33,636
J	−13704 *Bst*NI −16065 *Hin*fI	9 (7.7)	16 (9.1)	.042	14,483–19,091
K	−9052 *Hae*II	6 (5.1)	13 (7.4)	.039	13,448–17,727
Native American:[d]					
A	+663 *Hae*III	24 (28.6)	131 (39.1)	.075	25,862–34,091
B	9-bp-Deletion	19 (22.6)	83 (24.8)	.034	11,724–15,456
C	+13262 *Alu*I	25 (29.8)	61 (18.2)	.096	33,103–43,636
D	−5176 *Alu*I	16 (19.0)	60 (17.9)	.053	22,414–29,545

[a] Calculated by using the mtDNA evolution rate of 2.2%–2.9%/Myr (Torroni et al. 1994*d*).
[b] From Torroni et al. (1994*c*).
[c] From Torroni et al. (1994*b*).
[d] From Torroni et al. (1994*b*).

1994), 12% of the Khoisan populations from Namibia (Soodyall and Jenkins 1992), and 23%–89% of several Bantu-speaking populations from southern Africa (Johnson et al. 1983; Soodyall and Jenkins 1993). The finding of mtDNAs without the 3592 *Hpa*I site in sub-Saharan populations, which are unlikely to be genetically admixed with European populations, suggests that at least some of the mtDNAs lacking the 3592 *Hpa*I site in the Senegalese arose in Africa and are not the product of genetic admixture with populations from northern Africa, Europe, or Asia. Because of their widespread distribution in sub-Saharan populations, it is most likely that these mtDNAs have an ancient African origin. An African origin of the mtDNAs without the 3592 *Hpa*I site, their similarity to European and Asian mtDNAs, and the absence of mtDNAs defined by the HpaI site at np 3592 in non-African populations, appear to suggest that African mtDNAs without the 3592 *Hpa*I were the only mtDNAs that were carried from Africa by the *Homo sapiens sapiens* migrations, which ultimately gave rise to modern non-African populations.

Origin of Pygmy Populations

There are two contrasting hypotheses about the origin of the various Pygmy populations. Hiernaux (1977) has proposed, on the basis of anthropometric characters and few blood markers, that the different (Eastern and Western) Pygmy populations in Africa represent independent adaptation to the tropical forest environment. Cavalli-Sforza (1986*a*), by contrast, on the basis of many more markers, some of which are shared by different Pygmy groups and are not found in other populations, concluded that all Pygmy populations derive from a common ancestral Pygmy population and that the differences in gene frequencies observed between the Pygmy groups are best explained by different degree of genetic admixture with non-Pygmy populations and genetic drift. According to Cavalli-Sforza (1986*a*), the Eastern Pygmies, from Zaire, are probably the least admixed Pygmy population and the best representative of the ancestral Pygmy gene pool, while the Western Pygmies, from C.A.R., probably are hybrids of Eastern Pygmies-like ancestors with African farmers.

Historical evidence indicates some degree of genetic admixture between Pygmies and with non-Pygmies. However, in modern times, social constraints have meant that mixed marriages primarily occur between Pygmy women and non-Pygmy men, with the resulting children generally being integrated into the father's village. This process would result in the flow from Pygmy mtDNA into the adjacent farmer populations. Since marriage between a Pygmy man and a non-Pygmy woman is considered socially unacceptable (Cavalli-Sforza 1986*a*), gene flow from farmers to the Pygmies can occur only when children of mixed ancestry are integrated into the Pygmy group from which the mother originated. While this gene flow would affect nuclear gene frequencies in Pygmy populations, it would not affect the variation of the matrilineally transmitted mtDNA. Consequently, mtDNA analysis provides an independent and valuable approach to test the current hypotheses about the origin of Pygmy populations.

The present study shows that, although all Pygmy mtDNAs are members of the African haplogroup L, none of the mtDNA haplotypes is shared between Eastern and Western Pygmies. In addition, most of the Western Pygmy mtDNAs clustered in subhaplogroup L1, while most of the Eastern Pygmy mtDNAs clustered in sub-haplogroup L2. The only exception to the lack of overlap between Eastern and Western Pygmy mtDNAs is represented by the Eastern Pygmy haplotype AF60 and the Western Pygmy haplotype AF61. These two haplotypes share numerous mutations, including the 9-bp deletion between COII and $tRNA^{Lys}$ (fig. 2), and therefore they could represent haplotypes shared by descent by the two populations. However, they could also be the result of recent gene flow between the two populations, or they could both be acquired from non-Pygmies through genetic admixture.

Because of the lack of mtDNA data from other Pygmy populations and their neighboring farmer populations, it is currently not possible to discriminate between these possibilities. However, mtDNAs defined by the presence of the 9-bp deletion have been described in Nigerian populations (Merriwether et al. 1993), raising the possibility that haplotypes with the 9-bp deletion could also be present in the Bantu-speaking populations living in the same regions inhabited by the Pygmies and with whom the Pygmies could have genetically admixed.

To determine the extent of differentiation of the two Pygmy populations, we calculated their genetic distances and compared them with the respective genetic distances between Pygmies and Wolof and Pygmies and Mandenkalu (table 5). This comparison revealed that the sequence divergence between Eastern and Western Pygmies (0.316%) is comparable to that observed between Western Pygmies and Wolof (0.317%) and between Western Pygmies and Mandenkalu (0.304%), but it is somewhat higher than that observed between Eastern Pygmies and Wolof (0.244%) and Eastern Pygmies and Mandenkalu (0.270%). The observation that the genetic distance between the two Pygmy populations is similar to or even higher than those observed between Pygmy and non-Pygmy populations appears to support the hypothesis that Eastern and Western Pygmies are the result of independent adaptations to the same humid forest environment (Hiernaux 1977) rather that the hypothesis of a common ancestral origin of the Pygmy populations (Cavalli-Sforza 1986*a*).

However, these findings should be interpreted with some caution. Analysis of Amerind populations has

Table 5

Percent Sequence Divergence within and between African Populations

	Eastern Pygmies (*N* = 22)	Western Pygmies (*N* = 17)	Wolof (*N* = 20)	Mandenkalu (*N* = 60)
Eastern Pygmies	.203	.316	.244	.270
Western Pygmies		.284	.317	.304
Wolof			.221	.254
Mandenkalu				.256

NOTE.—Intrapopulation and interpopulation sequence divergence values are on the diagonal and above the diagonal, respectively.

shown that mtDNA variation in populations of limited size is particularly prone to genetic drift effects and that differential loss of specific mtDNA haplogroups can greatly affect the genetic distance between related populations (Torroni et al. 1993*a*, 1994*a*). Since Pygmies have small population sizes (Cavalli Sforza et al. 1986*b*), it is possible that genetic drift is partially responsible for the tendency of most of the Western and Eastern Pygmy mtDNAs to cluster in subhaplogroups L1 and L2, respectively.

Comparison of the Intrapopulation mtDNA Variation in Africans and Other World Populations

One of the arguments used to support the hypothesis of a recent African origin of all human mtDNAs is that Africans show the highest degree of mtDNA differentiation (Johnson et al. 1983; Cann et al. 1987; Scozzari et al. 1988). However, extensive comparisons of mtDNA variation in human populations are not easily accomplished, because of the different techniques and levels of resolution used in the different studies to acquire mtDNA data. The technique and level of resolution that we used in the analysis of Pygmy and Senegalese mtDNAs are the same of those employed in the mtDNA analysis of a large number of Asian, European, and Native American populations (Ballinger et al. 1992; Torroni et al. 1992, 1994*a*). This allows a direct comparison of African and non-African intrapopulation sequence divergence values.

Table 6 shows the results of this comparison. The highest value of intra-group sequence divergence was 0.284% and was observed in the Western Pygmies. The sequence divergence values of the other African groups are generally next, ranging between 0.256% and 0.203%.

In general, the Asian populations show less sequence divergence than the Africans. The exception is the Vietnamese, who harbor an intrapopulation sequence divergence (0.236%) comparable to that observed in some of the African populations. The lowest values of intragroup sequence divergence (0.143%–0.007%) are found in Native American tribes. Overall, these data confirm that African populations harbor the highest degree of mtDNA diversity and that a strong correlation exists between the extent of mtDNA variation observed in modern human populations and the archeological ages associated to the dispersal of modern *Homo sapiens* throughout the world.

Table 6

mtDNA Intrapopulation Sequence Divergences in World Populations

Population	No.	Sequence Divergence (%)
Africa:		
Western Pygmies	17	.284
Mandenkalu	60	.256
Wolof	20	.221
Eastern Pygmies	22	.203
Asia:[a]		
Vietnamese	28	.236
Malaysian Chinese	14	.196
Koreans	13	.185
Malays	14	.182
Sabah Aborigines	30	.180
Malay Aborigines	32	.148
Taiwanese Han	20	.145
Americas:[b]		
Dogrib	30	.016
Navajo	48	.062
Maya	27	.110
Mixe	16	.124
Pima	30	.143
Mixtec (Alta)	15	.102
Mixtec (Baja)	14	.080
Zapotec	15	.127
Teribe	20	.070
Guatuso	20	.040
Boruca	14	.067
Kuna	16	.007
Guaymi	16	.058
Bribri/Cabecar	24	.079
Yanomama	24	.114
Wapishana	12	.083
Ticuna	28	.138

[a] From Ballinger et al. (1992).

[b] From Torroni et al. (1992 and 1994*a*).

In conclusion, this study reveals that African mtDNAs are ~101,000–133,000 years old and that African mtDNA variation is the greatest of all populations, supporting an African origin for *Homo sapiens*. Furthermore, most African mtDNAs are members of the African-specific haplogroup L. This haplogroup appears to be the most ancient of all continent-specific haplogroups, having arisen 100,000–130,000 YBP. Such an age would indicate that haplogroup L mtDNAs originated before modern *Homo sapiens* expanded from Africa. Yet, this haplogroup is virtually absent in non-Africans. Therefore, mtDNAs of haplogroup L must not have been carried from Africa by the migrations that ultimately gave rise to Asians and Europeans. On the contrary, the close similarities between the African haplotypes lacking the 3592 *Hpa*I site and haplotypes observed in Asians and Europeans suggest that African haplotypes without the np 3592 site were the ones from which modern Asian and European mtDNAs were derived. Hence, further geographic localization of haplogroup L and non-L mtDNAs among various African populations could provide further insights into the origins of the first great migrations of our species.

Acknowledgments

The authors are indebted to Dr. L. Luca Cavalli-Sforza, Department of Genetics, Stanford University, for providing the Pygmy samples for analysis. This work was supported by NIH grant GM 46915 (to D.C.W.).

Appendix

Table A1

Polymorphic Restriction Sites That Define the 79 Distinct mtDNA Haplotypes (AF1–AF79) Observed in 140 Africans

Sites	Haplotypes 1111111111222222222233333333334444444444555555555566666666667777777777 1234567890123456789012345678901234567890123456789012345678901234567890123456789
102e	0000000000000000000000000000000000000000000000000001000000000011100000000000
173l	0000000000000000000000000000000000000000000000000000000000000000111111111
255k	0000000000000000000000000000000000000000000000000000000000000100000000000000
322e	1111111111111111111111111111111111111111111111100000000001111111111111111111
679c	1111111111111111111111111111111111111111111111100000000001111111111111111111
762k	0000000000000000000000000000000000000000000011000000000000000000000000000000
1169b	1111111111111111111111111101111111111111111111111111111111111111111111111111
1715c	1111111011111111111111111111111111111111111111111111111111111111111111111111
1899e	0000000010000000000000000000000000000000011000000000000000000000000000000000
2246a/2250a	0000000000000000000000000000000000000000001000000000000000000000000000000000
2349j	0000000000000000011110000000000000000000000000000000000000000000001111111111
2758k	1111111111111111111111111111111111111111111111111111111100000000000000000000
2904a	1111111111111111111111111111111111111111111111111111111111111111111111101111
3348j	1000000000000000000000000000000000000000000000000000000000000000000000000000
3534c	1111111111111111111111101111111111111111111111111111111111111111111111111111
3534c/3537a	1111111111111111111111111111111101111111111111111111111111111111111111111111
3592h	0000000000000000000001111111111111111111111111111111111111111111111111111111
3698f	1111111111111111111111111111111101111111111111111111111111111111111111111111
3842e	0000000000000000000000000000000000000000000000000000000100001100000000000
4064a	0000000000000000000000000000000000000000000000001000000000000000000000000000
4157a	0000000000000000000000000000000000000000000111000000000000000000000000000000
4310a	1111111111111111111111111111111111111111111111111111111000111111111111111111
4583a	1111111111111111111111111111111111111111111111111111111001111111111111111111
4709c/4711i	0000000000000000000000000000000000000000000000000000000000000000000000010000000
4769a[a]	0000000000000000000000000000000000000000000000000000000000000000000000000000000
4831k	0000000000000000000000000000001000000000000000000000000000000000000000000000000
4990a	1111111110111111111111111111111111111111111111111111111111111111011111111111111
5260h/5261e	0000000000000000000110000000000000000000000000000000000000000000000000000000000
5261e	1111111111111111111111111111111111111111111100111111111111111111111111111111111
5552c	1111111111111111111111111111111111111111111001111111111111111111111111111111111
5584a	1111111111111111111001111111111111111111111111111111111111111111111111111111111
5711o	0000000000000000000000000000000000000000000000000000000110000000000000000
5983g/5984b	1111111111111111111111111111111111111111111111111111111111000111111111111111
6014l	1111111111111111111111111111111101111111111111111111111111111111111111111111
6296c	1111111111111111111111011111111111111111111111111111111111111111111111111111
6377c	1111111111111111111111111111111111111111111111111111111111111111111111111001

(*continued*)

Table A1 (continued)

Sites	Haplotypes
	1111111111222222222233333333334444444444555555555566666666667777777777
	1234567890123456789012345678901234567890123456789012345678901234567890123456789
6625c	00000000000000000000000000000000000000000000000000000000000000000010000000000000
6931g	11111111111111111011111111111111111111111111111111111111111111111111111111111111
6957e	11111111111111111111111111111111111101111111111111111111111111111111111111111111
7025a[a]	11111111111111111111111111111111111111111111111111111111111111111111111111111111
7055a	11111111111111111111111111111111111111111111111111111111111000100000000000000000
7335l	11111111111111111111111111111111111111111111111111111111011111111111111111111111
7531e	00000000000000000000000000000000000000000000000000000000000000100000000000000000
7658j	11111111111111111111111111111111111111111111111111111111111111111111101111111111
7758c	11111111111111111111111111001111111111111111111111111111111111111111111111111111
7828f	00000000000000000000000000000000000000000000000000000000000000000000000000000001
7853o	11111111111111111111111111111100111111111111111111111111111111111111111111111111
7859j	11111111111111111111111111111111111111111111111111111111111111111111111110111111
7877g	00000000000000000000000000000000000000000000010000000000000000000000000000000000
8249b/8250e	00000001000000000000000000000000000000000000000000000000000000000000000000000000
8391e	11111111111111111011111111111111111111111111111111111111111111111111111111111111
8459a	00000000000000000000000000000000000000000000000000000000011000000000000000000000
8519g	00000000010000000000000000000010000000000000000000000000000000000000000000000000
8536j	00000000000000000000000000000000000000000000000000000100000000000000000000000000
8566b	00000000000000000000000000000000000000000000000000000010000000000000000000000000
8569c/8572e	00000000000000000000000000000000000000000000000000000100000000000000000000000000
8616j	11100010000111111111111111111111111111111111111111111111111111111111111111111111
8858f[a]	11111111111111111111111111111111111111111111111111111111111111111111111111111111
8865j	00000000000000000000010000000000000000000000000000000000000000000000000000000000
8872a	00000000000000000000000000000000000000010000000000000000000000000000000000000000
8988e	00000000000000000000000000000000000000000000000000000000000000001000000000000000
9052e/9053f	11111111111111111111111111111111111111111111111111111111111111111011111111111111
9070l	00000000000000000000000000000000000000000000000000000000000011111110000000000000
9153c	00001000000000000000000000000000000000000000000000000000000000000000000000000000
9266e	11111111111111111111111111111111111111111111111111111111111111111011111111111111
9368k	00000000000000000000000000000000000000000000000000000000000011000000000000000000
9380f	11111111111111111111111111111111011111111111111111111111111111111111111111111111
9404k	00000000000000000000000000000001000000000000000000000000000000000000000000000000
9438e	11111111111111111111111111111111001111111111111111111111111111111111111111111111
9493g	00000000000000000000000000000000000000000000000000000000010000000000000000000000
9644a	11111111111111111111111111111111111111111111111111111111111111111000111111111111
9811g	00000000000000000000000000000000010000000000000000000000000000000000000000000000
9820g	00000000000000000000010000000010000000000000000000000000000000000000000000000000
9985j	00000000000000000000000000010000000000000000000000000000000000000000000000000000
10084l	00000000001111111000000000000000000000000000000000000000000000000000000000000000
10135a	00000000000000000000000000001000000000000000000000000000000000000000000000000000
10254j	11011111111111111111111111111111111111111111111111111111111111111111111111111111
10319a	00000000000000000000000000000000000000000000000000000000011001110000000000000000
10356c	11111111111101111111111111111111111111111111111111111111111111111111111111111111
10394c	00011111111111111111111111111111111111111111111111111111111111111001111111111111
10397a	00000000000000000000000010000000000000000000000000000000000000000000000000000000
10407k	00000000000000000000000000011001000000000000001000100000000000000000000000000000
10598a	11111111111111111111111111111111111111111111101111111111111111111111111111111111
10644k	00000000000000001000000000000000000000000000000000000000000000000000000000000000
10806g	00000000000000000000001000000000000000000000000000000011111111111111111111111111
10907c	00000000000000000000000000000000000001100000000000000000000000000000000000000000
11001n/11002f	00000000000001111100000000000000000000000000000000000000000000000010000000000000
11641e	00000000000000000000000000000000000000000000000000000011100000000000000000000000
11766k	00000000000000000000000000000000000111110000000000000000000000000000000000000000

(*continued*)

Table A1 (continued)

Sites	Haplotypes
	1111111111222222222233333333334444444444555555555566666666667777777777
	1234567890123456789012345678901234567890123456789012345678901234567890123456789
12300e	1111111111111111111111101111111111111111111111111111111111111111111111111111
12406o/12406h	1111111101111111111011011111111111111111111111111111111111111111111111111111
12560a	1111111111111111111111111111111111111111111101111111111111111111111111111111
12765c	1111111111111111111111111111111111111111111111111111111111111110001111111111
12810k	0000000000000000000000000000000000000000000000000000000000000011111110000000000
12946c/12949n	0000000000000000000000010000000000000000000000000000000000000000000000000000000
/12950f	
13031g	1111111111111111111111111111111111111111111111111111111111111111111111011111
13041a	0000000000000000000000000000000000000000000010000000000000000000000000000000
13065c	1111111111111111111111111111111111111000101111111111111111111111111111111111
13366m/13367b	0000010000000000000000000000000000000000000000000000000000000000000000000000
/13367j	
13432k	0000000000000000000000000000000000000000000000010000000000000000000000000000
13702e[a]	0000000000000000000000000000000000000000000000000000000000000000000000000000
13803e	0000000000000000000000011111111011111111000010000000000000000000000000000000
13957e	1111111111111111111111111111111111111111111111000000000011111111111111111111
14149o[a]	0000000000000000000000000000000000000000000000000000000000000000000000000000
14268g[a]	1111111111111111111111111111111111111111111111111111111111111111111111111111
14368g[a]	0000000000000000000000000000000000000000000000000000000000000000000000000000
14304a	1111111111111111111111111111111111111111111111111111111111111111111111110111
14869j	1111111111111111101111111111111111111111111111111111111111111111111111111111
14976g	1111111111111111111111111111111111111111111111111111111011111111111111111111
15221j	0000000000000000000000000000000000000000000000000000000000000000000000000010
15247k	0000000000000000000001000000000000000000000000000000000000000000000000000000
15357j	1111110111111111111111111111111111111111111111111111111111111111111111111111
15460c	0000000000000000000000000000000000000000000000001000000000000000000000000000
15754c	0000000000000000000000000000001000000000000000000000000000000000000000000000
15882b/15883e	0000000000000010000000100000000000000000000000000000000000000000000000000000
15925i	1111111111111111111111111111111111111111111111111111111111101111111111111111
16049k	1111111111111011111001111111111111111111111111111111111111011111111111111111
16109e	0000000000000000000000000000000000000000000000000000100000000000000000000000
16170j	0000000000000000000000000000000010000000000000000000000000000000000000000000
16215j	0000000001000000000000000000000000000000000000000000000000000100000000000000
16224l	0000000000000000000000000000000000000000000000000000000000000001000000000
16318e	0000000000000000000000000000000000000000000001000000000000000000000000000000
16373e	0000000100000000000000000000000000000000000000000000000000000000000000000000
16389g/16390b	0000000000000000000000011111111111111111111111111111101000000000000000000
16398e	0000000000000000000000000000000000000000000001000000000000000000000000000000
16517e	0010111000101111111111100000001111111111111000000000011111111110011111111111
9-bp del	0000000000000000000000000000000000000000000000000000000000000110000000000000000
12bp ins	0000000000000000000000000000000000000000000000000000000000000010000000000000

NOTE.—A "1" indicates the presence of a site, and a "0" indicates the absence of a site, except for the COI-tRNA^{Tyr} 12-bp insertion and the COII-tRNA^{Lys} 9-bp deletion, where a "1" indicates the presence of either the deletion or the insertion, and a "0" indicates the absence of the deletion and the insertion. Sites are numbered from the first nucleotide of the recognition sequence according to the published sequence (Anderson et al. 1981). The restriction enzymes used in the analysis are designated by the following single-letter code: a = *Alu*I; b = *Ava*II; c = *Dde*I; e = *Hae*III; f = *Hha*I; g = *Hin*fI; h = *Hpa*I; i = *Msp*I; j = *Mbo*I; k = *Rsa*I; l = *Taq*I; m = *Bam*HI; n = *Hae*II; and o = *Hin*cII. Sites separated by a diagonal line indicate either simultaneous site gains or site losses for two different enzymes or a site gain for one enzyme and a site loss for another (in this case, the classification "0" or "1" refers to the first site), because of a single inferred nucleotide substitution; these sites are considered to be only one restriction site polymorphism in the parsimony analysis and sequence-divergence calculations. The haplotype of the Malay mtDNA used as Asian outgroup (MA12; Ballinger et al. 1992) was the following: −951j, +1063e, −7461l, +9326e/+9329f, +10143a, +10394c, +10397a, 16517e. The haplotype of the European mtDNA used as European outgroup (CA85; Torroni et al. 1994*b*) was the following: −1715c, −4529n, +7025a, +8249b/−8250e, +10028a, +10394c, +15754c, +16389m/+16390j/−16390b, +16517e.

[a] Site found to be present or absent in all samples, contrary to the published sequence.

References

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

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics 130:139–152

Bonne-Tamir B, Johnson MJ, Natali A, Wallace DC, Cavalli-Sforza LL (1986) Mitochondrial DNA types in two Israeli populations: a comparative study at the DNA level. Am J Hum Genet 38:341–351

Brown MD, Torroni A, Huoponen K, Chen Y-S, Lott M, Wallace D.C. (1994) Pathological significance of the mtDNA COX III mutation at nucleotide pair 9438 in Leber hereditary optic neuropathy. Am J Hum Genet 55:410

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cann RL, Wilson (1983) Length mutation in human mitochondrial DNA. Genetics 104:699–711

Cavalli-Sforza LL (1986*a*) African Pygmies: an evaluation of the state of research. In: Cavalli-Sforza LL (ed) African Pygmies. Academic Press, London, pp 361–422

Cavalli-Sforza LL (1986*b*) Demographic data. In: Cavalli-Sforza LL (ed) African Pygmies. Academic Press, London, pp 23–44

Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. Science 259:639–646

De Benedictis G, Rose G, Passarino G, Quagliarello C (1989) Restriction fragment length polymorphism of human mitochondrial DNA in a sample population from Apulia (south Italy). Ann Hum Genet 53:311–318

Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli Sforza LL, Wallace DC (1981) Ethnic variation in *Hpa*I endonuclease cleavage patterns of human mitochondrial DNA. Proc Natl Acad Sci USA 78:5768–5772

Excoffier L, Langaney A (1989) Origin and differentiation of human mitochondrial DNA. Am J Hum Genet 44:73–85

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti AS, Langaney A, Excoffier L. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol (in press)

Hertzberg M, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510

Hiernaux J (1977) Long-term biological effects of human migration from the African savanna to the equatorial forest: a case study of human adaptation to a hot and wet climate. In: Harrison GA (ed) Population structure and human variation. Cambridge University Press, London and New York, pp 187–217

Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL (1983) Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. J Mol Evol 19:255–271

McDermott F, Grun R, Stringer CB, Hawkesworth CJ (1993) Mass-spectrometry U-series dates for Israeli Neanderthal/early modern hominid sites. Nature 363:252–255

Merriwether DA, Rothhammer F, Ferrel RE (1993) Mitochondrial D-loop sequence variation in Native South Americans. Am J Hum Genet Suppl 53:833

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. Genetics 105:207–217

Newman NJ, Torroni A, Brown MD, Lott MT, Marquez-Fernandez M, Wallace DC, the Cuba Neuropathy Field Investigation Team (1994) Epidemic neuropathy in Cuba not associated with mitochondrial DNA mutations found in Leber's hereditary optic neuropathy patients. Am J Ophthalmol 118:158–168

Passarino G, Semino O, Modiano G, Santachiara-Benerecetti AS (1993) COII/tRNA^{Lys} intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have oriental affinities. Am J Hum Genet 53:609–618

Ritte U, Neufeld E, Preger FM, Gross M, Hakim I, Khatib A, Bonne-Tamir B (1993) Mitochondrial DNA affinities of several Jewish communities. Hum Biol 65:359–385

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, 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

Scozzari R, Torroni A, Semino O, Cruciani F, Spedini G, Santachiara Benerecetti AS (1994) Genetic studies in Cameroon: mitochondrial DNA polymorphisms in Bamileke. Hum Biol 66:1–12

Scozzari R, Torroni A, Semino O, Sirugo G, Brega A, Santachiara Benerecetti AS (1988) Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. Am J Hum Genet 43:534–544

Semino O, Torroni A, Scozzari R, Brega A, De Benedictis G, Santachiara-Benerecetti AS (1989) Mitochondrial DNA polymorphisms in Italy. III. Population data from Sicily: a possible quantitation of maternal African ancestry. Ann Hum Genet 53:193–202

Shields GF, Hecker K, Voevoda MI, Reed JK (1992) Absence of the Asian-specific region V mitochondrial marker in native Beringians. Am J Hum Genet 50:758–765

Soodyall H, Jenkins T (1992) Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa. Ann Hum Genet 56:315–324

——— (1993) Mitochondrial DNA polymorphisms in Negroid populations from Namibia: new light on the origins of the Dama, Herero, and Ambo. Ann Hum Biol 20:477–485

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Stoneking M, Wilson AC (1989) Mitochondrial DNA. In: Hill A, Serjeantson S (eds) The colonization of the Pacific: a genetic trail. Oxford University Press, Oxford, pp. 215–245

Swofford D (1993) Phylogenetic Analysis Using Parsimony (PAUP), Version 3.1. Illinois Natural History Survey, Champaign, Illinois

Templeton AR (1992) Human origins and analysis of mitochondrial DNA sequences. Science 255:737

Torroni A, Brown MD, Lott MT, Newman NJ, Wallace DC, the Cuba Neuropathy Field Investigation Team. African, Native American and European mtDNAs in Cubans from the Pinar del Rio Province and implications for the recent epidemic neuropathy in Cuba. Hum Mutat (in press)

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

Torroni A, Lott MT, Cabell MF, Chen Y-S, Lavergne L, Wallace DC (1994*b*) 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 55:760–776

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994*c*) 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, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994*d*) 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, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al (1993*a*) Asian affinities and continental radiation of the four founding Native American mitochondrial DNAs. Am J Hum Genet 53:563–590

Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, 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, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993*b*) Mitochondrial DNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53:591–608

Torroni A, Wallace DC (1994) Mitochondrial DNA variation in human populations and implications for detection of mtDNA mutations of pathological significance. J Bioenerg Biomembr 26:261–271

——— (1995) Mitochondrial DNA haplogroups in Native Americans. Am J Hum Genet 56:1234–1236

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Wallace DC, Lott MT, Torroni A, Brown MD, Shoffner JM (1994) Report of the committee on human mitochondrial DNA. In: Cuticchia AJ, Pearson PL (eds) Human gene mapping 1993. Johns Hopkins University Press, Baltimore, pp 813–845

Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Res 15:529–542