Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia

Éva-Liis Loogvåi,∗ Urmas Roostalu,∗ Boris A. Malyarchuk,† Miroslava V. Derenko,†
Toomas Kvistild,∗ Ene Metspalu,∗ Kristiina Tambets,∗ Maere Reidda,∗ Helle-Viví Tolkk,∗
Jüri Parik,∗ Erwan Pennarun,∗ Sirle Laos,* Arina Lunkina,† Maria Golubenko,∗†
Lovorka Barač,§§ Marijana Peričić,§§ Oleg P. Balanovsky,*||
Vladislava Gusar,∗©∗© Elsa K. Khusnutdinova,# Vadim Stepanov,‡ Valery Puzyrev,‡
Pavlo Rudan,# Elena V. Balanovska,∥ Elena Grechanka,∥∥ Christelle Richard,∗∗
Jean-Paul Moisan,∗∗ André Chaventré,∗ Nicholas P. Anagnou,†† Kalliopi I. Pappa,††
Emmanuel N. Michalodimitrakis,†† Mireille Claustrœ,§§ Mukaddes Gölge,||||
Ilia Mikerezi,¶¶ Esien Usanga,## and Richard Villems∗*

*Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocenter, Tartu, Estonia; †Genetics Laboratory, Institute of Biological Problems of the North, Russian Academy of Sciences, Magadan, Russia; ‡Institute of Medical Genetics, Tomsk Research Center, Russian Academy of Medical Sciences, Tomsk, Russia; §Institute for Anthropological Research, Zagreb, Croatia; ||Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia; ¶Kharkov Clinical Genetic and Prenatal Diagnostics Center, Kharkov, Ukraine; #Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa, Russia; ∥∥Laboratoire d’Etude du Polymorphisme de l’ADN, Faculté de Médecine, Nantes, France; ††Institute of Molecular Biology and Biotechnology and Department of Basic Sciences, University of Athens School of Medicine, Athens, Greece; ‡‡Department of Forensic Sciences and Toxicology, University of Crete School of Medicine, Heraklion, Greece; §§Laboratoire de Génétique Moléculaire, Institut Universitaire de Recherche Clinique IURC et CHU, Montpellier, France; ###Department of Physiology, University of Kiel, Kiel, Germany; ¶¶Department of Biology, Faculty of Natural Sciences, Tirana University, Tirana, Albania; and ##Department of Medical Laboratory Sciences, Kuwait University, Sulaibikhat, Kuwait

It has been often stated that the overall pattern of human maternal lineages in Europe is largely uniform. Yet this uniformity may also result from an insufficient depth and width of the phylogenetic analysis, in particular of the predominant western Eurasian haplogroup (Hg) H that comprises nearly a half of the European mitochondrial DNA (mtDNA) pool. Making use of the coding sequence information from 267 mtDNA Hg H sequences, we have analyzed 830 mtDNA genomes, from 11 European, Near and Middle Eastern, Central Asian, and Altaiian populations. In addition to the seven previously specified subhaplogroups, we define fifteen novel subclades of Hg H present in the extant human populations of western Eurasia. The refinement of the phylogenetic resolution has allowed us to resolve a large number of homoplasies in phylogenetic trees of Hg H based on the first hypervariable segment (HVS-I) of mtDNA. As many as 50 out of 125 polymorphic positions in HVS-I were found to be mutated in more than one subcluster of Hg H. The phylogeographic analysis revealed that sub-Hgs H1*, H1b, H1f, H2a, H3, H6a, H6b, and H8 demonstrate distinct phylogeographic patterns. The monophyletic subhaplogroups of Hg H provide means for further progress in the understanding of the (pre)historic movements of women in Eurasia and for the understanding of the present-day genetic diversity of western Eurasians in general.

Introduction

The mitochondrial DNA (mtDNA) sequences of Europeans are sorted into ten major phylogenetic clades, or haplogroups, alphabetically named H, J, K, N1, T, U4, U5, V, X, and W (Torroni et al. 1994, 1996; Macaulay et al. 1999; Richards et al. 2000). Haplogroup (Hg) H alone constitutes about one half of the European mtDNA pool and, along with other aforementioned lineages, is widespread also in western Asia (Macaulay et al. 1999; Richards et al. 2000; Tambets et al. 2000; Kvistild et al. 2003); Central Asia (Comas et al. 1998; Metspalu et al. 1999), Siberia (Saillard et al. 2000a; Derbeneva et al. 2002a; Derenko et al. 2003), southern Asia (Passarino et al. 1996; Kivisild et al. 1999, 2003; Bamshad et al. 2001), and northern Africa (Corte-Real et al. 1996; Rando et al. 1998; Stevanovitch et al. 2004; fig. 1A). At least 267 Hg H mtDNA genomes have been sequenced (nearly) completely (Reid, Vernham, and Jacobs 1994; Rieder et al. 1998; Levin, Cheng, and Reeder 1999; Ingman et al. 2000; Finnila, Lehtonen, and Majamaa 2001; Maca-Meyer et al. 2001; Herrnstadt et al. 2002, correction by Herrnstadt, Preston, and Howell 2003; Mishmar et al. 2003). Out of seven Hg H sub-Hgs defined so far, Hgs H1 and H2 (Finnila, Lehtonen, and Majamaa 2001) along with Hgs H3 and H4 (Herrnstadt et al. 2002) and Hgs H5, H6, and H7 (Quintans et al. 2004) cover 74% of Finnish, 68% of U.S./U.K. and 77% of Galician Hg H sequences, respectively.

Attempts to classify Hg H lineages by first hypervariable segment (HVS-I) have been hindered by a frequent occurrence of mutations at fast-evolving nucleotide sites—so-called mutational hot-spots (Richards et al. 2000; Allard et al. 2002). Furthermore, HVS-I sequence information...
leaves a substantial fraction of Hg H genomes phylogenetically unresolved: on average one-third of the lineages share a haplotype identical with the Cambridge reference sequence (CRS; Anderson et al. 1981). Therefore, progress in the analysis of Hg H diversity, and, indeed, of the understanding of the phylogeography of western Eurasian maternal lineages, depends critically on the use of full genome information in mtDNA samples, representative in size and geography.

Materials and Methods


The samples were selected at random from nine populations; 50 Finno-Ugric speakers from the Volga-Ural region (10 Udmurts, 10 Mokshas, 16 Erzyas, 7 Permyak Komis, 7 Zyrían Komis); 50 Estonians; 165 Eastern Slavs (127 Russians, 38 Ukrainians from various districts of Russia and the Ukraine); 50 Slovaks; 50 French from southern France, Lyon, Low Normandy, and Poitiers; 50 individuals from the Balkans (17 Croats, 17 Albanians, 16 Greeks); 50 Turks; 50 individuals from the Near and the Middle East (10 Jordanians, 8 Lebanese, 7 Saudis, 12 Syrians, 13 Iranians); 48 individuals from Central Asia (17 Altaïans, 11 Kirghiz, 3 Kazakhs, 11 Tajiks, 6 Uzbekks). Sixteen Russian and six Ukrainian HVS-I sequences have been published by Malyarchuk and Derenko (2001a), 33 Russian HVS-I and second hypervariable segment (HVS-II) sequences by Malyarchuk et al. (2002), and all of the Volga-Ural region mtDNA HVS-I sequences by Bernisheva et al. (2002). All the samples harbored a C at nucleotide position (np) 7028, which is diagnostic for Hg H and was inferred from the absence of the AluI restriction site at np 7025 (Torroni et al. 1994). All mutations and position numbers in this study are given with respect to Anderson et al. (1981) as revised by Andrews et al. (1999).

Four hundred forty-eight samples were screened for 14 polymorphisms in the mtDNA coding region and three in HVS-II in addition to HVS-I sequence variation. A hierarchical strategy was applied to 104 Russian and 11 Ukrainian mtDNAs (Appendix S2 in the Supplementary Material online). HVS-I variation for all of the samples was scored between np 16024–16383. Nucleotide changes at positions 73, 951, 3010, 4336, 4452, 4769, 4793, 5004, 8448, 9066, 9380, 13101, 13759, and 16482 were determined by restriction fragment length polymorphisms (RFLPs; Appendixes S1 and S2). Nucleotide states at positions 239, 456, 3915, and 6776 were detected by direct sequencing or allele-specific polymerase chain reaction (PCR; Appendixes S1 and S2). Nucleotide positions 239 and 3915 were sequenced in samples having 16362C and/or lacking a HincII restriction site at np 9380 and/or having a DdeI site at np 16478. We note that the transition at np 239 nearly always occurs with the 16362C allele, as it was not found in Hg H variants with 16362T in 2,350 published HVS-II sequences (Hofmann et al. 1997; Parson et al. 1998; Dimo-Simonin et al. 2000; Malyarchuk et al. 2003; Vanecek, Vorel, and Sip 2004; Pereira, Cunha, and Amorim 2004). Credible regions of the obtained haplogroup frequencies were computed with the Sampling program kindly provided by Vincent Macaulay.

The phylogeny of the samples was studied by the construction of a reduced median network (fig. 2A). In the network analysis 479 samples were included (see Appendix S1), including the 31 Finnish sequences taken from Finnilä, Lehtonen, and Majamaa (2001), while 115 Eastern Slav mtDNAs, which were analyzed hierarchically (see

![Fig. 1.—Spatial frequency distributions of haplogroup H (A) and its subhaplogroup H1b (B). H1b frequencies are given as percentages with respect to Hg H.](image-url)
Appendix S2), have not been included. The reduced median network (Bandelt et al. 1995; rho set at 2) was constructed with the Network 4.0.0.0. program (Fluxus Technology Ltd., Clare, Suffolk, UK, http://www.fluxus-engineering.com) followed by a median joining algorithm (Bandelt, Forster, and Röhl 1999; epsilon set at 0), as explained at the Fluxus-Engineering Web site. Nucleotide positions were divided into three classes of transition rates—fast (16093, 16129, 16189, 16304, 16311, and 16362), intermediate (16172, 16209, 16278, 16293), and slow (the remaining positions between 16024 and 16383)—and assigned class weights 1, 2, and 4, respectively. Transversions and coding region mutations were weighted 8.

To obtain the frequencies of Hg H and its sub-Hg H1b in different populations, we compiled a data set of 26105 HVS-I sequences from various sources listed in table S2 in the Supplementary Material online. The frequency data in individual populations was grouped into broader geographical regions (see table S2) and summary frequencies obtained were mapped (fig. 1). Maps were obtained using Surfer version 7 (Golden Software, Inc., Golden, Colo.) with the Kriging procedure. Estimates at each grid node were obtained by consideration of the entire data set.

Altogether, 830 mitochondrial genomes were included in the coalescence analysis. A subset of the obtained coalescence estimates are presented in table 1 and all of the results in table S1. An average transitional distance from the root haplotype (rho) was calculated. Coalescence time has been calculated taking one transitional step between nucleotide positions 16090–16365 (“HVS”) equal to 20,180 years (Forster et al. 1996) and one base substitution between nucleotide positions 577–16023 (“coding”) equal to 5,138 years (Mishmar et al. 2003). Standard deviation of the rho estimate (sigma) was calculated as in Saillard et al. (2000b), and SD denotes the deviation in years. The 115 Eastern Slav samples analyzed hierarchically and not shown in figure 2A have been included in the coalescence analysis. Note that the coding sequence data is derived mainly from European populations.

Results and Discussion

Figure 3 shows the backbone of the phylogenetic tree of Hg H subclades studied here. We have corrected the names of sub-Hgs H5 and H6 as defined by Quintans et al. (2004) as 4336C and 3915A, respectively, to H5a and H6a, following the hierarchical principle described by Richards et al. (1998). Note that the most parsimonious phylogenetic tree has two branching events based on shared HVS-I nucleotide transitions: one between sub-Hgs H1b and H1f and the second between sub-Hgs H6 and H8. Because the transitions, at nucleotide positions (nps) 16189 and 16362, involve mutational hot-spots, the indicated sub-Hgs, though monophyletic, are not necessarily sister clades, as depicted in figure 3.

The number of internal branches in Hg H is significantly higher than in other mtDNA haplogroups widespread in Europe. In the majority of European mtDNA variants—J, T, K, X and U5—the coding region variation is described by only a few extant basal subclades (Finnilä et al. 2000; Finnilä and Majamaa 2001; Herrnstadt et al. 2002; Reidla et al. 2003). In contrast, there are 57 basic branches stemming from the founder node of Hg H in the parsimonious phylogenetic tree relating 267 Hg H coding region sequences (fig. S1).

One hundred twenty-five variable positions were detected in 594 (563 + 31 Finnish sequences of Finnilä, Lehtonen, and Majamaa 2001) Hg H HVS-I sequences. Among them, recurrent transitions were observed in 50 positions (40%) in different subclades (table 2). The sites with the highest number of recurrences match the HVS-I hot-spot sites identified previously (Hasegawa et al. 1996; Finnilä et al. 2000; Hasegawa et al. 2002; Reidla et al. 2003).
The most variable positions, 16093 and 16311, had received parallel hits in seven different subclusters; 16189 in six; 16092, 16304, and 16362 each in five; and 16129, 16209, 16249, and 16325 each in four subclusters. Another 12 HVS-I mutations were found in three and 28 substitutions in two different phylogenetic contexts. Because quite a few of these hot-spot mutations are present in HVS-I haplotypes that have been highlighted as having founder status in Europe (Richards et al. 2000), our results document again that additional coding region information is essential and unavoidable in defining monophyletic subclades of Hg H reliably (Torroni et al. 1993; Bandelt et al. 2001; Kivisild et al. 2002). We also found that a reversion of A to the ancestral base G at np 73 of the HVS-II, noticed in Hg H first by Torroni et al. (1996), has occurred independently at least four times in Hg H phylogeny (see also Helgason et al. 2000).

In the coding region, a transition at np 3010 that defines sub-Hg H1 (Finnilä, Lehtonen, and Majamaa 2001) is phylogenetically equally problematic. The derived state at np 3010 has been detected in haplogroups C, D, H, J, L2, L3, and U, making this base pair one of the fastest evolving mtDNA coding region positions (Ingman et al. 2000; Finnilä, Lehtonen, and Majamaa 2001; Maca-Meyer et al. 2001; Torroni et al. 2001b; Herrnstadt et al. 2002). Character conflicts at np 3010 and at more conserved nps 1462 (occurs also in Hgs H*, H2, T), 6272 (H*; L3), 6776 (H3), 8470 (H3), 12172 (H*, L1, U2), and 14869 (H*, K2, L3) were found (fig. S1). Given the data, the number of independent 3010A incidences in Hg H may possibly be as many as four (fig. S1).

Sub-Hg H4 was previously defined by an array of eight mutations (nps 3992, 4024, 5004, 8269, 9123, 10044, 14365, and 14582) through an analysis of haplotypes that occurred in at least two individuals (Herrnstadt et al. 2002). However, re-examination of the sequence data of Herrnstadt et al. (2002) revealed that only six mutations at nps 3992, 4024, 5004, 9123, 14365, and 14582 appear to be necessary to characterize the clade (fig. S1). Consequently, here we name the bough defined by a G-to-A mutation at np 8269, which further embraces the 10044 twig, as H4a.

### Table 2

<table>
<thead>
<tr>
<th>Haplogroup H Subclusters</th>
<th>np(^a)</th>
<th>H1</th>
<th>H2</th>
<th>H3</th>
<th>H4</th>
<th>H5</th>
<th>H6</th>
<th>H7</th>
<th>H8</th>
<th>H11</th>
<th>H*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum(^b)</td>
<td>93</td>
<td>111</td>
<td>189</td>
<td>92</td>
<td>304</td>
<td>362</td>
<td>129</td>
<td>249</td>
<td>325</td>
<td>75</td>
</tr>
<tr>
<td>Haplogroup H Subclusters</td>
<td></td>
<td>169</td>
<td>172</td>
<td>192</td>
<td>218</td>
<td>234</td>
<td>243</td>
<td>261</td>
<td>291</td>
<td>294</td>
<td>301</td>
</tr>
<tr>
<td></td>
<td></td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
<td>354</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51</td>
<td>86</td>
<td>111</td>
<td>114</td>
<td>126</td>
<td>145</td>
<td>148</td>
<td>162</td>
<td>167</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td></td>
<td>188</td>
<td>193</td>
<td>240</td>
<td>256</td>
<td>259</td>
<td>265</td>
<td>266</td>
<td>270</td>
<td>271</td>
<td>274</td>
</tr>
<tr>
<td></td>
<td></td>
<td>278</td>
<td>288</td>
<td>290</td>
<td>293</td>
<td>296</td>
<td>299</td>
<td>300</td>
<td>344</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) HVS-I nucleotide position (minus 16000).

\(^b\) Total number of subhaplogroups where the particular transition has occurred. Analysis is based on mtDNA positions 16024-16365 in 594 haplogroup H mitochondrial genomes. Note that 115 Eastern Slav samples that are not shown in figure 2A are included here.

While applying the RFLP method we discovered three previously unknown mutations: a transition at np 13760 abolishing the AcI site at np 13757 defining sub-Hg H11, a transition at np 5005 eliminating the H4-defining Ddel site at np 5003, and a transition at np 8449 eliminating the H11-defining np 8446 SspI site. Therefore, we confirmed the presence of H4-specific T at np 5004 by...
the frequency of H1 does not exceed 6% (Pereira et al. 2004; Quintans et al. 2004). In the Near East Peninsula, covering about 46% of local Hg H lineages European mtDNA pool. H1 is most frequent in the Iberian graphic results. The largest subcluster is sub-Hg H1, which methods, like RFLP, should be backed-up by direct that classical indirect DNA polymorphism detection characteristic HVS-I mutation pattern. These results show established by the combination of two RFLPs and by the 5003 site. The monophyly of sub-Hg H11 is well tion for details of the studied samples.

H1a
H1b
H1c
H1f
H1i
H2a
H3
H4
H5
H5a
H6
H7
H8
H9
H11
H sample size
H frequency (%)
Total sample size

<table>
<thead>
<tr>
<th>Subclusters of Haplogroup H</th>
<th>VUF</th>
<th>Fin</th>
<th>Est</th>
<th>ESlav</th>
<th>Slk</th>
<th>Fre</th>
<th>Blk</th>
<th>Tur</th>
<th>NE</th>
<th>Asia</th>
<th>Her</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1a</td>
<td>14</td>
<td>2</td>
<td>10</td>
<td>8 + 28</td>
<td>3</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>69</td>
<td>165</td>
</tr>
<tr>
<td>H1b</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4 + 3</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>nd</td>
<td>24</td>
</tr>
<tr>
<td>H1c</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2 + 7</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>H1f</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>1 + 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>H2a*</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1 + 4</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>H2b</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>9 + 5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>H3</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3 + 4</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>25</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>H4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0 + 3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>H5a</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2 + 1</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>nd</td>
<td>19</td>
</tr>
<tr>
<td>H5b</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1 + 6</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>31</td>
</tr>
<tr>
<td>H6</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1 + 9</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>H7</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0 + 2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>H8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>H11</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2 + 8</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

Haplogroup H makes up 49% of the British population (Piercy et al. 1993; Richards et al. 1996; Helgason et al. 2001). The same percentage was used to estimate the frequency of haplogroup H in Western Europe, represented here by the data of Herrnstadt et al. (2002).

Table 3

Frequencies of Haplogroup H Subhaplogroups in Studied Populations

Note.—nd: no data.

Abbreviations for the populations are as follows: VUF: Volga-Ural region Finno-Ugric speakers; Fin: Finnish sequences are taken from Finnila¨, Lehtonen, and Majamaa (2001); Est: Estonians; ESlav: Eastern Slavs; Slk: Slovaks; Fre: French; Blk: Balkan peoples; Tur: Turks; NE: Near and Middle Easterners; Asia: Asian peo-ple; Her: Herrnstadt et al. (2002) coding mtDNA sequences represent the populations of the United Kingdom and United States. See the Materials and Methods sec-tion for details of the studied samples.

Subcluster H2* includes all members of H2 that do not belong to H2a.

Subcluster H2† includes all members of H2 that do not belong to H2a.

Subcluster H2‡ includes all members of H2 that do not belong to H2a.

Subcluster H2§ includes all members of H2 that do not belong to H2a.

Subcluster H2¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.

Subcluster H2‖¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶¶ includes all members of H2 that do not belong to H2a.
Peninsula, where H3 constitutes about 17% of Hg H and is the highest detected so far (Pereira et al. 2004; Quintans et al. 2004).

The coalescence ages of H2a1 and H3 fall to the period of postglacial recolonization in Europe (table 1), suggested first for mtDNA Hg V (Torroni et al. 1998, 2001a). We also note that mtDNA bearing “St. Luke motif,” 16235–16293 (Vernesi et al. 2001), belong to sub-Hg H2 (fig. 2A), being particularly frequent in Germany and Scotland (Helgason et al. 2001; Pfeiffer et al. 2001).

The Near Eastern samples cluster together with Central Asian mtDNAs in the sub-Hgs H6b and H8, which are very rare in Europe. The finding is demonstrating a separate flow of maternal lineages south of the Caspian and the Black Sea in addition to well-known long-lasting migrations of pastoral nomads alongside the steppe belt that connects the Danube Basin, over the Pontic-Caspian, with Central Asia, Altay, and Manchuria.

In contrast to that found in Europeans, sub-Hgs H6 and H8 among Central Asian/Altaian populations are characterized by distinctly divergent haplotypes (fig. 2A). This finding may reflect a long-time separation of Asian and European H6 and H8 mtDNA pools and/or an earlier expansion of H6 in the eastern part of its present range. Indeed, the coalescence age of H6 in Central Asians is very deep—40,400 years (SD 16,400 years; table S1). Because the Asian branches of sub-Hg H6 are highly divergent and seem to be among the oldest in Hg H (table S1), they pose an interesting problem, deserving specific study with a much larger sample size at hand.

The commonly used HVSI clock (Forster et al. 1996) places the initial expansion of Hg H in the Near East to about 23,000 to 28,000 years before the present (Richards et al. 2000). The ancestral clades of Hg H, pre-HV, and HV* have their combined present range predominantly in the Near and Middle East, and in the Caucasus (Metspalu et al. 1999; Richards et al. 2002), implying this could have been the region where the pre-HV/HV clade started to diversify and, possibly, where the earliest Hg H variants might have first appeared.

However, most subclusters of Hg H exhibit coalescence ages, corresponding to the beginning of their expansion in the Late Upper Paleolithic (tables 1 and S1). In this respect our results support an earlier proposal that Hg H was the major mtDNA haplogroup participating in the recolonization of Europe after the Last Glacial Maximum (Torroni et al. 1998; Richards et al. 2000). It is also important to note that the expansion time estimates derived from the coding region and HVSI of Hg H are often in reasonable agreement with each other (tables 1 and S1). Sub-Hgs H1 and H3 have their highest frequencies in the Iberian Peninsula. These sub-Hgs may have been the companions of mtDNA Hg V in the postglacial repopulation of Europe from a refuge area in Iberia (Torroni et al. 1998). However, in contrast to Hg V, suggested coalescence ages of H1 and H3—13,400 ± 3,000 and 8,600 ± 2,800 years ago, respectively (Pereira et al. 2004)—do not imply deeper phylogeny of H1 and H3 in Iberia compared to the rest of Europe (tables 1 and S1).

These results demonstrate that a seemingly uniform spread of this major human mtDNA clade in western Eurasian populations hides within itself a complex structure of phylogeographically informative subclades. However, it is evident that additional knowledge at the level of complete mtDNA sequences is still needed for a truly comprehensive cataloguing of Hg H diversity, in particular more effectively covering its variation in the Mediterranean, Near and Middle Eastern, and Central Asian/Altaian populations. Nevertheless, even now it is tempting to speculate that much deeper coalescence ages, close to/overlapping with the boundary between the Middle and Upper Paleolithic, for some Hg H branches in Central Asian/Altaian populations, suggest that the time depth of this predominant haplogroup may be much deeper than its apparent general signal for expansion in Europe. It is, therefore, possible that the carriers of pre-Aurignacian industry identified in Zagros as well as in Altay (Otte and Derevianko 2001) were anatomically modern humans already possessing Hg H.

Supplementary Material

Supplementary Appendixes S1 and S2, figure S1, and tables S1 and S2 are available at the journal’s Web site as well as the Web site of the University of Tartu, Department of Evolutionary Biology (http://www.evolutsoonn.ut.ee/mtDNA-H/).

Acknowledgments

We thank Ille Hilpus and Jaan Lind for technical assistance and we are grateful to Vladimir Ferík for providing Slovak samples. The research of R.V. was supported by Estonian basic research grant S14 and European Community grants ICA1CT20070006 and QLG2-CT-2002–90455. T.K. was supported by Estonian Science Fund grant 5574. The research of B.A.M. was supported by the Russian Foundation for Basic Research (project number 03–04–48162). P.R. received support from the Ministry of Science and Technology of the Republic of Croatia (project number 0196005). The research of V.S. and V.P. was supported by a grant (project number 03–04–49021) from the Russian Foundation for Basic Research and by grants from the President of the Russian Federation (projects number MD-88.2003.04 and NSh-840.2003.4).

Literature Cited


Arnason, U., X. Xu, and A. Gullberg. 1996. Comparison between the complete mitochondrial DNA sequences of Homo and the


Lisa Matisoo-Smith, Associate Editor
Accepted June 1, 2004