*Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, 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,

D5, 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; Kivisild et al. 2003), Central Asia (Comas et al. 1998; Metspalu et al. 1999), Siberia (Saillard et al. 2000; Derbeneva et al. 2002;

Derbeneva et al. 2003), southern Asia (Passarino et
Kivisild et al. 1999, 2003; Bamshad et al. 20
northern Africa (Côrte-Real et al. 1996; Rando et
et al. 2004, fig. 1*A*). At least 26
mtDNA genomes have been sequenced (nearly) co
(Reid, Vernham, and Jacobs 1994; Rieder et a
Levin, Cheng, and Reeder 1999; Ingman et a
Finnilä, Lehtonen, and Majamaa 2001; Maca-Me
2001; Herrnstadt et al. 2002, correction by H
Preston, and Howell 2003; Mishmar et al. 2003
seven Hg H sub-Hgs defined so far, Hgs H1
(Finnilä, Lehtonen, and Majamaa 2001) along wit
and H4 (Herrnstadt et al. 2002) and Hgs H5, H6
(Quintans et al. 2004) cover 74% of Finnish, 68%
U.K. and 77% of Galician Hg H sequences, respe

Attempts to classify Hg H lineages by firs
variable segmentHgRieder

Appendix S2), have not been included. The reduced median network (Bandelt et al. 1995; ε 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; ε', ε 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 remainder of the positions between 16024 and 16383)—and assigned


A
16355
d16183
16172
16266
16051
16189
16356
H1a
16162
16192
16362
16320
16093
16178
16300
Altaians
16311
16092
16249
16362
16167
16189
16092
16243
16169
16342
16301
16111

1993; Malyarchuk and Derenko 2001b; Allard et al. 2002). 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 ne8(LeI-8.2(rary-249(at)-254.3(tcaracte)ri-11.19(z)-252.37the)-249.8(aclad)-252.3((fig.)-TJ0 -1.104 TD[(s1)-6.4().)-3
s Gi-to- tutations-248.8(ati-248.8(Lp)-278.4(b269,-8.7(l,-279.92(wh)-8.6(bch-278.48(fur)-8.6(ehr)-928.1(oembr-8.2(raes)-2490the)]TJT*

sequencing the position in all 12 samples lacking the *D*≻ I 5003 site. The monophyly of sub-Hg H11 is well established by the combination of two RFLPs and by the characteristic HVS-I mutation pattern. These results show that classical indirect DNA polymorphism detection methods, like RFLP, should be backed-up by direct sequencing in order to avoid the ambiguous or even erroneous inference of phylogeny.

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, 2001). We also note that mtDNA bearing "St. Luke motif," 16235–16293 (Vernesi et al. 2001), belong to sub-Hg H2 (fig. 2*A*), 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. 2*A*). 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 HVS-I 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 proposition 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 HVS-I 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 repeopling 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.evolutsioon.ut.ee/mtDNA-H/).

Acknowledgments

We thank Ille Hilpus and Jaan Lind for technical assistance and we are grateful to Vladimir Ferák for providing Slovak samples. The research of R.V. was supported by Estonian basic research grant 514 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).

References

Allard, M. W., K. Miller, M. Wilson, K. Monson, and B. Budowle. 2002. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. Scientific Working Group on DNA Analysis Methods. J. Forensic Sci. : 1215–1223.

Anderson, S., A. T. Bankier, B. G. Barrell et al. (14 co-authors). 1981. Sequence and organization of the human mitochondrial genome. Nature :457–465.

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, and N. Howell. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. :147.

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

common chimpanzee based on nonchimeric sequences. J. Mol. Evol. , :145–152.

1996. Classification of European mtDNAs from an analysis of three European populations. Genetics :1835–1850.

Torroni, A., M. T. Lott, M. F. Cabell, Y. S. Chen, L. Lavergne, and D. C. Wallace. 1994. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic