Am. J. Hum. Genet. 62:1137–1152, 1998

mtDNA Analysis Reveals a Major Late Paleolithic Population Expansion from Southwestern to Northeastern Europe

Antonio Torroni,[1] Hans-Jürgen Bandelt,[3] Leila D'Urbano,[1] Päivi Lahermo,[4] Pedro Moral,[5] Daniele Sellitto,[2] Chiara Rengo,[1] Peter Forster,[3] Marja-Liisa Savontaus,[4] Batsheva Bonné-Tamir,[6] and Rosaria Scozzari[1]

[1]Dipartimento di Genetica e Biologia Molecolare and [2]Centro di Genetica Evoluzionistica, Universitá di Roma "La Sapienza," Rome; [3]Mathematisches Seminar, Universität Hamburg, Hamburg; [4]Department of Medical Genetics, University of Turku, Turku, Finland; [5]Departament de Biologia Animal, Universitat de Barcelona, Barcelona; and [6]Department of Human Genetics, Sackler Faculty of Medicine, Ramat-Aviv, Israel

Summary

mtDNA sequence variation was studied in 419 individuals from nine Eurasian populations, by high-resolution RFLP analysis, and it was followed by sequencing of the control region of a subset of these mtDNAs and a detailed survey of previously published data from numerous other European populations. This analysis revealed that a major Paleolithic population expansion from the "Atlantic zone" (southwestern Europe) occurred 10,000–15,000 years ago, after the Last Glacial Maximum. As an mtDNA marker for this expansion we identified haplogroup V, an autochthonous European haplogroup, which most likely originated in the northern Iberian peninsula or southwestern France at about the time of the Younger Dryas. Its sister haplogroup, H, which is distributed throughout the entire range of Caucasoid populations and which originated in the Near East ~25,000–30,000 years ago, also took part in this expansion, thus rendering it by far the most frequent (40%–60%) haplogroup in western Europe. Subsequent migrations after the Younger Dryas eventually carried those "Atlantic" mtDNAs into central and northern Europe. This scenario, already implied by archaeological records, is given overwhelming support from both the distribution of the autochthonous European Y chromosome type 15, as detected by the probes 49a/f, and the synthetic maps of nuclear data.

Received September 15, 1997; accepted for publication March 5, 1998; electronically published April 17, 1998.

Address for correspondence and reprints: Dr. Antonio Torroni, Dipartimento di Genetica e Biologia Molecolare, Universitá di Roma "La Sapienza," P.le Aldo Moro 5, 00185 Rome, Italy. E-mail: torroni@axcasp.caspur.it

0002-9297/98/6205-0021$02.00

Introduction

Molecular analyses have shown that most human mtDNA sequence variation has accumulated sequentially along radiating maternal lineages from sets of mtDNA founders, during and after the process of human colonization of different geographical regions of the world. This is the reason why haplogroups defined by high-resolution RFLP analysis are often found to be geographically or ethnically specific (Wallace 1995).

The first study directed at the identification of European mtDNA haplogroups was carried out by high-resolution RFLP analysis of individuals of European ancestry living in North America, and it revealed four European-specific haplogroups (H, I, J, and K; Torroni et al. 1994*a*). More recently, studies based on sequence data from the first hypervariable segment (HVS-I) of mtDNA (Richards et al. 1996) and RFLP haplotype data from the entire mtDNA in conjunction with sequence data from both HVS-I and HVS-II (Torroni et al. 1996, 1997) not only have confirmed the wide distribution of haplogroups H, I, J, and K among European populations but also have revealed the presence of some additional haplogroups. Torroni et al. (1996) named these additional haplogroups T, U, V, W, and X. Similar to the four haplogroups identified initially, it was observed that haplogroups T, V, and W are Caucasoid-specific, whereas haplogroup U is shared between Europeans and Africans (Torroni et al. 1996) and haplogroup X is shared between Europeans and northern Amerinds (Forster et al. 1996; Scozzari et al. 1997). These nine haplogroups, together with a few representatives of the Asian superhaplogroup M (Torroni et al. 1994*b*) and the African haplogroups L1 and L2 (Chen et al. 1995; Watson et al. 1997), were found to encompass virtually all mtDNAs in Europe (Torroni et al. 1996, 1997). Analyses of the frequency, variation, and distribution of these haplogroups have been used to evaluate current models concerning the process of colonization of the European continent (Richards et al. 1996) and have suggested that

mtDNA haplogroups can play an important role in modulating disease expression (Brown et al. 1997; Torroni et al. 1997).

The study by Torroni et al. (1996) showed the correlation between RFLPs in the mtDNA coding region and nucleotide changes in the mtDNA–control region sequences, for all European haplogroups except haplogroup V, which was absent in the population (Tuscan) in which the correlation was established.

In this study, we establish the correlation between the marker mutations of the mtDNA coding and control regions for haplogroup V and analyze the distribution and frequency of the sister haplogroups V and H in a wide range of Caucasoid populations. This analysis indicates that a Paleolithic population of southwestern Europe has largely contributed to the mitochondrial gene pool of all central and northern European populations, including the Saami and the Finns.

Subjects and Methods

A total of 419 unrelated subjects, belonging to eight Caucasoid populations (50 Adygei of the Caucasus region, 38 Yemenite Jews, 45 Druzes from northern Israel, 51 central Italians, 48 northern Sardinians from the province of Sassari, 50 Basques from the Guipuzcoa province, 49 Finns, and 37 Swedes) and one population from central Siberia (51 Yakuts), were analyzed for mtDNA haplotype variation by high-resolution restriction analysis. Control-region sequences for both mtDNA hypervariable segments (HVS-I and HVS-II) were then determined for 23 mtDNAs belonging to haplogroup V. These included 5 of the Basques and 3 of the Sardinians, also analyzed by high-resolution restriction analysis; three unrelated Finns affected by Leber hereditary optic neuropathy (LHON) (Lamminen et al. 1997); and 12 Saami (M.-L. Savontaus, unpublished data). Genomic DNAs from these samples were extracted either from buffy coats or lymphoblast cell lines by use of standard procedures.

To determine RFLP haplotypes, the entire mtDNA of each sample was amplified in nine overlapping fragments, by use of PCR and the primer pairs described in the study by Torroni et al. (1997). Each of the nine PCR segments was then 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) and was tested for the A→G sequence polymorphism at nucleotide position (np) 12308 in the $tRNA^{Leu}$ gene by use of a mismatched primer that generates a *Hin*fI site when the 12308G mutation is present (Torroni et al. 1996). In addition, all subjects were screened for the presence/absence of the *Bst*NI site at np 13704, the *Acc*I sites at nps 14465 and 15254, the *Bfa*I site at np 4917, and the *Nla*III sites at nps 4216 and 4577. Restriction fragments were resolved as described elsewhere (Torroni et al. 1997).

Sequences of HVS-I and HVS-II were determined from nps 16024–16383 and from nps 56–320, respectively, by use of the Sanger dideoxy chain–termination method and sequenase enzyme. The expansion/coalescence times of the haplogroups were estimated by calibrating the mean distance (ρ) from both the inferred founder(s)/root sequence and the founder(s)/root haplotype. For HVS-I sequences, one transition within nps 16090–16365 from an ancestral sequence was estimated to correspond to 20,180 years (Forster et al. 1996; Watson et al. 1997).

To relate HVS-I variation to the RFLP variation observed with the 14–restriction-enzyme system, Native American and Siberian mtDNAs (Torroni et al. 1993*a*, 1993*b*) characterized both by HVS-I sequences and by RFLP haplotype data were compared by use of the statistic ρ (Forster et al. 1996), for which the identification of founding sequences/haplotypes is required. These data included 15 mtDNAs from haplogroup A, 6 from haplogroup B, 13 from haplogroup C, and 11 from haplogroup D. The roots of RFLP haplogroups A, B, C, and D are founder haplotypes for this pooled data set (Torroni et al. 1993*a*, 1993*b*). With respect to HVS-I, haplogroups A, C, and D each have two founding sequences (Forster et al. 1996), and haplogroup B also may be subdivided, into ancestral B1, for which putative descendants are found in the data from the studies by Torroni et al. (1993*a*) and Ward et al. (1996), and derived B2. The variant nucleotides relative to the Cambridge Reference Sequence are the following: 16111C/T–16223T–16290T–16319A–16362C, for group A1/A2; 16189C–16217T/C, for group B1/B2; 16223T–16298C–16325T/C–16327T, for group C1/C2; and 16223T–16325T/C–16362C, for group D1/D2. When ρ is applied to HVS-I, the number of transitions, within nps 16090–16365, is counted from each sampled sequence to the corresponding founder sequence (ignoring transversions and indels); then, groupwise averages are taken. A similar procedure is adapted for the restriction-site data, for which longer indels (such as the 9-bp deletion between the COII and $tRNA^{Lys}$ genes), as well as the 16517*Hae*III variation, are disregarded; moreover, the double-site loss of 10394*Dde*I/10397*Alu*I is counted as one event. By this approach, we observed the following $\rho_{HVS\text{-}I}/\rho_{RFLP}$ ratios in the Native American/Siberian data set: 1.27 (1.27/1.00) for haplogroup A; 1.33 (0.67/0.50) for haplogroup B; 0.82 (1.08/1.31) for haplogroup C; and 1.40 (1.27/0.91) for haplogroup D. Averaging the ratios over the four groups yields 1.21, suggesting that 1.21 transitions can be equated with one restriction-site change. The same value is obtained by calculation of the ratio of the number ($S_{HVS\text{-}I}$) of segregating tran-

sitions in HVS-I and the number (S_{RFLP}) of segregating restriction sites, since the respective founders. We then counted 46 mutations in HVS-I and 38 restriction-site changes—that is, $\rho_{HVS\text{-}I}/\rho_{RFLP} \approx 1.21 \approx S_{HVS\text{-}1}/S_{RFLP}$. We then took this value as a current best estimate for translating RFLP variation into HVS-I variation. Exactly the same ratio is obtained when RFLP variation is compared with HVS-I variation in haplogroup V (see Results). By using $\rho_{HVS\text{-}I} \approx 1.21\rho_{RFLP}$ and the calibration of $\rho_{HVS\text{-}I} = 1$ corresponding to 20,180 years, we thus determined that one restriction-site change (other than 16517*Hae*III) from a founder haplotype requires 24,420 years, on average.

Previous estimations of coalescence times were based on mean divergence (π, estimated by use of the maximum likelihood procedure described by Nei and Tajima [1983]) and employed a 2.2%–2.9%/myr rate of mtDNA evolution (Torroni et al. 1994*c*). For a starlike haplogroup, such as the haplogroup V in this study, π indeed may be used as an estimator for coalescence time. In the case of haplogroup V, we calculated $\pi = 0.0332\%$ and then obtained a time estimate of 11,450–15,090 years. For the same data, we found $\rho_{RFLP} = 0.500$, which corresponds to 12,210 years according to the above calibration. This point estimate, in turn, is associated with a mutation rate of 2.72%/myr, which is within the range proposed by Torroni et al. (1994*c*). On the other hand, the rate of 2.55%/myr (the center of the range 2.2%–2.9%/myr) yields the time estimate of 13,020 years. We used this as a second calibration, by converting ρ_{RFLP} into time, so that we henceforth calibrated $\rho_{RFLP} = 1$, corresponding to 24,420 years/26,040 years, alternatively. Converted into HVS-I variation, $\rho_{HVS\text{-}I} = 1$ corresponds to 20,180 years/21,520 years.

In order to estimate the sampling error for ρ, we assumed a Poisson process, which is approximately justified for starlike haplogroups such as those described in this study. The SD is then calculated as $\sqrt{\rho/n}$, where n denotes the sample size. In the case of highly starlike haplogroups that are distributed over a wide geographical area (and, thus, are sampled from many distinct populations), the demographic error is expected to be negligible, compared with the sampling error. In order to quantify the starlikeness of a haplogroup, we defined the star index of a sample of sequences, relative to an estimated rooted tree, as the relative frequency of pairs of sequences in the sample that coalesce in the root (type) of the tree—that is, that are connected by a path passing through the root. A perfectly starlike group receives a star index of 1, and the less starlike the group, the closer the index is to 0. We regarded a haplogroup with a star index >.950 as highly starlike.

Results

The Population Distribution of Haplogroup V mtDNAs

The RFLP haplotype analysis showed that, among the 419 mtDNAs samples analyzed, 19 were from members of haplogroup V and encompassed eight haplotypes; these are given (along with three other haplotypes described in previous studies) in table 1. Haplotype 109, which is defined only by the basic combination −4577q/+7025a/−10394c, is the evident founder haplotype of this haplogroup. It encompasses 59.1% of the 22 haplogroup V mtDNAs; it is the only haplotype shared between populations; and it is central to the phylogeny of the haplogroup. Haplogroup V mtDNAs are not uniformly distributed. They were observed in 20.0% of the Basques, 10.4% of the northern Sardinians, 5.4% of the Swedes, and 4.1% of the Finns but were absent in the Tuscans, central Italians, Adygei, Druzes, Yemenites, and Yakuts (table 2).

To establish the correlation between RFLP haplotypes and control-region sequences for haplogroup V mtDNAs, we sequenced both hypervariable segments of the control regions from 23 mtDNAs belonging to this haplogroup (table 3). Haplogroup H and haplogroup V mtDNAs shared 73A in HVS-II. In addition, haplogroup V mtDNAs harbored the mutations 16298C in HVS-I and 72C in HVS-II (table 3). The stability of np 72 and, thus, its reliability as a marker for haplogroup V is confirmed by the observation that it has undergone, at most, one parallel mutation (Turk 8), reported in the data from the study by Calafell et al. (1996), in 1,166 worldwide HVS-I+II sequences, compiled from the studies by Vigilant et al. (1991), Ginther et al. (1993), Piercy et al. (1993), Santos et al. (1994), Batista et al. (1995), Graven et al. (1995), Kolman et al. (1995), Mountain et al. (1995), Reed et al. (1995), Calafell et al. (1996), Torroni et al. (1996), and Lee et al. (1997). Moreover, in an additional 241 worldwide HVS-II sequences (Jorde et al. 1995), 72C appears only once, in a French individual who is very likely to be a haplogroup V member, according to np 73 status.

To better define the distribution of haplogroup V, we then carried out an extensive survey of previously published data sets (mainly HVS-I) for the 16298C mutation. The survey of Caucasoid mtDNAs revealed the presence of the 16298C mutation in virtually all western, central, and northern European populations (table 4). Among these mtDNAs, there were only three British mtDNAs for which both HVS-I and HVS-II data were available (Piercy et al. 1993). Similar to our haplogroup V mtDNAs, all three harbored the motif 72C–73A in HVS-II (table 3). For all other European mtDNAs harboring 16298C, HVS-II data were not available to sup-

Table 1

RFLP Haplotypes of the mtDNAs Belonging to Haplogroups V and H

Haplotype[a]	Restriction-Site Change(s)[b]	No. of mtDNAs, by Population[c]									Total No. of mtDNAs
		1	2	3	4	5	6	7	8	9	
Haplogroup V (all −4577q):											
36	+1941f, +9209g	.	1	.	.	.	.	.	.	.	1
37	−8838e	.	1	.	.	.	.	.	.	.	1
38	−5261e	1	.	.	.	.	.	.	.	.	1
40	−14015a	1	.	.	.	.	.	.	.	.	1
109[d]	...	.	.	8	4	.	.	.	.	.	13
110	−8592j	.	.	1	.	.	.	.	.	.	1
111	+12026o/+12026h	.	.	.	1	.	.	.	.	.	1
112	+14276t/+14279e	.	.	1	.	.	.	.	.	.	1
US10[e]	−5261e, +16517c	.	.	.	.	.	.	.	.	.	1
US36[e]	+12810k, +16216o, +16517e	.	.	.	.	.	.	.	.	.	1
Total	...	2	2	10	5	.	.	.	.	.	22
Haplogroup H (all −7025a):											
1	...	2	.	5	.	1	1	.	.	.	9
2	+16517c	9	4	9	13	7	8	.	.	3	53
3	+1842k, +16517e	.	2	.	.	.	.	.	.	.	2
4	+10097a, +16517e	1	.	.	.	.	.	.	.	.	1
5	+8268a, +16517e	1	.	.	.	.	.	.	.	.	1
6	−5003c/ +5004r	1	.	.	.	.	.	1	.	.	2
7	−5003c/+5004r, +15882b/−15883e, +14347k	1	.	.	.	.	.	.	.	.	1
8	−13325k, +16517e	1	.	.	.	.	.	.	.	.	1
9	−5003c/+5004r, +14347k	.	1	.	.	.	.	.	.	.	1
10	+4769a	.	1	.	1	.	.	.	.	.	2
11	−951j, +4769a	2	1	.	.	.	1	.	.	.	4
12	+16170j, +16517e	1	.	.	.	.	.	.	.	.	1
13	−9644a	.	1	.	.	.	.	.	.	.	1
14	−9380f	.	1	.	.	.	.	.	.	.	1
15	+4332b/+4336q, −16303k	.	2	1	.	.	.	.	.	.	3
16	+14493c, −16303k	.	1	.	.	.	.	.	.	.	1
17	−9266e, +16215j, +16517c	.	1	.	.	.	.	.	.	.	1
18	−16049k, +16517e	1	.	.	.	.	.	.	.	.	1
53	+717g, +16517e	.	.	.	.	1	.	.	.	.	1
54	−3537a, +9156e	.	.	.	.	1	.	.	.	.	1
55	+7316j/+7319c, +16517e	.	.	.	.	1	.	.	.	.	1
56	+54k, +16517c	.	.	.	1	.	.	.	.	.	1
57	+7852j/−7853o, +16517e	.	.	.	1	.	.	.	.	.	1
58	+16254a, +16517e	.	.	.	2	.	.	.	.	.	2
59	+4793e, +16517e	.	.	.	2	.	.	.	.	.	2
60	+15614e/+15615f, +16517e	.	.	.	1	.	.	.	.	.	1
61	+13702e, +16517e	.	.	.	1	.	.	.	.	.	1
62	+9985j, +14223n/ +14224f, +16517c	.	.	5	.	.	.	.	.	.	5
63	−12406h	.	.	1	.	.	.	.	.	.	1
64	+3846c/−3849e, +16517e	.	.	1	.	.	.	.	.	.	1
65	+3846c/−3849e, −9553e, +16517e	.	.	1	.	.	.	.	.	.	1

(*continued*)

Table 1 (continued)

Haplotype[a]	Restriction-Site Change(s)[b]	No. of mtDNAs, by Population[c]									Total No. of mtDNAs
		1	2	3	4	5	6	7	8	9	
66	+10524k, −16208k, +16517e	.	.	1	.	.	.	.	.	.	1
67	+4332b/+4336q, +15882b/−15883e, −16303k, +16517c	.	.	1	.	.	.	.	.	.	1
122	+13100i	.	.	.	.	.	.	1	.	.	1
123	+9493g, −9553e, +16517e	.	.	.	.	.	.	1	.	.	1
124	+9336k (heteroplasmic), +16517e	.	.	.	.	.	.	1	.	.	1
125	+4793e, −6296c, +16517c	.	.	.	.	.	.	1	.	.	1
126	−14258m, +16517e	.	.	.	.	.	.	1	.	.	1
127	+9253e	.	.	.	.	.	2	.	.	.	2
128	−951j, +4769a, −5176a, +16517c	.	.	.	.	.	1	.	.	.	1
129	−16065g (heteroplasmic), +16517e	.	.	.	.	.	1	.	.	.	1
130	−1715c, +4793e, +16517e	.	.	.	.	.	1	.	.	.	1
169	−5176a	.	.	.	.	.	.	.	.	1	1
170	+4745k, −13325k, +14368g, −14869j, −7335l	.	.	.	.	.	.	.	.	1	1
171	+4769a, −16208k	.	.	.	.	.	.	.	1	.	1
Total	...	20	15	25	22	11	15	6	1	5	120

[a] Haplotypes 1–18, 36–38, and 40 have been reported elsewhere by Torroni et al. (1996).

[b] The restriction-site changes are relative to the reference sequence (Anderson et al. 1981). 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; o = *Hin*cII; q = *Nla*III; r = *Bfa*I; and t = *Bst*oI.

[c] 1 = Finnish; 2 = Swedish; 3 = Basque; 4 = northern Sardinian; 5 = central Italian; 6 = Adygei; 7 = Druze; 8 = Yemenite; and 9 = Yakut.

[d] An additional subject (an Italian LHON subject from Liguria [northwestern Italy]) harboring haplotype 109 was described elsewhere by Torroni et al. (1997). This additional haplogroup V mtDNA also was used for estimation of the age of the haplogroup.

[e] These mtDNAs are from Caucasians of North America and have been described elsewhere by Torroni et al. (1994*a*). They are included here because they have been used for estimation of the age of haplogroup V.

port their inclusion within haplogroup V. However, mutation 16298C in all these HVS-I sequences was never associated with the HVS-I motifs that define other European haplogroups (Richards et al. 1996; Torroni et al. 1996). In addition, we observed that np 16298 is not a fast-evolving site. Indeed, a survey of the 407 African HVS-I sequences reported by Watson et al. (1997) revealed only one subject, from Nigeria, harboring this mutation. In contrast, mutation 16298C is commonly observed in Asian/Native American mtDNAs. However, all these Asian/Native American mtDNAs harbor the distinguishing motifs 16223T–16298C, 16223T–16298C–16327T, and 16223T–16298C–16325C–16327T (Torroni et al. 1993*a,* 1993*b;* Kolman et al. 1996) and share mutation 16298C by descent. The latter two motifs characterize the Asian/Native American–specific haplogroup C, whereas the former motif is considered to be the ancestral Asian prehaplogroup C motif, as suggested by the phylogenetic tree in the study by Vigilant et al. (1991). None of these Asian motifs were observed among the published European HVS-I sequences harboring 16298C; thus, we included all of them within haplogroup V.

Figure 1 shows that haplogroup V mtDNAs reach particularly high frequencies in some Iberian populations (Basques and Catalonians). However, this haplogroup is also very common among the Berbers of North Africa (8%–11%) and, amazingly, reaches its highest frequencies among the Scandinavian Saami (40.9%). In contrast, haplogroup V appears to be absent among the populations of southeastern Europe and the Near East, with the exception of one Turk subject.

By considering both the HVS-I sequences obtained in this study and those published previously, we found 25

Table 2

Population Distribution of Haplogroup V mtDNAs

Region and Population	*n*[a]	%	Reference(s)
India:			
Southern Indian	98	...	Mountain et al. (1995)
Siberia:			
Yakut	51	...	Present study
Near East:			
Yemenite	38	...	Present study
Druze	45	...	Present study
Mixed Middle East	42	...	Di Rienzo and Wilson (1991)
Turkish	96	1.0	Calafell et al. (1996); Comas et al. (1996); Richards et al. (1996)
Caucasus:			
Adygei	50	...	Present study
Southeastern Europe:			
Bulgarian	30	...	Calafell et al. (1996)
Italy:			
Central Italian	51	...	Present study
Tuscan	48	...	Torroni et al. (1996)
Italian LHON subjects	37	2.7	Torroni et al. (1997)
South Tyrol/Trentino	70	4.3	Stenico et al. (1996)
Northern Sardinian	48	10.4	Present study
Mixed Sardinian	69	1.4	Di Rienzo and Wilson (1991)
North Africa:			
Algerian Berber	85	8.2	Côrte-Real et al. (1996)
Moroccan Berber	18	11.1	Pinto et al. (1996)
Iberia:			
Basque (Alava/Vizcaya)	61	3.3	Côrte-Real et al. (1996)
Basque (Guipuzcoa)	45	11.1	Bertranpetit et al. (1995); Côrte-Real et al. (1996)
Basque (Guipuzcoa)	50	20.0	Present study
Catalonian	15	26.7	Côrte-Real et al. (1996)
Canary Islands	54	3.7	Pinto et al. (1996)
Mixed Spanish	56	5.4	Côrte-Real et al. (1996)
Portuguese	54	3.7	Côrte-Real et al. (1996)
Central-northern Europe:			
Swiss	74	5.4	Pult et al. (1994)
Bavarian	49	6.1	Richards et al. (1996)
Northern German	107	4.7	Richards et al. (1996)
British	100	3.0	Piercy et al. (1993)
Welsh	92	3.3	Richards et al. (1996)
Cornish	69	1.4	Richards et al. (1996)
Danish	33	3.0	Richards et al. (1996)
Icelandic	53	1.9	Sajantila et al. (1995); Richards et al. (1996)
Swedish	37	5.4	Torroni et al. (1996); present study
Finnish	49	4.1	Torroni et al. (1996); present study
Finnish	29	3.4	Richards et al. (1996)
Finnish	32	9.4	Lahermo et al. (1996)
Estonian	28	...	Sajantila et al. (1995)
Volga Finnic	34	2.9	Sajantila et al. (1995)
Karelian	83	6.0	Sajantila et al. (1995)
Saami	115	40.9	Sajantila et al. (1995)

[a] Number of subjects analyzed for each population.

different types of haplogroup V sequences, encompassing a total of 129 mtDNAs (table 4). Sequence type 1, defined only by mutation 16298C is by far the most common (61.2% of all haplogroup V mtDNAs) and is central to the phylogeny of the haplogroup. Thus, it appears that 16298C is the founding motif of the haplogroup V HVS-I sequences (fig. 2). The motif 16298C–16153A (sequence type 2) is also common (12.4%) and is shared among populations. Its presence in the Berbers, Germans, Finns, Volga Finnics, and Saami could indicate that it predated the geographical diffusion of haplogroup V and that motifs 16298C and 16298C–16153A were carried together by the same expanding population(s). However, the lack of additional HVS-I variation accumulated in the basic motif 16298C–16153A, and the absence of this motif in the Iberian peninsula, could also indicate that mutation 16153A occurred more recently, possibly in a central-northern European population, and was carried to North Africa in historical times. Sequence type 9 is shared between two Finns and the only Near Eastern subject (Turk) harboring a haplogroup V sequence, suggesting some gene flow from central-northern Europe to Turkey. Some additional and less common motifs (16298C–16189C, sequence types 3 and 24; 16298C–16362C, sequence types 4 and 20; 16298C–16216G, sequence types 5 and 9; and 16298C–16311C, sequence types 6 and 21) are also shared by geographically distant populations. However, parallelisms at nps 16189, 16311, and 16362 cannot be excluded, since these are fast-evolving sites (Hasegawa et al. 1993; Wakeley 1993).

The Population Distribution of Haplogroup H mtDNAs

Among the 419 mtDNA samples studied by RFLP analysis, 120 were from members of haplogroup H and encompassed 45 haplotypes (table 1). Haplogroup H has two candidate founding haplotypes (haplotypes 1 [−7025a/−10394c] and 2 [−7025a/−10394c, +16517e]) distinguished by the hypervariable *Hae*III site at np 16517. Haplotypes 1 and 2 together represent >50% of haplogroup H mtDNAs and are central to the phylogeny of the remaining haplotypes. This haplogroup was common in almost all populations studied by RFLP analysis, particularly in those that also harbored haplogroup V mtDNAs, and showed a clinal distribution. Haplogroup H characterized 50.0% of the Basques, 45.8% of the Sardinians, 40.5% of the Swedes, 40.8% of the Finns, 21.6% of the central Italians, 30.0% of the Adygei, 13.3% of the Druzes, 2.6% of the Yemenites, and 9.8% of the Yakuts of Siberia (table 5).

A survey of published control-region sequences, similar to that carried out for haplogroup V, was also possible for haplogroup H by considering as members of

Table 3

HVS-I and HVS-II Variation of Haplogroup V mtDNAs

Sequence Type	Polymorphic Positions: 111111 666666 011122 758849 538988	112223333 779590060011 233050432955 aaab	No. of mtDNAs, by Population[a]: 123456	RFLP Haplotype(s)[b]
CRS[c]	TGCTCT	TAACTATA----		
1	-----C	CA-T---G-CC-	2-----	109, 112
2	-----C	CA-T---G-CCC	1-----	109
3	C--C-C	CA--CG-G-CC-	1-----	110
4	-----C	CA-----G-CC-	111---	109, 109, ND
5	--T--C	CA-----G--C-	-1----	109
6	-----C	CA-----G--C-	-1-28-	111, ND, ND
7	-A---C	CAG----G-CC-	--1---	ND
8	-----C	CA-----GCCC-	--1---	ND
9	-----C	CA----CG--C-	----2-	ND
10	----TC	CAG-C--G-C--	-----1	ND
11	-----C	CA-----GCC--	-----2	ND

[a] 1 = Basque; 2 = northern Sardinian; 3 = Finnish; 4 = Norwegian Saami; 5 = Skolt Saami; and 6 = British. The British data are from the study by Piercy et al. (1993).

[b] The RFLP haplotypes of the Finns, Saami, and British were not determined ("ND"). However, all Finnish and Saami mtDNAs were screened for the 4577 *Nla*III site, the absence of which defines haplogroup V; all of these mtDNAs lacked this site.

[c] Reference sequence (Anderson et al. 1981).

this haplogroup the control-region sequences that harbored 73A in HVS-II and that lacked 16298C in HVS-I. This additional data further support a clinal distribution of haplogroup H mtDNAs (table 5 and fig. 1). Haplogroup H showed its highest frequencies (40%–60%) in western and northern Europe, intermediate frequencies (20%–40%) in southern Spain, North Africa, central Italy, eastern Europe, Turkey, and the Caucasus, and frequencies <20% in the Near East, India, and central Siberia.

The Relationship of Haplogroups V and H

V and H are the only haplogroups harboring 73A in HVS-II and differ by only two RFLP sites in the coding regions (*Alu*I site at np 7025 and *Nla*III site at np 4577) and by two nucleotide positions in the control region (np 16298 in HVS-I and np 72 in HVS-II). In addition, Lamminen et al. (1997) recently have shown that both these haplogroups are also characterized by the transition T→C at np 14766, a mutation that is absent in all other European haplogroups and that eliminates an *Mse*I site. These observations reveal that haplogroups V and H are closely related to each other and suggest two possible explanations for this relatedness. They may be two sister groups that originated from a common ancestor only defined by 14766C and 73A. Alternatively, haplogroup V could have originated from a haplogroup H mtDNA, which has again acquired the *Alu*I site at np 7025. The latter possibility is, however, less parsimonious than the first, since it requires one additional mutational event. Moreover, screening for the *Mse*I-site loss at np 14766 has revealed that some mtDNAs with the sequence features of the common ancestor of haplogroups V and H are still present in Druzes, central Italians, Sardinians, and Finns.

Time Depths of Haplogroups V and H

For haplogroup V, π measured by RFLP analysis is 0.0332%, and the mean distance from the evident ancestral haplotype (haplotype 109) is $\rho_{RFLP} = 0.500 \pm 0.151$ (table 6). Therefore, the age of the haplogroup is estimated to be 12,200 ± 3,700 years/13,000 ± 3,900 years (see Subjects and Methods). On the other hand, the 82 HVS-I sequences sampled from all populations, except the Saami, which belong to haplogroup V (as judged from the sequence motif), show a mean distance of $\rho_{HVS\text{-}I} = 0.610 \pm 0.086$ from the inferred haplogroup V root sequence (no. 1 in table 4 and fig. 2). This yields an estimated age of 12,300 ± 1,700/13,100 ± 1,900 years (see Subjects and Methods), which is in perfect agreement with the estimate based on RFLPs. The numerous Saami mtDNAs belonging to haplogroup V (table 4) were not included in the above estimation, because their lack of variation suggests a very recent founder

Table 4

HVS-I Sequence Variation of Haplogroup V mtDNAs

Sequence Type	Mutation(s)[a]	No. of mtDNAs, by Population[b]																													Total No. of mtDNAs
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	
1	298	3	1	4	.	1	.	1	2	.	4	.	2	3	1	1	1	2	1	2	.	1	1	.	2	2	.	2	.	42	79
2	298, 153	.	.	.	.	.	.	.	.	2	1	.	.	.	.	.	2	.	.	.	.	.	.	.	2	3	1	1	.	4	16
3	298, 189	.	.	.	.	.	.	.	.	.	1	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3
4	298, 362	.	.	.	1	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3
5	298, 216	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
6	298, 311	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
7	298, 093	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	3
8	298, 257	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	2
9	298, 216, 261	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	1	.	3
10	298, 162	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
11	298, 231	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1
12	298, 325	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1
13	298, 291	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1
14	298, 169	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	1
15	298, 248	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1
16	298, 240	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
17	298, 188	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
18	298, 368	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
19	298, 086	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
20	298, 124, 362	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	1
21	298, 242, 311	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
22	298, 249, 362	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
23	298, 075, 240	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
24	298, 075, 189	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
25	298, 204t, 213t, 255t	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
Total	...	5	2	5	4	3	1	2	2	2	7	1	3	4	3	3	5	3	1	3	1	1	2	1	8	5	1	3	1	47	129

[a] All mutations listed are relative to CRS (Anderson et al. 1981), less 16,000, and are transitions, except for those with the suffix "t," which indicates a transversion; the combination of transversions in type 25 possibly is artifactual.

[b] 1 = Basque (Bertranpetit et a1. 1995); 2 = Basque (Côrte-Real et al. 1996; Richards et al. 1996); 3 = Basque (this study); 4 = Catalonian (Côrte-Real et al. 1996); 5 = Spanish (Côrte-Real et al. 1996; Richards et al. 1996); 6 = Spanish (Pinto et al. 1996); 7 = Portuguese (Côrte-Real et al. 1996; Richards et al. 1996); 8 = Canary Islands (Pinto et al. 1996); 9 = Morroccan Berber (Pinto et al. 1996); 10 = Algerian Berber (Côrte-Real et al. 1996); 11 = Sardinian (Di Rienzo and Wilson 1991); 12 = northern Sardinian (this study); 13 = Swiss (Pult et al. 1994); 14 = Mocheni (Stenico et al. 1996); 15 = Bavarian (Richards et al. 1996); 16 = northern German (Richards et al. 1996); 17 = British (Piercy et al. 1993); 18 = Cornish (Richards et al. 1996); 19 = Welsh (Richards et al. 1996); 20 = Icelandic (Richards et al. 1996); 21 = Danish (Richards et al. 1996); 22 = Swedish (Sajantila et al. 1996); 23 = Finnish (Richards et al. 1996); 24 = Finnish (Sajantila et al. 1995; Lahermo et al. 1996); 25 = Karelian (Sajantila et al. 1995); 26 = Volga Finnic (Sajantila et al. 1995); 27 = Finnish LHON patients (Lamminen et al. 1997); 28 = Turkish (Richards et al. 1996); and 29 = Saami (Sajantila et al. 1995; Lahermo et al. 1996).

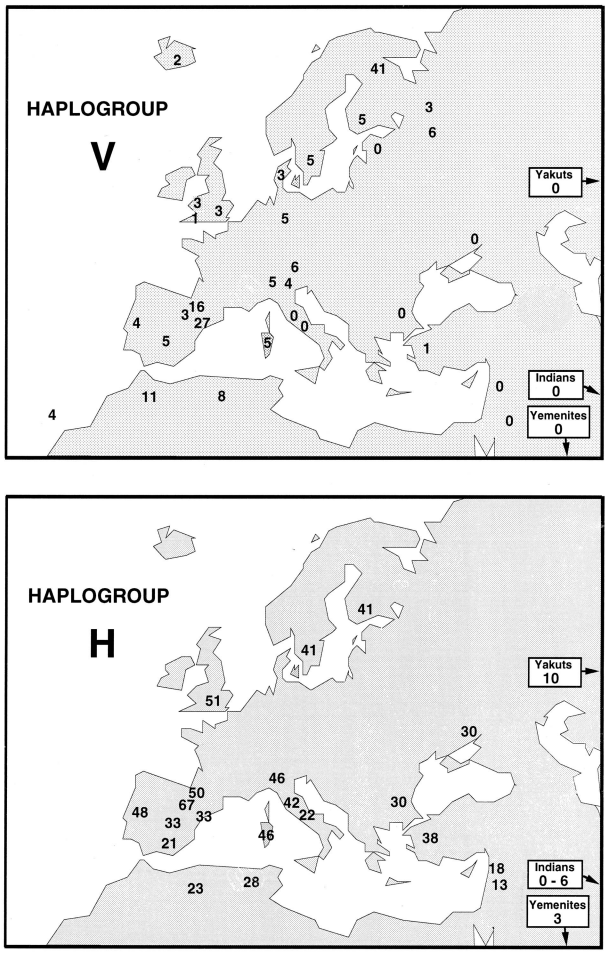

Figure 1 Maps illustrating the percent frequencies of mtDNAs belonging to haplogroups V and H.

event (see Discussion). The star index of haplogroup V (with respect to the sites scored for ρ) equals 0.996 for the RFLPs and 0.981 for the HVS-I sequences (or 0.969 when Saami sequences are excluded).

For haplogroup H, π inferred from RFLP data is 0.0574%, which is almost twice as high as that for haplogroup V, and corresponds to an age of 19,800–26,100 years, when the mtDNA evolution rate of 2.2%–2.9%/myr is used. Obvious candidates for the root haplotype of H are the two most frequent haplotypes (nos. 1 and 2, table 1 and fig. 3), which are distinguished only by the hypervariable *Hae*III site at np 16517. This site is ignored in the calculation of ρ_{RFLP} and the star index of haplogroup H (with respect to the restriction sites scored for ρ) = 0.991. Thus, the mean distance to the root haplotype is $\rho_{RFLP} = 0.750 \pm 0.079$, yielding the ages of 18,300 ± 1,900 years/19,500 ± 2,100 years. However, there is considerable difference in the divergence of H for the populations studied, suggesting a geographical pattern. The Near East populations, represented by the Druzes and the Yemenite Jews, show a total absence of the otherwise shared haplotypes 1 and 2 and give a value of $\rho_{RFLP} = 1.143 \pm 0.404$. In contrast, the western/southern European populations (Basques, northern Sardinians, and central Italians) give a value of $\rho_{RFLP} = 0.569 \pm 0.099$, whereas the northern European populations (Finns and Swedes) are in between, with $\rho_{RFLP} = 0.886 \pm 0.159$. These European populations share, with the Near East populations, three branches with the characteristic sites +4769a, −5003c/+5004r, and +4793e (table 1 and fig. 3). In a first attempt to dissect haplogroup H according to geography, we grouped all northwestern European mtDNAs, except for those 10 (haplotypes 6, 7, 9, 10, 11, and 59) located in the three branches shared by northwestern Europeans and the Near East, into a group with potentially western European affinity. This subsample comprised 83 individuals and gave a value of $\rho_{RFLP} = 0.578 \pm 0.083$, whereas the other individuals together yielded a value of $\rho_{RFLP} = 1.135 \pm 0.175$, which is close to the value for the small Near East sample alone. The age of haplogroup H in western Europe then would be estimated to be 14,100 ± 2,000 years/15,100 ± 2,200 years, whereas the age of haplogroup H in the rest of Eurasia is estimated to be 27,700 ± 4,300 years/29,600 ± 4,600 years (table 6). However, the age estimate for the expansion time of haplogroup H in western Europe

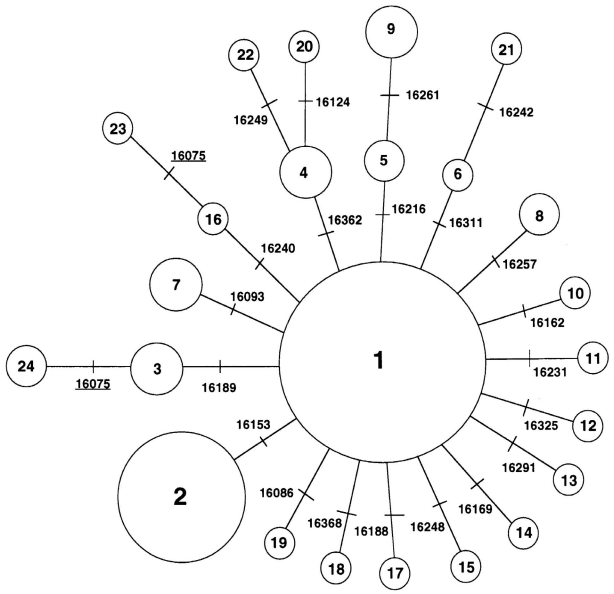

Figure 2 Unique, most parsimonious tree for HVS-I haplogroup V. The designation of sequence types and variants is according to table 4. The areas of the circles are proportional to the number of individuals. Underlining indicates a parallelism. Haplogroup V mtDNAs from the Saami were not included (see text).

Table 5

Population Distribution of Haplogroup H mtDNAs

Region and Population	n[a]	%	Reference(s)
India:			
Andhra Pradesh	40	...	Passarino et al. (1996)
Southern Indian	98	1.0	Mountain et al. (1995)
New Delhi	56	1.8	Passarino et al. (1996)
Punjab	67	6.0	Passarino et al. (1996)
Siberia:			
Yakut	51	9.8	Present study
Near East:			
Yemenite	38	2.6	Present study
Druze	45	13.3	Present study
Lebanese	50	18.0	Passarino et al. (1996)
Turkish	29	37.9	Calafell et al. (1996)
North Africa:			
North African	39	28.2	Passarino et al. (1996)
Algerian Berber	85	23.5	Côrte-Real et al. (1996)
Caucasus:			
Adygei	50	30.0	Present study
Southeastern Europe:			
Bulgarian	30	30.0	Calafell et al. (1996)
Northern Europe:			
Finnish	49	40.8	Torroni et al. (1996); present study
Swedish	37	40.5	Torroni et al. (1996); present study
British	100	51.0	Piercy et al. (1993)
Italy:			
Northern Italian	22	45.5	Passarino et al. (1996)
Northern Sardinian	48	45.8	Present study
Tuscan	48	41.7	Torroni et al. (1996)
Central Italian	51	21.6	Present study
Iberia:			
Basque (Alava/Vizcaya)	61	67.2	Richards et al. (1996)
Basque (Guipuzcoa)	50	50.0	Present study
Portuguese	54	48.1	Côrte-Real et al. (1996)
Northern Spanish	30	33.3	Côrte-Real et al. (1996)
Catalonian	15	33.3	Côrte-Real et al. (1996)
Mixed Spanish	9	33.3	Côrte-Real et al. (1996)
Andalusian	15	21.4	Côrte-Real et al. (1996)

[a] Number of subjects analyzed for each population.

could be even lower. Indeed, there is some evidence for another minor haplogroup H–founder haplotype in western Europe. This possible root haplotype is characterized by −16303k (owing to a T→C transition at np 16304) and gave rise to two branches that encompass haplotypes 15 (two Swedes and one Basque), 16 (one Swede), and 67 (one Basque) (fig. 3). Although this node is not represented in this sample, HVS-I data does indeed support its existence in Spain and among the Basques (Côrte-Real et al. 1996; Richards et al. 1996). If we include this as a potential root, then the ρ_{RFLP} would be reduced to a value of 0.518 ± 0.079, giving an age estimate of 12,600 ± 1,900 years/13,500 ± 2,100 years for the expansion time of haplogroup H. Averaging over the age estimates for haplogroup V (both RFLPs and HVS-I sequences) and the western branch of haplogroup H, we obtain an age of 12,300 years/13,200 years, which corresponds to $\rho_{HVS\text{-}I} = 0.611 \pm 0.057$ scaled to HVS-I, for the combined sample of $n = 187$ sequences/RFLPs. Appreciating the sampling error for ρ (± 0.057), all resulting age estimates fall into the range of 11,100–14,400 years.

Discussion

The Origin and Diffusion of Haplogroups V and H

Haplogroup H is the most common haplogroup in all European populations and reaches its highest frequencies (40%–60%) in western and northern Europe. This haplogroup is also common in the Caucasoid populations of the Near East and North Africa and is also observed in northern India and among the Yakuts. Even though this haplogroup is more common in Europe than in the Near East, analysis of sequence divergence appears to indicate that haplogroup H harbors a much higher diversity in the Near East than in Europe (table 6). These divergence values suggest that haplogroup H originated in the Near East ~25,000–30,000 years ago and expanded into Europe before the Second Pleniglacial (15,000–20,000 years ago; Otte 1990). The Turkish sequences (Calafell et al. 1996) belonging to haplogroup H (characterized by 73A) could serve to provide an upper time boundary for the expansion of haplogroup H into Europe. These sequences yield an HVS-I value of $\rho = 1.200 \pm 0.346$, which correspond to 24,200 ± 7,000 years/25,800 ± 7,500 years. The origin of haplogroup H in the Near East is in agreement with a Near East origin for most of the European gene pool and is compatible with its presence also in India and central Asia. However, an expansion of haplogroup H into Europe 20,000–25,000 years ago would place this expansion at a time that is intermediate between the appearance of modern *Homo sapiens* in Europe (>40,000 years ago) and the Neolithic expansion (starting ~10,000 years ago) (Cavalli-Sforza et al. 1994) and suggests that haplogroup H could represent a second Paleolithic wave (Richards et al. 1996) that was contemporary with the diffusion of the Gravettian technology (20,000–25,000 years ago).

Haplogroup V has a much more limited geographical distribution, and it is observed only in northwestern Eu-

Table 6

Expansion Time Estimates, for Haplogroups V and H, Based on RFLPs

	Haplogroup V	Haplogroup H	
		Western Europe[a]	Other Eurasia
Sample size	22	83	37
No. of haplotypes	10	27	20
ρ_{RFLP}[b]	.500 ± .151	.518 ± .079 or .578 ± .083	1.135 ± .175
Age (years)[c]	12,200 ± 3,700/13,000 ± 3,900	12,600 ± 1,900/13,500 ± 2,100 or 14,100 ± 2,000/15,100 ± 2,200	27,700 ± 4,300/29,600 ± 4,600

[a] Comprising haplotypes of populations 1–5 from table 1, except for haplotypes 6, 7, 9, 10, 11, and 59 (see text).

[b] The variation of the *Hae*III site at np 16517 was disregarded.

[c] Calculated by calibrating $\rho_{RFLP} = 1$ to 24,420 years/26,040 years.

rope and North Africa. It reaches high frequencies in some Iberian populations, is also very common among the Berbers of North Africa, but showed its highest frequencies (40.9%) among the Scandinavian Saami. Thus, this distribution indicates that, in contrast with haplogroup H, haplogroup V did not originate in the Near East but either in Europe or North Africa. In addition, the sequence divergence estimates suggest that haplogroup V is more recent than haplogroup H and originated only 12,200 ± 3,700 years/13,000 ± 3,900 years ago. The observed frequency values of haplogroup V could be interpreted to suggest three most likely origins

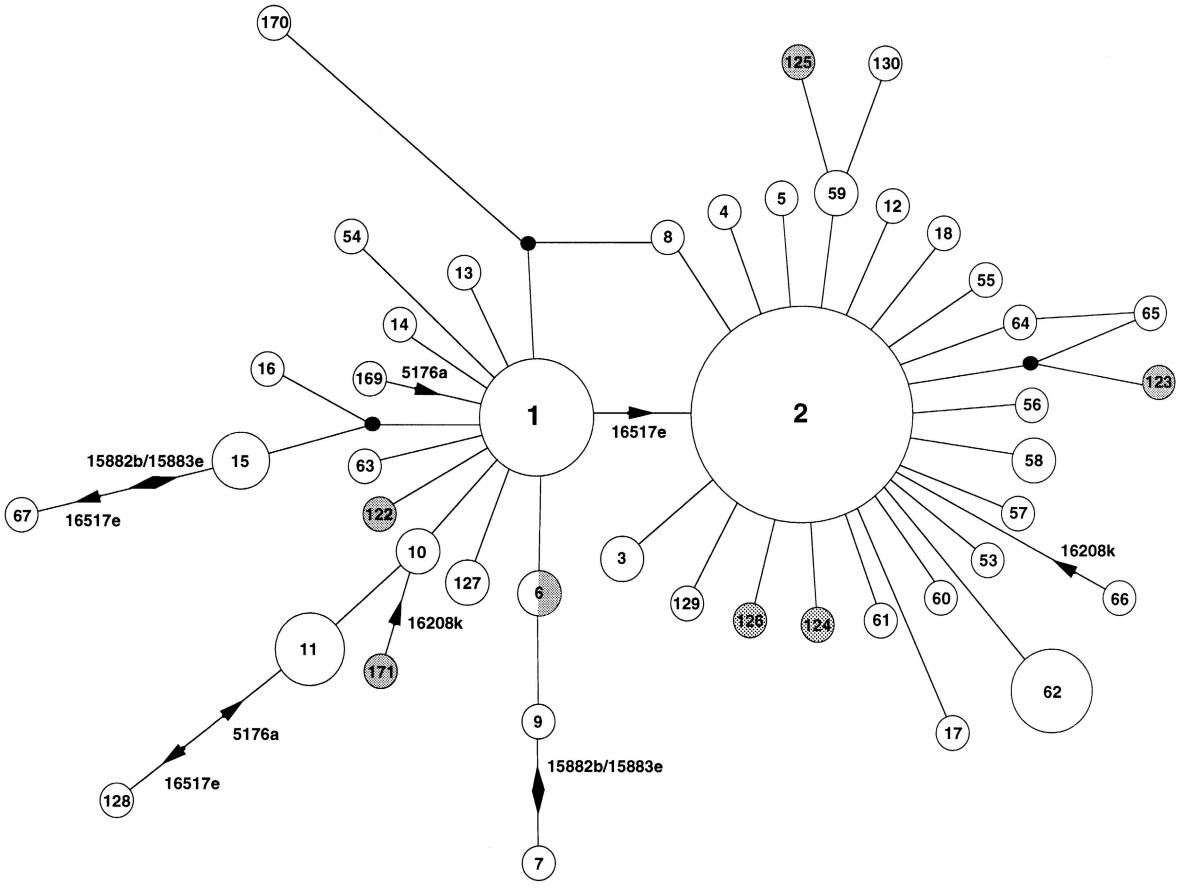

Figure 3 Network for RFLP haplogroup H, comprising the most parsimonious trees. The designation of haplotypes and restriction sites is according to table 1. The areas of the circles are proportional to the number of sampled individuals. Shaded circles indicate the haplotypes observed in the Near East. Recurrent mutations postulated by parsimony are indicated along the links, where each arrow points to the presence of the site. However, it is most likely that not all recurrent mutations at the hypervariable site 16517e can be reconstructed. Haplotype 1 is the probable root of haplogroup H.

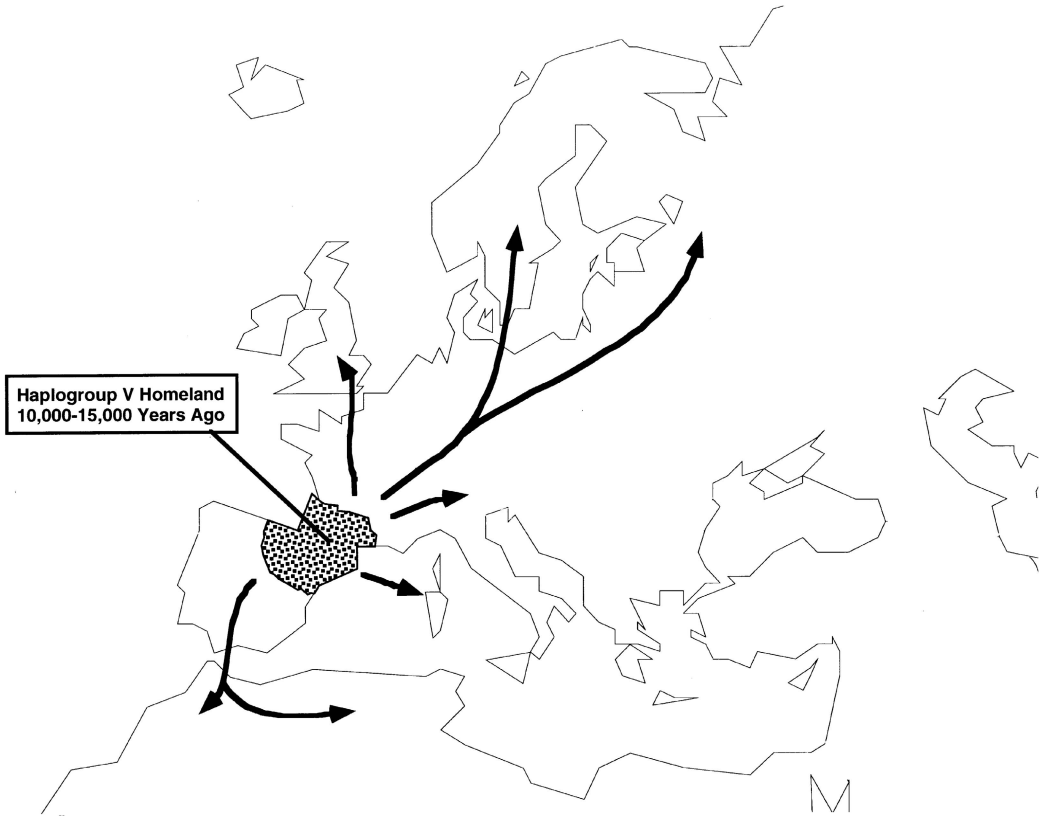

Figure 4 Map of Europe depicting the most likely homeland of haplogroup V and its pattern of diffusion.

for this haplogroup: North Africa, the Iberian peninsula, and the Saami. However, mtDNA haplogroup frequencies can be strongly affected by genetic drift and founder events; therefore, the identification of the most likely homeland of haplogroup V requires analysis of the extent of variation accumulated within the haplogroup, in its entire range of geographical distribution. The average number of HVS-I nucleotide differences from the founder–haplogroup V sequence is 0.09 ± 0.04 in the Saami, 0.56 ± 0.25 in North Africa, 0.77 ± 0.19 in the Iberian peninsula, and 0.63 ± 0.11 in a pool of all remaining European populations. This observation suggests that the Iberian peninsula is the most likely homeland of haplogroup V.

An origin of haplogroup V in the Iberian peninsula is supported not only by the highest degree of diversity but also by the climatic and demographic situation of Europe at the time in which haplogroup V originated. Indeed, an age of 12,200–13,000 years would place its origin during the Younger Dryas glacial interlude or earlier, during the first warm period after the Second Pleniglacial. This warm period is termed the Bölling/Alleröd period and occurred ~14,700–12,600 calendar years ago (Björck et al. 1996). The Second Pleniglacial was associated with extremely cold and dry climate conditions that, by 18,000 years ago, caused an almost complete retreat of people from the central plains of Europe. Southern France and the Iberian peninsula, in western Europe, and southern Ukraine in eastern Europe served as refuges, with people in the two areas being isolated from each other (Otte 1990). During the Bölling/Alleröd period, people returned to the northern areas, and Upper Magdalenian industries similar to those of southwestern France began to diffuse into northern France, Belgium, the Rhine region, the Swiss and Swebian plateaus, Bavaria, Thuringia, Bohemia, Moravia, and Lower Poland. The diffusion of this industry has led Otte (1990) to hypothesize a major population expansion from southwestern to central Europe, at the end of the Second Pleniglacial. This hypothesis is in complete agreement with the population distribution of haplogroup V, its coalescence time, and its diversity in different geographical regions. Indeed, all of these data suggest that this haplogroup originated in a population of the Iberian peninsula/southern France and expanded into central-northern Europe after the end of the Second Pleniglacial (fig. 4).

Obviously, haplogroup V was not the only haplogroup present in the western Europe refuge, at the end of the Second Pleniglacial. The age estimates indicate

that haplogroup H arrived in Europe from the Near East before the Second Pleniglacial. This haplogroup is common in all European populations but reaches the highest frequencies (50%–60%) among the Basques. Thus, it is most likely that haplogroup H was the most common haplogroup in the populations living in southwestern Europe during the Second Pleniglacial and that it expanded again into central-northern Europe together with haplogroup V, during the Bölling/Alleröd period. Indeed, a particularly high incidence of haplogroup H in this expanding population fits with the observation that all modern central-northern European populations harboring haplogroup V mtDNAs are also defined by very high frequencies (40%–50%) of haplogroup H.

Other Genetic Evidence for a Late Paleolithic Population Expansion from Southwestern to Northeastern Europe

Cavalli-Sforza et al. (1994) have recently summarized genetic data available from populations all over the world. These authors have introduced the methodology of synthetic genetic maps, which are geographic maps of isopleths of principal components values (PCs). With regard to Europe, PCs were calculated as optimized linear functions of the gene frequencies at 95 loci. Their second synthetic map of Europe shows a concentric gradient centered in the Iberian peninsula. To explain this result, Menozzi et al. (1978) and Cavalli-Sforza et al. (1994) took into account the possibility of a demic expansion from Iberia, but they considered an interpretation of this map on the basis of a migration from this area unlikely, because it was not substantiated by other nongenetic evidence. Thus, they considered the gradient most likely to be due to either climatic and ecological effects or a gene flow caused by one or more migrations of Uralic speakers from northwestern Asia.

Y-chromosome data are also in agreement with a demic expansion from the same region. Semino et al. (1996) have recently reported the population distribution of haplotype 15, detected by probes 49a/f. This haplotype is virtually absent in the Near East and eastern Europe, but it shows a clear gradient of frequency centered in western Europe. The peak of this gradient is again in the Iberian Peninsula/southwestern France, where haplotype 15 shows a frequency of ~60% among the Basques. Thus, it appears that mtDNA, autosomal, and Y-chromosome data are all concordant in indicating that a Paleolithic population living in the Iberian peninsula/southwestern France has contributed substantially to the gene pool of modern western and northern Europeans.

The mtDNA Variation in the Saami, the Finns, and the Yakuts

The Saami, together with a few other European populations including the Finns, speak a language belonging to the Uralic linguistic group (Ruhlen 1987). Linguistic and nuclear-gene data suggest that these populations originated in western Siberia and extensively admixed with European populations, with the Saami having 82% European and 18% Samoyed (Uralic speaking) admixture and the Finns harboring 90% European and 10% Uralic genes (Guglielmino-Matessi et al. 1990; Cavalli-Sforza and Piazza 1993; Cavalli-Sforza et al. 1994). Our study shows that >40% of modern Saami harbor mtDNAs belonging to haplogroup V and that this haplogroup is found only in western and northern European populations. However, this high incidence of haplogroup V in the Saami is associated with very low sequence variation and with only the two most common HVS-I sequences (nos. 1 and 2; table 4), encompassing virtually all subjects. These findings suggest that the Saami have acquired haplogroup V recently from some other northern European population(s) and that the introduction of this haplogroup has been associated with a strong founder event in the maternal lineage. Sajantila et al. (1995) observed that, in addition to the motif defined by 16298C (haplogroup V), the Saami also are characterized by a very high incidence (37%) of the motif 16144C–16189C–16270T. This Saami motif is a direct derivative of the motif 16189C–16270T, which is observed in a wide range of European populations (Côrte-Real et al. 1996; Richards et al. 1996; Stenico et al. 1996) and defines a subset of haplogroup U (Torroni et al. 1996). Thus, two limited subsets (haplogroup V and a subgroup of haplogroup U) of the European mtDNA variation appear to encompass $\geqslant$78% of the Saami mtDNAs, further indicating that founder effects and genetic drift have played a major role in shaping the mtDNA variation of this population.

An almost completely European origin of the Saami mtDNAs is also supported by the data from Lahermo et al. (1996), which have shown that $\leqslant$6.3% of the Saami mtDNAs are members of Asian superhaplogroup M. The European origin of mtDNA is even clearer for the Finns. This population harbors all nine European mtDNA haplogroups at frequencies that are very similar to those observed in other central-northern European populations (Richards et al. 1996; Torroni et al. 1996), and only 2.0% of their mtDNAs belong to haplogroup M (Lahermo et al. 1996; Torroni et al. 1996).

A completely, or almost completely, European origin of the Saami and Finnish mtDNAs might appear to be in partial disagreement with the linguistic and nuclear-gene data. However, recent Y-chromosome data may have resolved part of this incongruency. Zerjal et al.

(1997) have observed a novel Y-chromosomal T→C transition in some central-eastern Asian populations, including the Yakuts, in which this mutation reaches its highest frequency (85.7%). This transition is also very common in the Saami (25.0%) and the Finns (52.4%), but it is absent in the non–Uralic-speaking populations of western and central Europe. Thus, Y-chromosome data indicate that the Saami, the Finns, and other Uralic-speaking populations share part of their gene pool with western-central Asian populations, but the almost complete absence of Asian mtDNAs in the Finns and the Saami also suggests that the gene flow between the proto–Uralic-speaking population(s) and the ancestors of some western-central Asian populations could have been mainly male-mediated.

To confirm this hypothesis, it was necessary to determine whether the Yakut mtDNAs belonged to either European or Asian haplogroups. In the course of this study, we have observed that European haplogroups encompass only 17.6% of the Yakut mtDNAs. In addition to haplogroup H (9.8%), the Yakuts also harbor low frequencies of mtDNAs belonging to haplogroups J (3.9%), T (2.0%), and U (2.0%). In contrast, the various subsets of superhaplogroup M, including haplogroups C and D (Torroni et al. 1994*b*), encompass 66.7% of the Yakut mtDNAs. Thus, in contrast with the Finns and the Saami, a large majority of the Yakut mtDNAs are of Asian origin. This finding further supports a western European origin of the Saami and the Finnish mtDNAs.

In conclusion, this analysis confirms that a prior phylogenetic dissection of human mtDNA variation into haplogroups, followed by a differentiated and detailed analysis of each haplogroup, is necessary to determine the time and patterns of human expansions into different regions of the world and the genetic origin of modern human populations (Bandelt and Forster 1997). In particular, we show that a population living in the Iberian peninsula/southern France before the Younger Dryas has contributed substantially to the gene pool of all modern populations of central-northern Europe, including the Saami and the Finns. This late Paleolithic population expansion from southwestern to northeastern Europe is not only supported by mtDNA data but also by Y-chromosome and autosomal data and by archaeological records.

Acknowledgments

We are indebted to Kenneth K. Kidd and Judith R. Kidd (Yale University) for providing the Yakut, Yemenite, and Druze DNA samples and for their helpful comments on the manuscript. We also thank Martin Richards and Vincent Macaulay (Oxford, U.K.) for their critical advice and suggestions. The work was supported by Téléthon grant 921 (to A.T.), P.F. Beni Culturali contract 96.01182.PF36 (to R.S.), Grandi Progetti Ateneo (Ministero dell'Università e della Ricerca Scientifica e Tecnologica, 60% to R.S.), and Italian Consiglio Nazionale delle Ricerche grant 97.04297.CT04 (to A.T.).

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

Bandelt H-J, Forster P (1997) The myth of bumpy hunter-gatherer mismatch distributions. Am J Hum Genet 61: 980–983

Batista O, Kolman CJ, Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panama. Hum Mol Genet 4:921–929

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

Björck S, Kromer B, Johnsen S, Bennike O, Hammarlund D, Lemdahl G, Possnert G, et al (1996) Synchronized terrestrial-atmospheric deglacial records around the North Atlantic. Science 274:1155–1160

Brown MD, Sun F, Wallace DC (1997) Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 and 14484 mutations on an mtDNA lineage. Am J Hum Genet 60:381–387

Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann Hum Genet 60:35–49

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Cavalli-Sforza LL, Piazza A (1993) Human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur J Hum Genet 1:3–18

Chen Y-S, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57:133–149

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

Côrte-Real HBSM, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, et al (1996) Genetic diversity in the Iberian peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60:331–350

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

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin

and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, et al (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. In: Penna SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) DNA fingerprinting: state of the science. Birkhäuser, Basel, pp 211–219

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti AS, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345

Guglielmino-Matessi CR, Piazza A, Menozzi P, Cavalli-Sforza LL (1990) Uralic genes in Europe. Am J Phys Anthropol 83: 57–68

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol 37:347–354

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, et al (1995) Origin and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngöbé Amerinds of Panama. Genetics 140:275–283

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

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

Lamminen T, Huoponen K, Sistonen P, Juvonen V, Lahermo P, Aula P, Nikoskelainen E, et al (1997) mtDNA haplotype analysis in Finnish families with Leber hereditary optic neuropathy (LHON). Eur J Hum Genet 5:271–279

Lee SD, Shin CH, Kim KB, Lee YS, Lee JB (1997) Sequence variation in mitochondrial DNA control region in Koreans. Forensic Sci Int 87:99–116

Menozzi P, Piazza A, Cavalli-Sforza LL (1978) Synthetic maps of human gene frequencies in Europe. Science 201:786–792

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992

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

Otte M (1990) The northwestern European plain around 18,000 BP. In: Soffer O, Gamble C (eds) The world at 18,000 BP. Vol 1. Unwin Hyman, London, pp 54–68

Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS (1996) Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. Am J Hum Genet 59:927–934

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

Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60: 321–330

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

Reed AJ, Takezaki N, Sherry ST, McGarvery ST, Sofro AS, Stoneking M (1995) Evolutionary history of the COII/t-RNA^{Lys} intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

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

Ruhlen M (1987) A guide to the world languages. Stanford University Press, Stanford, CA

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

Sajantila A, Salem A-H, Savolainen P, Bauer K, Gierig C, Pääbo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci USA 93:12035–12039

Santos M, Ward RH, Barrantes R (1994) mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. Hum Biol 66:963–977

Scozzari R, Cruciani F, Santolamazza P, Sellitto D, Cole DEC, Rubin LA, Labuda D, et al (1997) mtDNA and Y chromosome–specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. Am J Hum Genet 60:241–244

Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the Neolithic demic diffusion in Europe through two Y chromosome–specific markers. Am J Hum Genet 59:964–968

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

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

Torroni A, Lott MT, Cabell MF, Chen Y-S, Lavergne L, Wallace DC (1994*a*) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am J Hum Genet 55:760–776

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994*b*) 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*c*) 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, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, Carrara F, Carducci C, et al (1997) Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. Am J Hum Genet 60: 1107–1121

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 mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993*b*) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53: 591–608

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

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

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

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA Jr, Santos RV (1996) Mitochondrial DNA polymorphism in three Brazilian Indian tribes. Am J Hum Biol 8:317–323

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

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, et al (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183