

Report

The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool

Alessandro Achilli,¹ Chiara Rengo,¹ Chiara Magri,¹ Vincenza Battaglia,¹ Anna Olivieri,¹ Rosaria Scozzari,^{2,3} Fulvio Cruciani,² Massimo Zeviani,⁴ Egill Briem,⁴ Valerio Carelli,⁵ Pedro Moral,⁶ Jean-Michel Dugoujon,⁷ Urmaz Roostalu,⁸ Eva-Liis Loogväli,⁸ Toomas Kivisild,⁸ Hans-Jürgen Bandelt,⁹ Martin Richards,¹⁰ Richard Villems,⁸ A. Silvana Santachiara-Benerecetti,¹ Ornella Semino,¹ and Antonio Torroni¹

¹Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy; ²Dipartimento di Genetica e Biologia Molecolare, Università "La Sapienza," and ³Istituto di Biologia e Patologia Molecolare, Consiglio Nazionale delle Ricerche, Rome; ⁴Unit of Molecular Neurogenetics, Pierfranco and Luisa Mariani Center for the Study of Children's Mitochondrial Disorders, National Neurological Institute Carlo Besta, Milan; ⁵Dipartimento di Scienze Neurologiche, Università di Bologna, Bologna; ⁶Departament de Biologia Animal, Universitat de Barcelona, Barcelona; ⁷Centre d'Anthropologie, UMR 8555, Toulouse, France; ⁸Department of Evolutionary Biology, University of Tartu and Estonian Biocentre, Tartu, Estonia; ⁹Fachbereich Mathematik, Universität Hamburg, Hamburg; and ¹⁰Schools of Biology and Computing, University of Leeds, Leeds

Complete sequencing of 62 mitochondrial DNAs (mtDNAs) belonging (or very closely related) to haplogroup H revealed that this mtDNA haplogroup—by far the most common in Europe—is subdivided into numerous sub-haplogroups, with at least 15 of them (H1–H15) identifiable by characteristic mutations. All the haplogroup H mtDNAs found in 5,743 subjects from 43 populations were then screened for diagnostic markers of subhaplogroups H1 and H3. This survey showed that both subhaplogroups display frequency peaks, centered in Iberia and surrounding areas, with distributions declining toward the northeast and southeast—a pattern extremely similar to that previously reported for mtDNA haplogroup V. Furthermore, the coalescence ages of H1 and H3 (~11,000 years) are close to that previously reported for V. These findings have major implications for the origin of Europeans, since they attest that the Franco-Cantabrian refuge area was indeed the source of late-glacial expansions of hunter-gatherers that repopulated much of Central and Northern Europe from ~15,000 years ago. This has also some implications for disease studies. For instance, the high occurrence of H1 and H3 in Iberia led us to re-evaluate the haplogroup distribution in 50 Spanish families affected by nonsyndromic sensorineural deafness due to the A155G mutation. The survey revealed that the previously reported excess of H among these families is caused entirely by H3 and is due to a major, probably nonrecent, founder event.

For most of human evolution, and particularly during the recent process of diffusion from Africa to the other continents, the relatively fast evolution of human mitochondrial DNAs (mtDNAs) has occurred in a context of small founding populations. Thus, founder events and genetic drift have played a major role in shaping hap-

lotype frequencies, giving rise to haplogroups and sub-haplogroups that are often restricted to specific geographic areas and/or population groups. In Europe, with the exception of U5 and V, which most likely arose in situ, all mtDNA haplogroups (H, I, J, K, T, U2e, U3, U4, X, and W) are most likely of Middle Eastern origin and were introduced by either the protocolonization ~45–40 thousand years ago (kya), by later arrivals in the Middle/Late Upper Paleolithic, Neolithic dispersals, or by more recent contacts (Torroni et al. 1998; Richards et al. 2000). For some haplogroups, particularly the more common ones, multiple chronologically distinct arrivals to Europe are extremely likely. In addition, the

Received July 16, 2004; accepted for publication August 30, 2004; electronically published September 20, 2004.

Address for correspondence and reprints: Dr. Antonio Torroni, Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Ferrata 1, 27100 Pavia, Italy. E-mail: torroni@ipvgen.unipv.it

© 2004 by The American Society of Human Genetics. All rights reserved. 0002-9297/2004/7505-0019\$15.00

genetic landscape of Europe has probably been further confounded by the major climatic changes that have occurred since the arrival of the first modern humans. In particular, the early Paleolithic populations of Northern and Central Europe either became extinct or retreated to the south during the Last Glacial Maximum (LGM) ~20 kya, and there was a gradual re peopling from southern refuge areas only when climatic conditions improved, from ~15 kya. This scenario is supported not only by recent work on archaeological dating (Housley et al. 1997; Richards 2003) but also by the phylogeographic evidence provided by mtDNA haplogroup V (Torroni et al. 1998; 2001a) and Y-chromosome haplogroups R1b and I1b2 (Semino et al. 2000; Cinnioglu et al. 2004; Rootsi et al. 2004).

Among the mtDNA haplogroups of Europe, haplogroup H displays two unique features: an extremely wide geographic distribution and a very high frequency in most of its range. Indeed, it is by far the most prevalent haplogroup in all European populations except the Saami, is very common in North Africa and the Middle East, and retains frequencies of 5%–10% even in northern India and Central Asia, at the edges of its distribution range (Richards et al. 2002).

Previous studies have proposed that haplogroup H (i) originated in the Middle East ~30–25 kya; (ii) expanded into Europe in association with a second Paleolithic wave, possibly contemporary with the diffusion of the Gravettian technology (25–20 kya); and (iii) was strongly involved in the late-glacial expansions from ice-age refugia after the LGM (Torroni et al. 1998; Richards et al. 2000). In addition, because of its high frequency and wide distribution, haplogroup H most likely participated in all subsequent episodes of putative gene flow in western Eurasia, such as the Neolithic diffusion of agriculture from the Near East, the expansion of the Kurgan culture from southern Ukraine, and the recent events of gene flow to northern India.

As a result, it is likely that the dissection of H into subhaplogroups of younger age might reveal previously unidentified spatial frequency patterns, which in turn could be correlated to prehistoric and historical migratory events. However, until now, haplogroup H has been only partially resolved genealogically (Herrnstadt et al. 2002) allowing for the identification of 11 subclades (H1–H11) (Quintáns et al. 2004; Loogväli et al. 2004), the phylogeography of which has been evaluated only in rare instances (Tambets et al. 2004). Therefore, the objective of this study is to provide new information concerning the molecular dissection of haplogroup H and to determine whether its subhaplogroups do indeed show such spatial patterns.

To achieve this objective, the first step consisted in the complete sequencing of 62 mtDNAs performed as described by Torroni et al. (2001b). Fifty-four of the

mtDNAs that were chosen for complete sequencing harbored –7025 *AluI* and –14766 *MseI*, two well-known diagnostic RFLP markers of haplogroup H. In addition, for the choice of these mtDNAs, we also took into account the nature and extent of the sequence variation observed in a preliminary sequence analysis restricted to the control region; the objective being to include the widest possible range of haplogroup H internal variation. The remaining eight mtDNAs were chosen because the RFLP analysis and control-region sequencing had suggested that they belonged to haplogroups that were closely related to H. Thus, their complete sequences would allow the definition of the branching order of the entire superhaplogroup HV.

A tree of the 62 complete mtDNA sequences (authors' Web site; GenBank) is illustrated in figure 1, which also incorporates information from previous studies about shared mutations in minor subbranches (Ingman et al. 2000; Fennilä et al. 2001; Herrnstadt et al. 2002, 2003; Mishmar et al. 2003; Coble et al. 2004). The phylogeny reveals that superhaplogroup pre-HV first splits into the minor haplogroup (pre-HV)1 and the major haplogroup HV, in which pre-V, HV1, and other branches of HV are all sister haplogroups of H (Macaulay et al. 1999; Torroni et al. 2001a).

As for the haplogroup H mtDNAs, the phylogenetic analysis confirmed a very large number of independent basal branches, some giving rise to subclades that have several basal subbranches themselves (fig. 1). Among these subclades, representatives of all previously proposed subhaplogroups (H1–H11) were present. However, we noticed that subhaplogroup H9 of Loogväli et al. (2004), is actually a subclade of H6, as attested by the control-region mutational motif of sequence S092 reported by Howell et al. (2003). In addition, 18 of our H sequences did not fit in any of the known subhaplogroups, but 7 harbored control-region and/or coding-region mutations already seen in H mtDNAs from published and unpublished data sets, thus suggesting that those mutations might characterize additional relatively common clades. To these clades, we assigned the following novel subhaplogroup names: H9 (3591-4310-13020-16168), thus replacing the H9 proposed by Loogväli et al. (2004); H12 (3936-14552-16287); H13 (2259-4745-13680-14872); H14 (7645-11377); and H15 (55-57-6253). For the time being, taking into account the possibility that their mutational motifs could be very uncommon, no names were assigned to the remaining 11 H sequences shown in figure 1.

Among the 15 defined subhaplogroups of H, 2 appeared by far the most frequently in our sample; these were H1 and H3, encompassing 12 and 10 mtDNAs, respectively. This suggested that, if the high incidence of H1 and H3 was a real feature of haplogroup H and was not restricted to our selected H sample, a detailed phy-

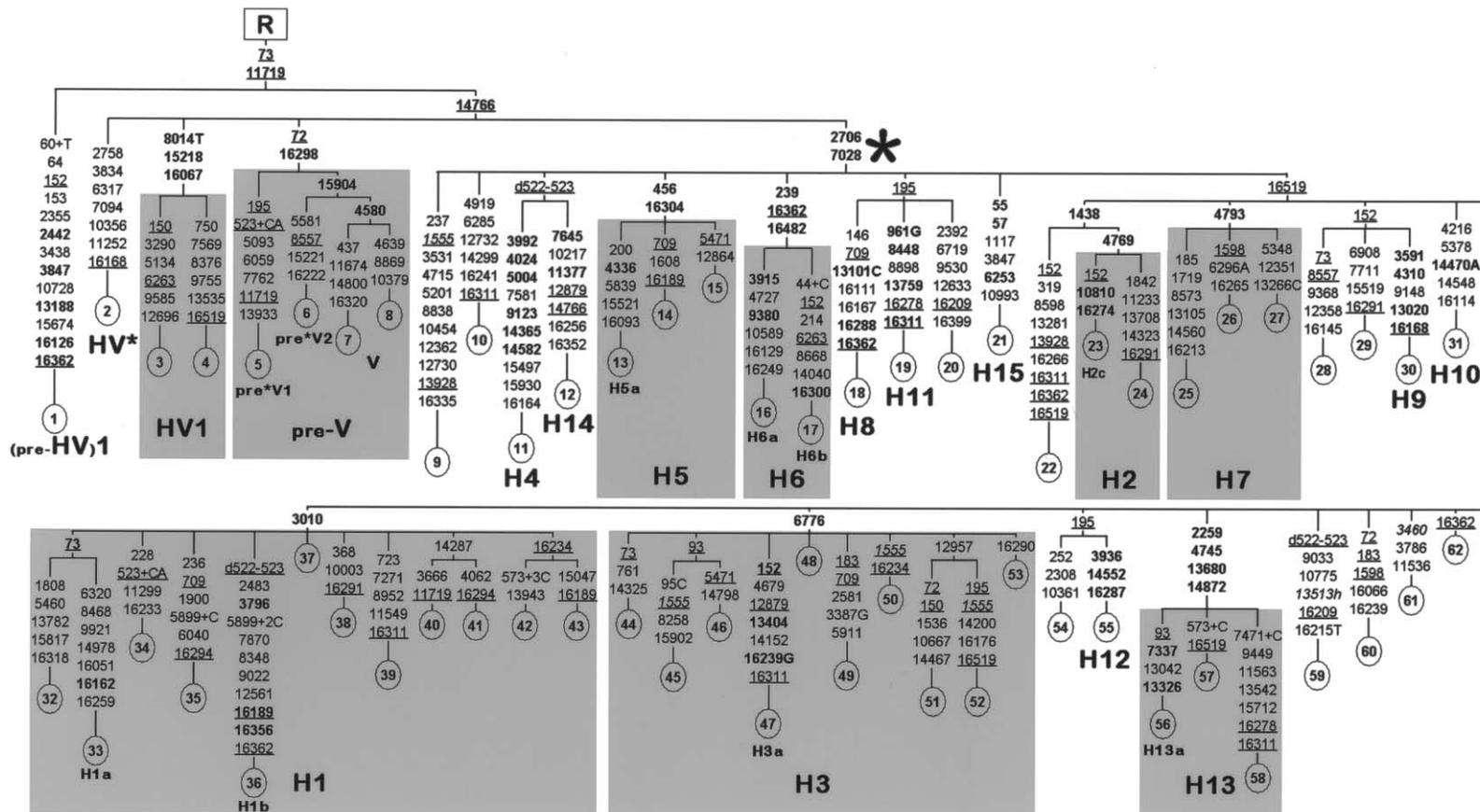


Figure 1 Most-parsimonious tree of complete (pre-HV)1, HV*, HV1, pre-V, and H mtDNA sequences. The tree, rooted in haplogroup R, includes 62 mtDNAs (1–62) sequenced in this study and illustrates subhaplogroup affiliations. Phylogeny construction was performed by hand, following a parsimony approach, and was confirmed by use of the program Network 4.0. We have applied the reduced median algorithm ($r = 2$) (Bandelt et al. 1995), followed by the median-joining algorithm ($\epsilon = 0$) (Bandelt et al. 1999) and the MP (maximum parsimony) calculation option, as explained at the Fluxus Engineering Web site. For the phylogeny construction, half weight was assigned to the control-region positions, and the pathological mutations 1555, 3460, and 13513 (shown in italics) were excluded. Mutations are shown on the branches; they are transitions, unless the base change is explicitly indicated. Deletions are indicated by a “d” preceding the deleted nucleotides. Insertions are indicated by a “+” preceding the inserted nucleotide(s). Heteroplasmy is indicated by an “h” following the nucleotide position. Underlining indicates recurrent mutations, whereas mutations in boldface are diagnostic of the haplogroup/subhaplogroup. The asterisk (*) indicates the most recent common ancestor of the H mtDNAs. This differs from the reference sequence (Andrews et al. 1999), which belongs to H2, by mutations at the following positions: 263, 315+C, 750, 1438, 4769, 8860, and 15326. The variation in number of Cs at np 309 was not included in the phylogeny (mtDNAs 1, 4, 8, 12, 15, 24, 26, 29, 31–32, 34–35, 37, 39, 41, 45–48, 56, and 58–62 harbored 309+C, whereas the mtDNAs 2–3, 5–7, 14, 17–18, 20, 42, 53, and 57 harbored 309+CC). mtDNAs 2, 4–8, 10–21, 23, 25–31, 33, 35–37, 39–44, 46–47, 49, 51, and 54–62 are from Italian subjects; 9, 24, 45, 48, 50, and 52 are from Spanish subjects; 32 and 53 are from Georgian subjects; 22 and 38 are from Iraqi subjects; 1 is from a Sindhi (Pakistan); 3 is from a Druze subject; and 34 is from a Berber of Egypt.

Table 1
Population Distribution and Frequencies of Haplogroup H, H1, and H3 mtDNAs

REGION, ID NUMBER, AND POPULATION	NO. OF SUBJECTS	H FREQUENCY (%)	SUBHAPLOGROUP FREQUENCY (%)	
			H1	H3
Africa:				
1. Senegal	100
2. Berbers (Morocco)	125	36.8	19.2	1.6
3. Algeria	82	25.6	9.8	...
4. Tunisia	83	26.5	9.6	1.2
5. Berbers (Egypt)	71	1.4	1.4	...
Europe:				
6. Andalusia	103	43.7	24.3	4.9
7. Spain (miscellaneous)	132	39.4	18.9	3.8
8. Galicia ^a	266	45.1	17.7	8.3
9. Pasiegos (Cantabria)	51	35.3	23.5	2.0
10. Basques (Spain)	108	51.9	27.8	13.9
11. Basques (France)	40	40.0	17.5	5.0
12. Béarnaise	27	25.9	14.8	7.4
13. France ^b	106	47.2	12.3	5.7
14. Netherlands	34	38.2	8.8	2.9
15. Austria ^c	277	44.8	14.4	2.2
16. Italy (north)	322	46.9	11.5	5.0
17. Italy (center)	208	35.6	6.3	3.8
18. Sardinia	106	42.5	17.9	8.5
19. Italy (south)	206	37.9	8.7	2.4
20. Sicily	90	48.9	10.0	2.2
21. Greece (mainland)	79	41.8	6.3	1.3
22. Greece (Aegean islands)	247	44.1	1.6	.4
23. Macedonia (Northern Greece)	52	40.4	7.7	...
24. Albania	105	48.6	2.9	...
25. Croatia	84	44.0	8.3	6.0
26. Hungary	130	42.3	12.3	6.2
27. Slovaks ^b	119	42.0	7.6	.8
28. Czech Republic	102	41.2	10.8	2.0
29. Poland	86	37.2	9.3	3.5
30. Ukraine ^d	191	40.8	9.9	2.1
31. Russia ^b	312	40.1	13.5	1.6
32. Estonia ^b	114	43.9	16.7	2.6
33. Saami	57	5.3
34. Volga-Ural Finnic speakers ^b	125	40.0	13.6	...
Caucasus:				
35. Caucasus (north)	68	27.9	8.8	...
36. Caucasus (south)	132	21.2	2.3	.8
Middle East:				
37. Turkey ^e	242	26.0	5.0	...
38. Druze	58	17.2	3.4	...
39. Iraq	206	19.9	1.9	...
40. Arabian Peninsula	94	10.6
Asia:				
41. Pakistan	100	12.0
42. Central Asia ^b	445	11.2	.7	.2
43. Yakutia	58	8.6	1.7	1.7

^a From Quintáns et al. (2004).
^b This data set is from Loogväli et al. (2004).
^c From Brandstätter et al. (2003).
^d Includes 100 mtDNAs from Loogväli et al. (2004).
^e Includes 192 mtDNAs from Loogväli et al. (2004).

logeographic analysis focused on these two subhaplogroups could be particularly informative in revealing spatial patterns.

To evaluate this possibility, we performed a detailed molecular survey of all the H mtDNAs found in 5,743 subjects (who had signed appropriate informed con-

sents) from 43 populations of Europe, North Africa, the Middle East, the Caucasus, and Central Asia (table 1 and fig. 2). The H mtDNAs were screened for the presence of the transitions G3010A and T6776C, which mark haplogroups H1 and H3, respectively. The 3010 mutation was detected as a *TaqI* site loss at np 3008 by

use of the mismatched primer 2988FOR (5'-cgatgttgatcaggacatctc), whereas the 6776 mutation was identified as an *Nla*III site gain at np 6773 by use of the mismatched primer 6807REV (5'-gtgtgtctacgtctattctactgtaaaca).

The results of this survey are reported in table 1 and are illustrated in the spatial distribution of figure 3. Subhaplogroup H1 turned out to encompass a large proportion of H in the western part of its distribution range. It has a frequency peak among the Basques of Spain (27.8%) and very high frequencies in the rest of Iberia (17.7%–24.3%), Morocco (19.2%), and Sardinia (17.9%). The spatial pattern depicted in figure 3 appears to indicate the presence of an overall gradient for H1, with a peak centered at the most southwestern edge of Europe and in Morocco and declining frequencies towards both the northeast and southeast.

Compared to H1, subhaplogroup H3 represents a much smaller fraction of H (table 1). However, its highest frequencies are found among the Basques of Spain (13.9%), in Galicia (8.3%), and, again, in Sardinia (8.5%)—in other words, in the same areas where H1 is also most frequent.

The frequency decline of both H1 and H3 from their peaks centered in southwestern Europe is not completely uniform, but a few intermediate local peaks are also observed. Both Austria and Estonia harbor peaks for haplogroup H1 (14.4% and 16.7%, respectively), whereas a local maximum of H3 is observed in Hungary (6.2%). Some intermediate peaks are indeed expected, as a result of random genetic drift. However, in some instances, these could also indicate a more direct genetic link of the populations living in these areas with those of southwestern Europe than with their current surrounding neighbors.

Thus, although the frequency distribution of haplogroup H overall in Europe is rather uniform (fig. 2), those of H1 and H3 harbor clear-cut patterns, with peaks both centered in Iberia and surrounding areas. We noted with great interest that such frequency patterns are extremely similar to that previously described for 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 (Torroni et al. 1998, 2001a; Richards 2003). The distribution of haplogroup V was attributed to a major Paleolithic/Mesolithic population expansion from southwestern Europe, which occurred 13–10 kya and eventually carried those mtDNAs into Central and Northern Europe following the postglacial improvement of the climate conditions.

To determine if the distributions of H1 and H3 could be attributed to the same phenomenon, we estimated the coalescence ages of the two subhaplogroups from the sequence data reported in figure 1. To obtain the esti-

mates, only the coding-region data were used, according to Mishmar et al. (2003). When all 54 H sequences were included, the coalescence estimate for the entire haplogroup H was 18.4 ± 2.0 kya (table 2); a value that, taking into account that the large majority of the H mtDNAs we sequenced are from Western Europe, is in good agreement with those proposed in the past (Torroni et al. 1998; Richards et al. 2000; Loogväli et al. 2004). As expected, when only mtDNAs belonging to H1 and H3 were included, younger coalescence times were obtained. They were 12.8 ± 2.4 kya for H1, and 10.3 ± 2.4 kya for H3. These coalescence ages are very similar, and they become even more similar (10.8 ± 1.1 kya and 11.0 ± 1.4 kya, respectively) when the estimates are obtained by inclusion of previously published H1 and H3 sequences (table 2). Furthermore, they overlap to the coalescence time (11.2 ± 2.7 kya) estimated for haplogroup V from control-region data (Torroni et al. 2001a) and the coalescence time (12.4 ± 2.5 kya) that can now be estimated from the pool of the 66 available coding-region sequences belonging to V (table 2). Then, the assumption of a common origin and spread for H1, H3, and V would give an averaged age of 11.3 ± 0.9 kya. This estimate would be consistent with an origin of these haplogroups, say, in the terminal Pleistocene (16–11.5 kya), with major expansion in the early Holocene (perhaps ~10 kya, when vegetation stabilized) (Roberts 1998).

The finding that H1 and H3 show a nonuniform geographic distribution led us to reanalyze in greater detail the mtDNAs of 50 Spanish families previously reported by Torroni et al. (1999). These families were affected by nonsyndromic sensorineural deafness due to the mtDNA A1555G mutation in the 12S rRNA gene (MIM 561000). That study revealed an excess of affected families harboring haplogroup H (38 of 50; $P = 1.3 \times 10^{-5}$). However, the A1555G mutation was the result of 30 independent mutational events among the 50 families, and the distribution of the A1555G mutational events was still compatible with a random occurrence of the mutation on different haplogroups, supporting the conclusion that mtDNA backgrounds do not play a significant role in the expression of the A1555G mutation (Torroni et al. 1999).

Having determined that H1 and H3 are two common subhaplogroups of H in Iberia, we surveyed the 50 mtDNAs from the deafness families for the presence of the H1 (G3010A) and H3 (T6776C) diagnostic markers (table 3). We found that the frequency of H1 in patients (22.0%) was virtually identical to that observed among Spanish controls (21.1%). In contrast, 15 of the affected families harbored an H3 mtDNA, corresponding to an incidence of 30.0%; a value significantly higher ($P = 6.5 \times 10^{-6}$) than that observed in the general Spanish population (7.3%). Essentially, the excess of haplogroup

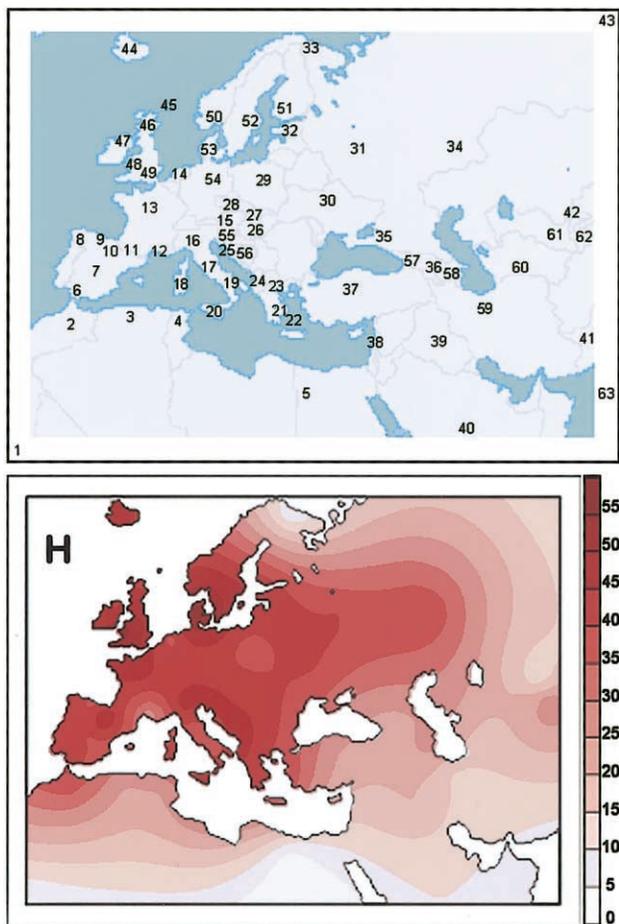


Figure 2 Geographical locations of populations surveyed for haplogroup H (top) and its spatial frequency distribution (bottom). Frequency values for populations 1–43 are from table 1, whereas those for populations 44–63 are from the literature, as follows: 44–46 from Helgason et al. 2001; 47–49 and 53 from Richards et al. 2000; 50 from Richards et al. 2000 and Passarino et al. 2002; 51 from Finnilä et al. 2001; 52 from Torroni et al. 1996 and Richards et al. 2000; 54 from Baasner et al. 1998, Lutz et al. 1998, Pfeiffer et al. 1999 and 2001, and Richards et al. 2000; 55–56 from Malyarchuk et al. 2003; and 57–63 from Quintana-Murci et al. 2004. The frequency map was obtained using Surfer version 6.04 (Golden Software), with the Kriging procedure, and estimates at each grid node were inferred by consideration of the entire data set.

H among the affected subjects is caused entirely by a marked excess of H3. Three possible explanations can be envisioned for such an observation, taking into account that population substructure could be excluded because of the rather heterogeneous geographic origin of the affected Spanish families (Torroni et al. 1999). First, subhaplogroup H3 might increase the penetrance and/or the expressivity of the A1555G mutation, thus increasing its detection rate in subjects with an H3 mtDNA. Second, one A1555G mutation occurred on a specific H3 mtDNA, and this was later transmitted by

descent through numerous generations to families that show the disease phenotype, but their members are not aware of the fact that they are maternally related to each other—a classical example of a founder event. Finally, the mtDNA background which may be modulating the expression of A1555G is not the entire subhaplogroup H3 but only a specific subset within it—for instance, a specific haplotype.

To evaluate the three alternative scenarios, we completely sequenced and included in the haplogroup H phylogeny (fig. 1) three H3 mtDNAs (45, 50, and 52) harboring the A1555G mutation. The three sequences were very divergent from each other, and the only mutation that they shared was T6776C, the diagnostic marker of H3. This transition is in the cytochrome c oxidase subunit I gene, but it is synonymous and does not affect the corresponding histidine. Thus, this finding appears to exclude the first scenario, that subhaplogroup H3 might play a role in the expression of the A1555G mutation. We then checked whether there was some evidence for a founder event among the 15 H3 mtDNAs with the A1555G mutation. Nine of them harbored the *RsaI* site at np 8255 (T8258C) and the control-region motif 16519-93-95C (Torroni et al. 1999), which are observed also in sequence 45 in figure 1. This motif has never been reported in H mtDNAs from Spanish controls (or, in fact, from any other European population), thus strongly supporting the validity of the second scenario,

Table 2

Age Estimates of Haplogroups H, H1, H3, and V

Haplogroup and Source of Data	No. of mtDNAs	ρ^a	σ^b	$T \pm \Delta T$ (kya) ^c
H:				
This study	54	3.574	.395	18.4 ± 2.0
H1:				
This study	12	2.500	.464	12.8 ± 2.4
Total ^d	134	2.112	.219	10.8 ± 1.1
H3:				
This study	10	2.000	.458	10.3 ± 2.4
Total ^d	50	2.140	.279	11.0 ± 1.4
V:				
Total ^d	66	2.409	.490	12.4 ± 2.5

^a The average number of base substitutions in the mtDNA coding region (between nps 577 and 16023) from the ancestral sequence type.

^b Standard error calculated from an estimate of the genealogy, in the manner of Saillard et al. (2000).

^c Estimate of the time to the most recent common ancestor of each cluster, using an evolutionary rate estimate of $1.26 \pm 0.08 \times 10^{-8}$ base substitutions per nucleotide per year in the coding region (Mishmar et al. 2003), corresponding to 5,140 years per substitution in the whole coding region.

^d The sequences from this study plus the coding-region sequences from the studies by Ingman et al. (2000), Finnilä et al. (2001), Herrnstadt et al. (2002, 2003), Mishmar et al. (2003), and Coble et al. (2004).

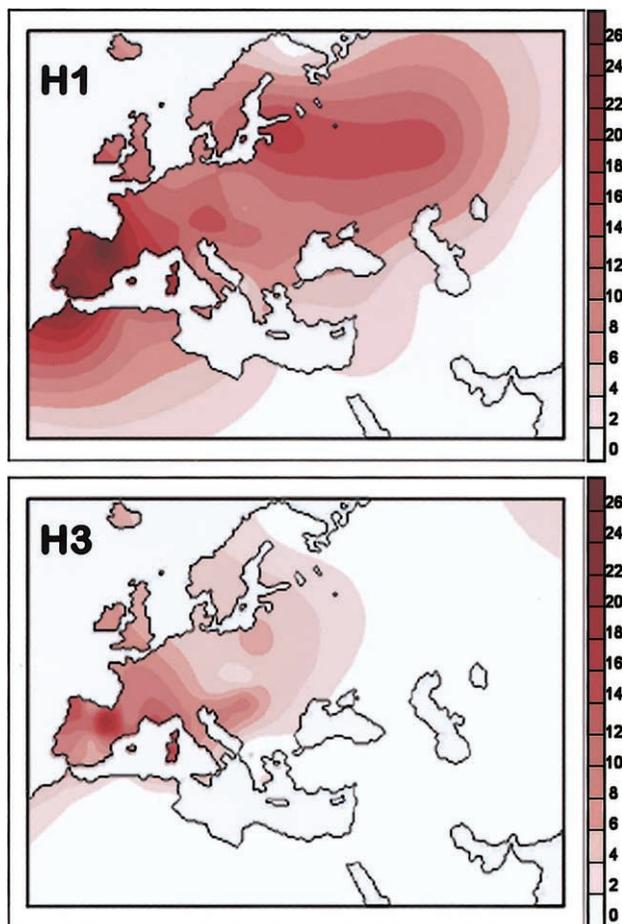


Figure 3 Spatial frequency distributions of subhaplogroups H1 and H3. Frequency values are from table 1. Maps were obtained as in figure 2.

an important founder event involving a rare H3 haplotype. Furthermore, sequence 45 shows a homoplasmic transition (A15902G) in the tRNA^{Thr} gene that we surveyed in the other eight mtDNAs with the motif 16519-93-95C; we found that none of them harbored it. This observation suggests that the founder event may not be recent, concordant with the apparent unrelatedness of the affected families.

Finally, to determine whether the haplotype shared by the nine mtDNAs could, by itself, play a role in the expression of the A1555G mutation, we analyzed in detail its shared mutations: A93G, A95C, and T8258C. The first two (93 and 95C) are located in the control region, and their association is sporadically seen in other Eurasian haplogroups but very frequently in some African haplogroups—for example, in L0a, L1c, and L2c (Alves-Silva et al. 2000; Ingman et al. 2000; Mishmar et al. 2003)—whereas the 8258 mutation causes a phenylalanine-to-leucine amino acid change in a noncon-

served position of the cytochrome c oxidase subunit II gene. Most importantly, none of them is located in the 12S rRNA gene—the gene affected by the A1555G mutation. Thus, although a role of this rare haplotype in the modulation of the A1555G expression cannot be completely ruled out, both the location and the features of the mutational motif indicate that such a possibility is extremely unlikely.

In conclusion, our analysis of complete mtDNA sequences reveals that haplogroup H, the most common haplogroup in western Eurasia, can be subdivided into numerous sister clades. Among these, two—H1 and H3—were particularly common in our sample of H sequences, suggesting that a phylogeographic study focusing on the two subhaplogroups could be particularly informative. Indeed, the survey of a wide range of western Eurasian and North African populations revealed that, in contrast to haplogroup H as a whole, which harbors a rather uniform frequency within Europe, both subhaplogroups H1 and H3 are characterized by frequency peaks centered in Iberia and surrounding areas and by declining distributions toward the northeast and southeast. This pattern is extremely similar to that previously reported for mtDNA haplogroup V. However, not only the frequency distributions of H1, H3, and V resemble each other; also, the coalescence ages of H1 and H3 are close to that of V. Thus, it is now clear that the scenario proposed to explain the age and distribution of haplogroup V can be directly transposed to subhaplogroups H1 and H3. This suggests that the Franco-Cantabrian refuge area was indeed the source of late-glacial expansions of hunter-gatherers that repopulated much of Central and Northern Europe from ~15 kya. This picture, now supported by three of the clades of the mtDNA phylogeny, is also in perfect agreement with the synthetic map of the second principal component of variation in 95 classical genetic markers (Cavalli-Sforza et al. 1994), and the distributions of the Y-chromosome haplogroups R1b and I1b2 (Semino et al. 2000; Rootsi et al. 2004).

Table 3

Frequency of H, H1, and H3 among Spanish Subjects Harboring the Deafness mtDNA Mutation A1555G

POPULATION	NO. OF SUBJECTS	HAPLOGROUP FREQUENCY (%)		
		H ^a	H1 ^b	H3 ^c
Deaf Spanish subjects	50	76.0	22.0	30.0
Spanish controls ^d	660	44.1	21.1	7.3

^a $P = 1.3 \times 10^{-5}$ (Fisher's exact test, two-tailed).

^b $P = .858$ (Fisher's exact test, two-tailed).

^c $P = 6.5 \times 10^{-6}$ (Fisher's exact test, two-tailed).

^d Populations 6, 7, 8, 9, and 10 from table 1.

Finally, this study confirms that the molecular dissection of major mtDNA haplogroups into clades of younger ages and more restricted geographic distributions can be extremely informative. Applied extensively at the highest level of resolution—that of complete mtDNA sequences—it is likely to reveal spatial patterns attributable not only to primary colonization events and late-glacial expansions from ice-age refugia but also to Neolithic dispersals and more recent events of gene flow. Such spatial patterns, as is attested by our survey of mtDNAs harboring the deafness A1555G mutation, could also have important implications for disease studies.

Acknowledgments

We are grateful to all the donors, for providing blood samples, and to the people who contributed to their collection. In particular, we thank Ignacio del Castillo, Felipe Moreno, and Xavier Estivill (for the Spanish samples harboring the A1555G mutation); Anne Cambon-Thomsen (for the French Basque and Bearnais samples); Mauricio DeGrado (for the Pasiego samples); Marc Fellous (for the samples from Algeria and Tunisia); M. Mohamed Melhaoui, M. Abdellatif Baali, and M. Mohamed Cherkaoui (for the Berber samples from Morocco); Farha El Chennawi (for the Berber samples from Egypt); Costas Thriantafillidis (for the samples from Northern Greece); and Dragan Primorac (for the samples from Croatia). This research received support from Progetto CNR-MIUR Genomica Funzionale-Legge 449/97 (to A.T.), the Italian Ministry of the University (Progetti Ricerca Interesse Nazionale 2002, 2003) (to A.T. and R.S.), Grandi Progetti di Ateneo (to R.S.), the Istituto Pasteur Fondazione Cenci Bolognetti (to R.S.), Fondo Investimenti Ricerca di Base 2001 (to A.T.), Estonian Science Foundation grant 5574 (to T.K.), European Commission grants ICA1CT20070006 and QLG2-CT-2002-90455 (to R.V.), Fondo d'Ateneo per la Ricerca dell'Università di Pavia (to A.T. and A.S.S-B), the Progetto Finalizzato CNR "Beni Culturali" (to A.S.S-B), and Fondo Giovani Ricercatori dell'Università di Pavia (to A.A.). The sampling of the Berbers from Morocco and Egypt was made within the framework of the action OMLL (The Origin of Man, Language, and Languages) (EUROCORES Programme) and benefited from funding of the CNRS (Centre National de la Recherche Scientifique) and the Conseil Régional de Midi-Pyrénées, Toulouse (France).

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Authors' Web site, http://ipvgen.unipv.it/docs/projects/torroni_data/torroni_sequences.html (for the complete mtDNA sequences)

Fluxus Engineering, <http://www.fluxus-engineering.com> (for NETWORK 4.0)

GenBank Overview, <http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html> (for the complete mtDNA sequences [accession numbers AY738940–AY739001])

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/>

References

- Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444–461 (erratum 67:775)
- Andrews RM, Kubacka I, Chinnery PE, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
- Baasner A, Schäfer C, Junge A, Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. *Forensic Sci Int* 98:169–178
- Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753
- Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117:291–298
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ
- Cinnioğlu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 114:127–148
- Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *Int J Legal Med* 118:137–146
- Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. *Am J Hum Genet* 68:1475–1484
- Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, Sykes B (2001) mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 68:723–737
- Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal FM, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70:1152–1171 (erratum 71:448–449)
- Herrnstadt C, Preston G, Howell N (2003) Errors, phantom and otherwise, in human mtDNA sequences. *Am J Hum Genet* 72:1585–1586
- Housley RA, Gamble CS, Street M, Pettitt P (1997) Radiocarbon evidence for the lateglacial human recolonisation of northern Europe. *Proc Prehist Soc* 63:25–54
- Howell N, Oostra RJ, Bolhuis PA, Spruijt L, Clarke LA, Mackey DA, Preston G, Herrnstadt C (2003) Sequence anal-

- ysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy. *Am J Hum Genet* 72:1460–1469
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713
- Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, et al (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* (in press)
- Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111:67–77 (errata 111:286 and 112:145–150)
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232–249
- Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Drobnic K, Miścicka-Śliwka D (2003) Mitochondrial DNA variability in Bosnians and Slovenians. *Ann Hum Genet* 67:412–425
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171–176
- Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL, Underhill PA (2002) Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *Eur J Hum Genet* 10:521–529
- Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112:291–298
- Pfeiffer H, Forster P, Ortmann C, Brinkmann B (2001) The results of an mtDNA study of 1,200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. *Int J Legal Med* 114:169–172
- Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Qasim Mehdi S, Torroni A, McElreavey K (2004) Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* 74:827–845
- Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 140:251–257
- Richards M (2003) The Neolithic invasion of Europe. *Annu Rev Anthropol* 32:135–162
- Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67:1251–1276
- Richards M, Macaulay V, Torroni A, Bandelt H-J (2002) In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet* 71:1168–1174
- Roberts N (1998) *The Holocene—an environmental history*. 2nd ed. Blackwell, Oxford
- Roots S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, et al (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75:128–137
- Saillard J, Forster P, Lynnerup N, Bandelt H-J, Nørby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718–726
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159
- Tambets K, Roots S, Kivisild T, Help H, Serk P, Loogväli EL, Tolk HV, et al (2004) The western and eastern roots of the Saami—the story of genetic “outliers” told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* 74:661–682
- Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850
- Torroni A, Bandelt H-J, D’Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonnè-Tamir B, Scozzari R (1998) MtDNA analysis reveals a major late Palaeolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137–1152
- Torroni A, Cruciani F, Rengo C, Sellitto D, Lopez-Bigas N, Rabionet R, Govea N, Lopez De Munain A, Sarduy M, Romero L, Villamar M, del Castillo I, Moreno F, Estivill X, Scozzari R (1999) The A1555G mutation in the 12S rRNA gene of human mtDNA: recurrent origins and founder events in families affected by sensorineural deafness. *Am J Hum Genet* 65:1349–1358
- Torroni A, Bandelt H-J, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, et al (2001a) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69:844–852
- Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001b) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69:1348–1356