*Am. J. Hum. Genet. 73:1178–1190, 2003*

## Report

# Origin and Diffusion of mtDNA Haplogroup X

Maere Reidla,[1] Toomas Kivisild,[1] Ene Metspalu,[1] Katrin Kaldma,[1] Kristiina Tambets,[1] Helle-Viivi Tolk,[1] Jüri Parik,[1] Eva-Liis Loogväli,[1] Miroslava Derenko,[2] Boris Malyarchuk,[2] Marina Bermisheva,[1,3] Sergey Zhadanov,[1,4] Erwan Pennarun,[1,5] Marina Gubina,[1,4] Maria Golubenko,[1,6] Larisa Damba,[1,4] Sardana Fedorova,[1,7] Vladislava Gusar,[1,8] Elena Grechanina,[8] Ilia Mikerezi,[9] Jean-Paul Moisan,[5] André Chaventré,[5] Elsa Khusnutdinova,[3] Ludmila Osipova,[4] Vadim Stepanov,[6] Mikhail Voevoda,[4] Alessandro Achilli,[10] Chiara Rengo,[10] Olga Rickards,[11] Gian Franco De Stefano,[11] Surinder Papiha,[12] Lars Beckman,[13] Branka Janicijevic,[14] Pavao Rudan,[14] Nicholas Anagnou,[15] Emmanuel Michalodimitrakis,[16] Slawomir Koziel,[17] Esien Usanga,[18] Tarekegn Geberhiwot,[19] Corinna Herrnstadt,[20] Neil Howell,[20] Antonio Torroni,[10] and Richard Villems[1]

[1]Department of Evolutionary Biology, Institute of Molecular and Cell Biology, Tartu University and Estonian Biocentre, Tartu, Estonia; [2]Genetic Laboratory, Institute of Biological Problems of the North, Magadan, Russia; [3]Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa, Russia; [4]Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia; [5]Laboratoire d'Etude du Polymorphisme de l'ADN, Faculté de Médecine, Nantes, France; [6]Institute of Medical Genetics, Siberian Branch of the Russian Academy of Medical Sciences, Tomsk, Russia; [7]Yakut Scientific Center, Russian Academy of Medical Sciences, and Government of Republic Sakha (Yakutia), Yakutsk, Russia; [8]Center of Clinical Genetics and Prenatal Diagnostics, Kharkov, Ukraine; [9]Department of Biology, Faculty of Natural Sciences, Tirana University, Tirana, Albania; [10]Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy; [11]Dipartimento di Biologia, Università "Tor Vergata," Rome; [12]Department of Human Genetics, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne; [13]Gotland University, Visby, Sweden; [14]Institute for Anthropological Research, Zagreb, Croatia; [15]Institute of Molecular Biology and Biotechnology and Department of Basic Sciences and [16]Department of Forensic Sciences and Toxicology, University of Crete School of Medicine, Heraklion, Greece; [17]Institute of Anthropology, Wroclaw, Poland; [18]Department of Haematology, University of Calabar, Calabar, Nigeria; [19]Birmingham and Solihull Teaching Hospital, Birmingham; and [20]MitoKor, San Diego, CA

**A maximum parsimony tree of 21 complete mitochondrial DNA (mtDNA) sequences belonging to haplogroup X and the survey of the haplogroup-associated polymorphisms in 13,589 mtDNAs from Eurasia and Africa revealed that haplogroup X is subdivided into two major branches, here defined as "X1" and "X2." The first is restricted to the populations of North and East Africa and the Near East, whereas X2 encompasses all X mtDNAs from Europe, western and Central Asia, Siberia, and the great majority of the Near East, as well as some North African samples. Subhaplogroup X1 diversity indicates an early coalescence time, whereas X2 has apparently undergone a more recent population expansion in Eurasia, most likely around or after the last glacial maximum. It is notable that X2 includes the two complete Native American X sequences that constitute the distinctive X2a clade, a clade that lacks close relatives in the entire Old World, including Siberia. The position of X2a in the phylogenetic tree suggests an early split from the other X2 clades, likely at the very beginning of their expansion and spread from the Near East.**

mtDNA and the nonrecombining part of the Y chromosome are widely used in archaeogenetic studies (Renfrew 2000; Cavalli-Sforza and Feldman 2003) that aim to reveal the human past. The uniparental inheritance and complete linkage of mutations in these two loci allow an unambiguous determination of the phylogenetic relationships between individual lineages. However, the putative genetic histories of the lineages that are obtained do not fully reflect the complex dynamics of ancient populations; thus, the data must be interpreted carefully. Phylogenetic clustering of mtDNA haplogroups has been found to be congruent with geography—there are haplogroups specific to African (Chen et

Received June 11, 2003; accepted for publication August 27, 2003; electronically published October 20, 2003.

Address for correspondence and reprints: Dr. Maere Reidla, Department of Evolutionary Biology, Tartu University, Estonian Biocentre, Riia 23, Tartu, 51010, Estonia. E-mail: mreidla@ebc.ee


0002-9297/2003/7305-0020$15.00

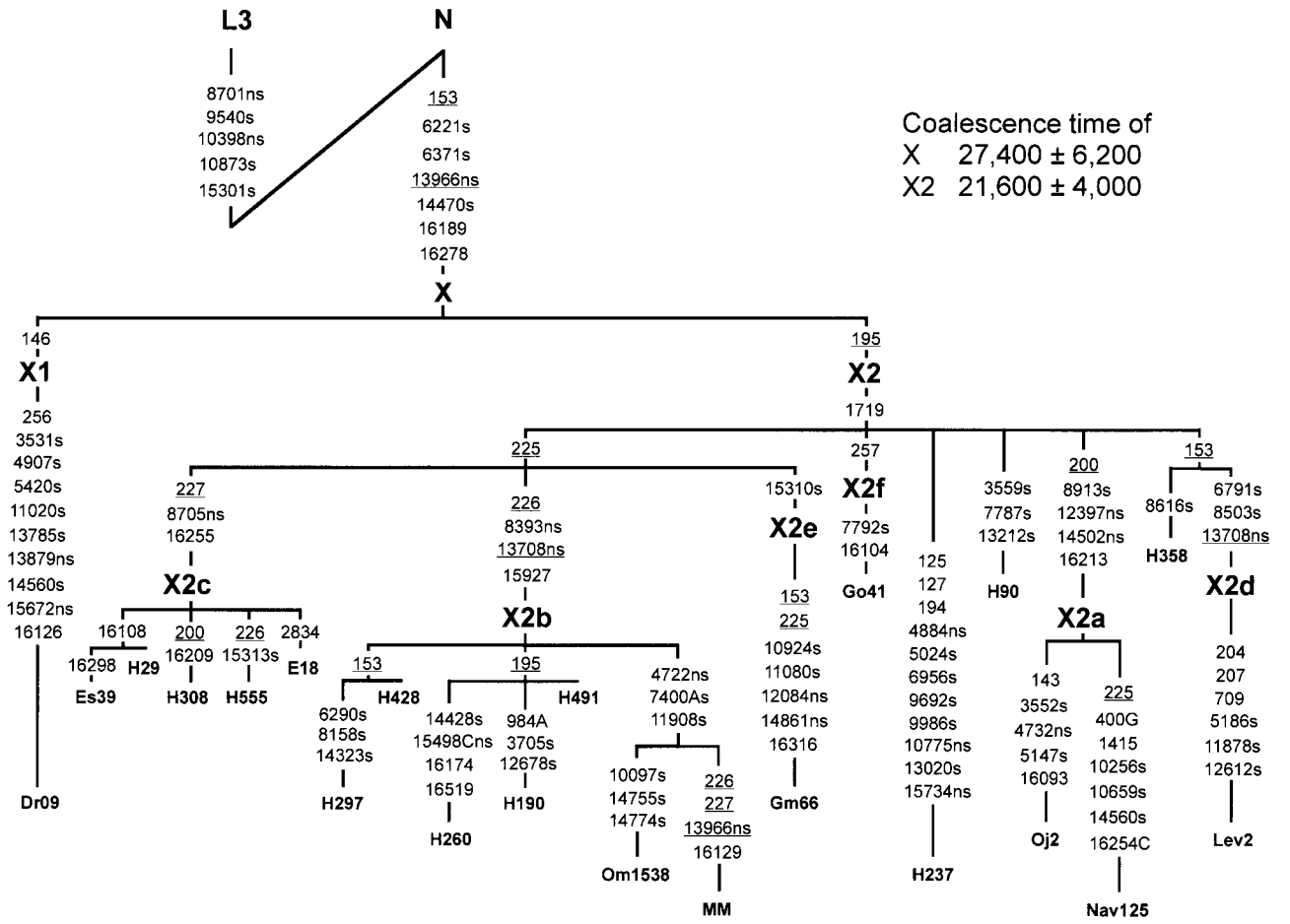
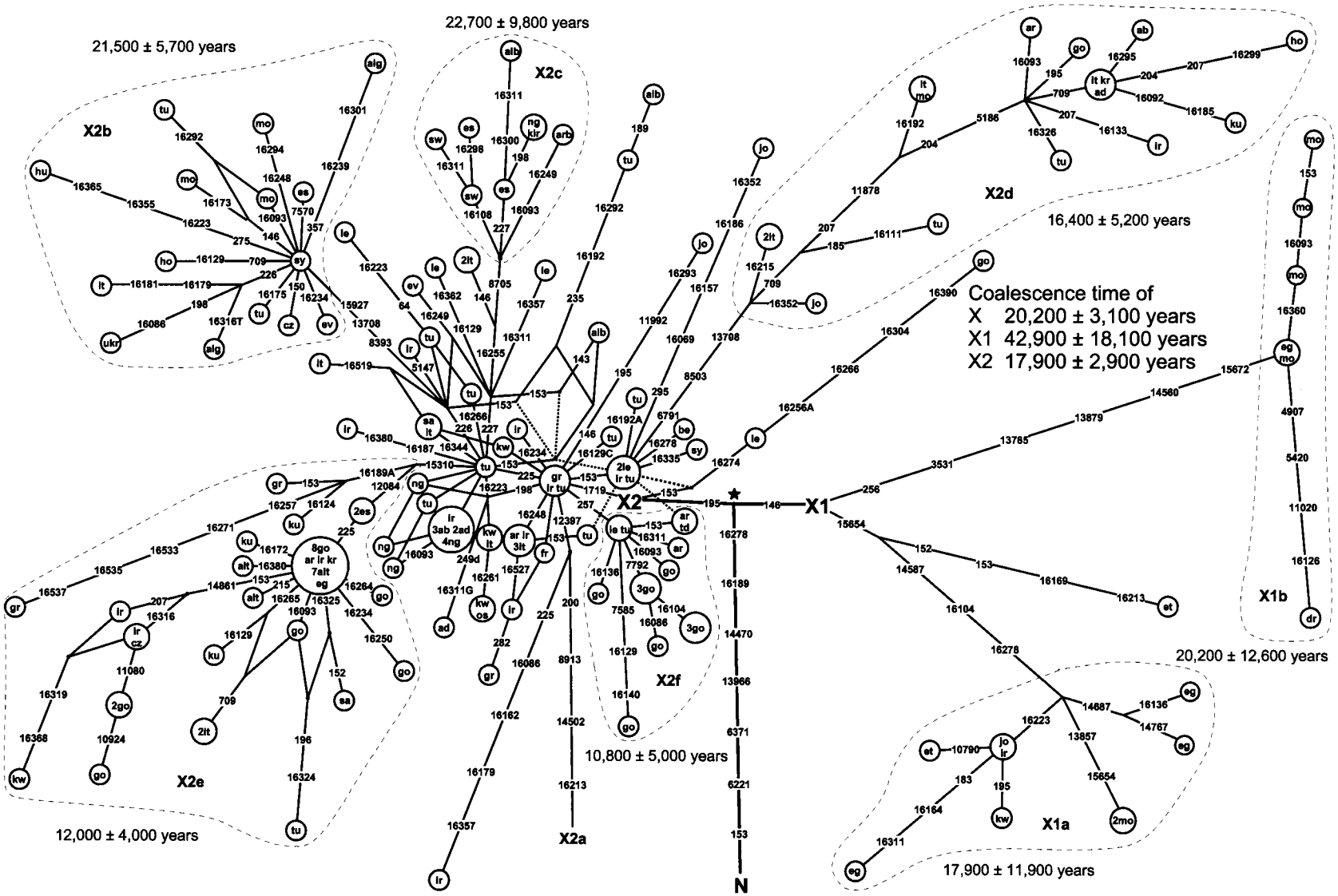



**Figure 1** Maximum parsimony phylogenetic tree of complete mtDNA sequences belonging to haplogroup X. Five mtDNAs were selected for complete sequence analysis in the course of the present study (Dr09, Es39, Gm66, Go41, and Om1538), 11 coding region sequences were from the work of Herrnstadt et al. (2002), but their control region sequences have now been added, and 5 complete sequences (Lev2, MM, E18, Nav125, and Oj2) were taken from the literature (Levin et al.1999; Maca-Meyer et al.2001; Mishmar et al. 2003; Bandelt et al., in press). The diagnostic mutations relative to the revised reference sequence (Andrews et al. 1999) are indicated on the branches. Transversions are specified by suffixes, and underlined mutations appear more than once in the tree. For protein-coding genes, "ns" indicates nonsynonymous and "s" synonymous mutations. For coalescence-time estimates, a mutation rate of 1 mutation per 5,139 years in the coding region (nps 577–16023) was used (Mishmar et al. 2003). mtDNA sequence data are available at the Estonian Biocentre Web page.

al. 1995; Watson et al. 1997), Asian (Ballinger et al. 1992; Torroni et al. 1994; Kivisild et al. 2002), European/West Eurasian (Torroni et al. 1996; Macaulay et al. 1999), and Native American (Torroni et al. 1993) populations.

Haplogroup X is an exception to this pattern of limited geographical distribution. It is found, generally at low frequencies, in both West Eurasians (Richards et al. 2000) and some northern groups of Native Americans (Ward et al. 1991; Forster et al. 1996; Scozzari et al. 1997; Brown et al. 1998; Smith et al. 1999; Malhi et al. 2001), but, intriguingly, it is absent in modern north Siberian and East Asian populations (Brown et al. 1998; Starikovskaya et al. 1998; Schurr et al. 1999), which are genetically and geographically closest to those of Native Americans. Among Siberians, haplogroup X mtDNAs have only been detected in some Altaian populations of southwestern Siberia (Derenko et al. 2001).

When the sequence variation of the first hypervariable segment (HVS-I) of the control region is analyzed, haplogroup X mtDNAs from Europe and the Near East are found to yield similar coalescence times: 17,000–30,000 years before present (YBP) and 13,700–26,600 YBP, respectively (Richards et al. 2000). These estimates are consistent with a pre-Holocene origin and spread of this haplogroup into West Eurasia. For Native Americans, the relatively old presence of haplogroup X is confirmed by the analysis of ancient human remains (Stone and Stoneking 1999; Malhi and Smith 2002). Moreover, Native American haplogroup X mtDNAs form a clade dis-


21,500 ± 5,700 years
22,700 ± 9,800 years
X2b
X2c
X2d
16,400 ± 5,200 years
Coalescence time of
X 20,200 ± 3,100 years
X1 42,900 ± 18,100 years
X2 17,900 ± 2,900 years
X1
X2
X1b
20,200 ± 12,600 years
X2e
12,000 ± 4,000 years
X2f
10,800 ± 5,000 years
X2a
X1a
17,900 ± 11,900 years
N

tinct from that of West Eurasians and with coalescence time estimates varying widely depending on both the method of estimation and the number of assumed founders. Thus, the coalescence times ranged from 12,000–17,000 YBP to 23,000–36,000 YBP, times that are consistent with both a pre- and a postglacial population diffusion (Brown et al. 1998).

To obtain further information about the extent of haplogroup X diversity, 5 mtDNAs (from 1 Druze, 1 Estonian, 2 Georgians, and 1 Omani) were completely sequenced and were compared with 16 previously published X sequences (fig. 1). These latter sequences included the 11 haplogroup X coding sequences published by Herrnstadt et al. (2002) that have now been completed with the sequencing of the control region.

A maximum parsimony tree of the 21 haplogroup X sequences revealed that one nonsynonymous (13966) and three synonymous (6221, 6371, and 14470) substitutions in the coding region, as well as three transitions in the control region (153, 16189, and 16278), distinguish haplogroup X from the root of superhaplogroup N. Moreover, haplogroup X is subdivided into two major subhaplogroups, designated "X1" and "X2." Subhaplogroup X1, represented by a single Druze mtDNA in figure 1, differs from the root of haplogroup X by eight coding and three control region transitions and lacks the two transitions (195 and 1719) that characterize X2. These two nucleotides are rather mutable (Finnilä et al. 2001; Herrnstadt et al. 2002); thus, it cannot be completely ruled out that X1 is indeed a subset of X2 that reverted at both nucleotide positions. However, this possibility appears very unlikely, especially when one considers the time depth and the distinct geographic distribution of X1 (see below).

In contrast with X1, X2 is well represented in the tree, and it is further subdivided into at least six major clades (X2a–X2f), which include clades X1 and X2 as defined by Herrnstadt et al. (2002). All 20 X2 sequences, including the 2 Native American X2a sequences, share transitions at nucleotide positions (nps) 195 and 1719. A recurrence of the nonsynonymous substitution at 13708 was observed (clades X2b and X2d). In addition, the nonsynonymous transition 13966 was found to have reverted in the Moroccan X2b sequence. These were the only recurrent mutations found among the 67 variable positions in the coding region sequences. The ratio of nonsynonymous to synonymous substitutions was 0.40 (18/45). Six mutations were located in RNA genes.

The data obtained from the analyses of complete mtDNA sequences belonging to haplogroup X were then used to survey 13,589 mtDNAs (21,682 when mtDNA data from the literature are included) from 66 populations of Eurasia and North Africa (table 1). A total of 175 mtDNAs were found to harbor the four coding region transitions—at 6221, 6371, 13966, and 14470—that define haplogroup X. The four markers were always found in association and, by combination of the control and coding variation, all 175 mtDNAs could be apportioned to subhaplogroups X1 and X2 (table 2). No other main branches occur, and the root haplotype was not present among the sequences. The adenine at np 153 appears relatively conserved in the phylogenetic context of 376 complete human mtDNA sequences taken worldwide, exhibiting a change to guanine only in haplogroups A and X (Ingman et al. 2000; Finnilä et al. 2001; Maca-Meyer et al. 2001; Derbeneva et al. 2002; Mishmar et al. 2003). In contrast, this position shows a high level of variation in the background of haplogroup X as both the 153A and 153G alleles are present in its different subclades. It is possible that the A→G transition at 153 arose only once in the haplogroup X ancestor and the recurrent reverse mutations in 11 branches in figure 2 bear witness to the process of favored fixation of the more stable A allele.

Subhaplogroup X1 was found to be largely restricted to the Afro-Asiatic–speaking populations of North Africa and neighboring areas, including Ethiopia, suggesting a possible geographic diffusion of X1 alongside the

**Figure 2** Median-joining (MJ) network of 175 haplogroup X partial sequences. The MJ (Bandelt et al. 1999) algorithm was implemented within the Network 3111 program (A. Röhl; Shareware Phylogenetic Software Web site). Default settings were used for HVS-I (nps 16024–16383), HVS-II (nps 16518–310), and coding-region sequence variation as given in table 2. Highly variable (Hasegawa et al. 1993) positions of HVS-I (16093, 16126, 16129, 16187, 16189, 16223, 16234, 16278, 16292, 16293, 16311, 16325, 16355, and 16362) were assigned a weight of 1, other HVS-I and HVS-II sites were assigned a weight of 2, and coding-region sites were assigned a weight of 10. Areas of the circles are proportional to haplotype frequencies. Populations are indicated by the following abbreviations: ab = Abazin, ad = Adygei, alb = Albanian, alg = Algerian, alt = Altaian, ar = Armenian, arb = Arab from Uzbekistan, be = Bengali, cz = Czech, dr = Druze, eg = Egyptian, es = Estonian, et = Ethiopian, ev = Evenk, fr = French, go = Georgian, gr = Greek, ho = Croat, hu = Hungarian, ir = Iranian, it = Italian, jo = Jordanian, kir = Kyrgyz, kr = Karachay, ku = Kumyk, kw = Kuwaiti, le = Lebanese, mo = Moroccan, ng = Nogay, os = Ossetian, sa = Saudi Arabian, sw = Swede, sy = Syrian, td = Tadjik, tu = Turk, and ukr = Ukrainian. Variant bases are numbered (Anderson et al. 1981) and are shown along links between haplotypes. Nucleotide changes are specified by suffixes only for transversions, and a "d" indicates a deletion. The node marked with a large asterisk (★) matches the root type of haplogroup X. The coalescence times of the clades are shown near the clades. Coalescence times of HVS-I clusters were calculated by means of $\rho$, the average mutational distance to the founder haplotype of the cluster, by using a mutation rate of 1 transition per 20,180 years in the segment between nps 16090 and 16365 (Forster et al. 1996). Standard deviations for $\rho$ were calculated as in the work of Saillard et al. (2000), a procedure which ignores the variance due to molecular clock calibration. mtDNA sequence data are available at the Estonian Biocentre Web page.

## Table 1

**Frequency and Diversity of Subhaplogroups X1 and X2 in Eurasian and African Populations**

| Region and Population[a] | Sample Size | Value for Subhaplogroup[b,c] | | | |
|---|---|---|---|---|---|
| | | X1 | | X2 | |
| | | Mean Frequency [95% CR] (%) | Diversity | Mean Frequency [95% CR] (%) | Diversity |
| North and northeastern Africa: | 1,606 | .8 [.5–1.4] | .936 | .9 [.5–1.5] | .978 |
| Ethiopians[1] | 270 | .7 [.2–2.6] | | 0 [.0–1.1] | |
| Egyptians[1] | 193 | 2.1 [.8–5.2] | | .5 [.1–2.8] | |
| Libyans[1] | 72 | 0 [.0–4.0] | | 0 [.0–4.0] | |
| Tunisians[1] | 417 | .2 [.1–1.3] | | 1.7 [.8–3.4] | .952 |
| Algerians[1] | 124 | 0 [.0–2.4] | | 1.6 [.5–5.7] | |
| Moroccans[1] | 530 | 1.1 [.5–2.4] | | .8 [.3–1.9] | |
| West and southern Africa: | 537 | 0 [.0–.6] | | 0 [.0–.6] | |
| *Senegalese[27] | 121 | 0 [.0–2.4] | | 0 [.0–2.4] | |
| *17 ethnic groups in Mozambique[24,32] | 416 | 0 [.0–.7] | | 0 [.0–.7] | |
| Near East: | 2,299 | .5 [.3–.9] | .564 | 2.9 [2.3–3.6] | .887 |
| Yemenis[1] | 116 | 0 [.0–2.5] | | .9 [.2–4.7] | |
| Omanis[1] | 78 | 0 [.0–3.7] | | 1.3 [.3–6.9] | |
| Saudi-Arabians[1] | 204 | 0 [.0–1.5] | | 1.5 [.5–4.2] | |
| Kuwaitis[1] | 202 | .5 [.1–2.7] | | 2.0 [.8–5.0] | |
| *Israeli Palestinians[9,29] | 117 | 0 [.0–2.5] | | 3.4 [1.4–8.5] | |
| Israeli Druze[17] | 45 | 15.6 [7.8–28.9] | 0 | 11.1 [4.9–23.6] | |
| Jordanians[1] | 202 | .5 [.1–2.7] | | 1.5 [.5–4.3] | |
| Lebanese[1] | 172 | 0 [.0–1.7] | | 5.8 [3.2–10.4] | .867 |
| Syrians[1] | 219 | 0 [.0–1.4] | | 1.8 [.7–4.6] | |
| *Iraqis[29] | 116 | .9 [.2–4.7] | | .9 [.2–4.7] | |
| Turks[1] | 388 | 0 [.0–.8] | | 4.4 [2.8–6.9] | .919 |
| Iranians[1] | 440 | .2 [.1–1.3] | | 3.0 [1.7–5.0] | .833 |
| Mediterranean Europe: | 2,900 | .0 [.0–.2] | | 2.5 [2.0–3.2] | .859 |
| Cypriots[1] | 179 | 0 [.0–1.7] | | 6.7 [3.9–11.4] | .864 |
| Greeks from mainland and Crete[1] | 273 | 0 [.0–1.1] | | 4.4 [2.5–7.5] | .455 |
| Albanians[1] | 199 | 0 [.0–1.5] | | 2.5 [1.1–5.7] | |
| Croats[1] | 884 | 0 [.0–.3] | | .9 [.5–1.8] | .786 |
| Italians from mainland and Sicily[1,19,33] | 859 | 0 [.0–.3] | | 2.9 [2.0–4.3] | .923 |
| *Basques[3,4,29] | 147 | 0 [.0–2.0] | | 1.4 [.4–4.8] | |
| Spaniards[5] | 118 | 0 [.0–2.5] | | 4.2 [1.9–9.5] | |
| Portugese[23] | 241 | .4 [.1–2.3] | | 1.7 [.7–4.2] | |
| Northwestern Europe: | 1,775 | 0 [.0–.2] | | 1.7 [1.2–2.5] | .665 |
| French[1,30] | 398 | 0 [.0–.7] | | .8 [.3–2.2] | |
| *English[12,26,28] | 334 | 0 [.0–.9] | | .9 [.3–2.6] | |
| *Scots[12] | 891 | 0 [.0–.3] | | 1.6 [.9–2.6] | .692 |
| *Orkney inhabitants[12] | 152 | 0 [.0–1.9] | | 7.2 [4.1–12.5] | .473 |
| Scandinavia: | 1,271 | 0 [.0–.2] | | .9 [.5–1.5] | .491 |
| Swedes[1] | 318 | 0 [.0–.9] | | .6 [.2–2.2] | |
| *Norwegians[12,20,29] | 559 | 0 [.0–.5] | | .4 [.1–1.3] | |
| Icelanders[11] | 394 | 0 [.0–.8] | | 1.8 [.9–3.6] | 0 |
| Alps: | 255 | 0 [.0–1.2] | | .8 [.2–2.8] | |
| Swiss[8] | 154 | 0 [.0–1.9] | | .6 [.2–3.5] | |
| Austrians[22] | 101 | 0 [.0–2.9] | | 1.0 [.2–5.3] | |
| North-Central Europe: | 1,419 | 0 [.0–.2] | | 1.3 [.8–2.0] | .948 |
| Poles[1,18] | 547 | 0 [.0–.5] | | 1.6 [.9–3.1] | .972 |
| *Germans[13,16,25,28] | 532 | 0 [.0–.6] | | 1.1 [.5–2.4] | |
| Czechs[1] | 94 | 0 [.0–3.1] | | 2.1 [.7–7.4] | |
| Slovaks[1] | 130 | 0 [.0–2.3] | | 0 [.0–2.3] | |
| Hungarians[1] | 116 | 0 [.0–2.5] | | .9 [.2–4.7] | |
| Northeastern Europe: | 1,611 | 0 [.0–.2] | | 1.1 [.7–1.8] | .902 |
| Finns[10] | 192 | 0 [.0–1.5] | | 2.1 [.8–5.2] | |
| Estonians[1] | 401 | 0 [.0–.7] | | 1.2 [.5–2.9] | |
| Latvians[1] | 192 | 0 [.0–1.5] | | .5 [ 0.1–2.9] | |
| Russians[1,18,21] | 826 | 0 [.0–.4] | | 1.0 [.5–1.9] | .964 |
| Volga-Ural region: | 1,037 | 0 [.0–.3] | | 0 [.0–.3] | |
| Komis[2] | 136 | 0 [.0–2.2] | | 0 [.0–2.2] | |
| Udmurds[2] | 101 | 0 [.0–2.9] | | 0 [.0–2.9] | |
| Maris[2] | 136 | 0 [.0–2.2] | | 0 [.0–2.2] | |
| Mordvins[2] | 102 | 0 [.0–2.9] | | 0 [.0–2.9] | |
| Nenets[31] | 58 | 0 [.0–5.0] | | 0 [.0–5.0] | |
| Chuvashis[2] | 55 | 0 [.0–5.2] | | 0 [.0–5.2] | |
| Tatars[2] | 228 | 0 [.0–1.3] | | 0 [.0–1.3] | |
| Bashkirs[2] | 221 | 0 [.0–1.3] | | 0 [.0–1.3] | |

(*continued*)

## Table 1 (continued)

| Region and Population[a] | Sample Size | Value for Subhaplogroup[b,c] X1 Mean Frequency [95% CR] (%) | X1 Diversity | X2 Mean Frequency [95% CR] (%) | X2 Diversity |
|---|---|---|---|---|---|
| Southeastern Europe: | 357 | 0 [.0–.8] | | 2.0 [1.0–4.0] | .952 |
| Ukrainians[1] | 357 | 0 [.0–.8] | | 2.0 [1.0–4.0] | .952 |
| Northern Caucasus: | 838 | 0 [.0–.4] | | 2.7 [1.8–4.1] | .798 |
| Nogays[1] | 213 | 0 [.0–1.4] | | 4.2 [2.3–7.8] | .722 |
| Adygeis[1] | 159 | 0 [.0–1.9] | | 2.5 [1.0–6.3] | |
| Karachays[1] | 106 | 0 [.0–2.8] | | 1.9 [.6–6.6] | |
| Abazins[1] | 63 | 0 [.0–4.6] | | 6.3 [2.6–15.2] | |
| Kabardins[1] | 66 | 0 [.0–4.4] | | 0 [.0–4.4] | |
| Kumyks[1] | 111 | 0 [.0–2.6] | | 3.6 [1.5–8.9] | |
| Kalmyks[1] | 120 | 0 [.0–2.4] | | 0 [.0–2.4] | |
| South Caucasus: | 782 | 0 [.0–.4] | | 4.3 [3.1–6.0] | .797 |
| Georgians[1] | 340 | 0 [.0–.9] | | 7.6 [5.3–11.0] | .738 |
| Southern Ossetians[1] | 201 | 0 [.0–1.5] | | .5 [.1–2.7] | |
| Armenians[1] | 193 | 0 [.0–1.5] | | 2.6 [1.1–5.9] | |
| *Azeris[29] | 48 | 0 [.0–5.9] | | 4.2 [1.3–14.0] | |
| Central Asia: | 1,036 | 0 [.0–.3] | | .8 [.4–1.5] | .929 |
| Uighurs from Kazakhstan[1] | 122 | 0 [.0–2.4] | | .8 [.2–4.4] | |
| Kazakhs[1] | 495 | 0 [.0–.6] | | .6 [.2–1.8] | |
| Kyrgyz[1] | 105 | 0 [.0–2.8] | | 1.0 [.2–5.1] | |
| Tadjik[1] | 77 | 0 [.0–3.8] | | 1.3 [.3–6.9] | |
| Uzbeks[1] | 160 | 0 [.0–1.8] | | .6 [.2–3.4] | |
| Arabs from Uzbekistan[1] | 77 | 0 [.0–3.8] | | 1.3 [.3–6.9] | |
| India: | 1,010 | 0 [.0–.3] | | .2 [.1–.7] | |
| Western Bengalis[1] | 106 | 0 [.0–2.8] | | .9 [.2–5.1] | |
| Gujaratis and Konkanastha Br.[1] | 111 | 0 [.0–2.6] | | 0 [.0–2.6] | |
| Punjabis[1] | 112 | 0 [.0–2.6] | | 0 [.0–2.6] | |
| Moors[1] | 50 | 0 [.0–5.7] | | 0 [.0–5.7] | |
| Singhalese from Sri Lanka[1] | 82 | 0 [.0–3.5] | | 0 [.0–3.5] | |
| Uttar Pradesh[14] | 122 | 0 [.0–2.4] | | .8 [.2–4.4] | |
| Telugus[14] | 250 | 0 [.0–1.2] | | 0 [.0–1.2] | |
| Chenchu[15] | 96 | 0 [.0–3.0] | | 0 [.0–3.0] | |
| Koya[15] | 81 | 0 [.0–3.6] | | 0 [.0–3.6] | |
| Siberia: | 2,949 | 0 [.0–.1] | | .4 [.2–.7] | .455 |
| Altaians[1,7] | 481 | 0 [.0–.6] | | 1.9 [1.0–3.5] | 0 |
| Buryats[7] | 105 | 0 [.0–2.8] | | 0 [.0–2.8] | |
| Tuvins[1,7] | 314 | 0 [.0–.9] | | 0 [.0–.9] | |
| Koryaks[1,7] | 105 | 0 [.0–2.8] | | 0 [.0–2.8] | |
| Evens[7] | 65 | 0 [.0–4.4] | | 0 [.0–4.4] | |
| Evenks[1,7] | 185 | 0 [.0–1.6] | | 1.1 [.3–3.8] | |
| Yakuts[1,7] | 340 | 0 [.0–.9] | | 0 [.0–.9] | |
| Khakassians[7] | 54 | 0 [.0–5.3] | | 0 [.0–5.3] | |
| Shors[1,7] | 207 | 0 [.0–1.4] | | 0 [.0–1.4] | |
| Sojots[7] | 34 | 0 [.0–8.2] | | 0 [.0–8.2] | |
| Kets[1] | 66 | 0 [.0–4.4] | | 0 [.0–4.4] | |
| Selkups[1] | 120 | 0 [.0–2.4] | | 0 [.0–2.4] | |
| Nenets[1] | 79 | 0 [.0–3.7] | | 0 [.0–3.7] | |
| Dolgans[1] | 130 | 0 [.0–2.3] | | 0 [.0–2.3] | |
| Nanais[1] | 88 | 0 [.0–3.3] | | 0 [.0–3.3] | |
| Komis[1] | 78 | 0 [.0–3.7] | | 0 [.0–3.7] | |
| Nganasans[1] | 107 | 0 [.0–2.7] | | 0 [.0–2.7] | |
| Khants[1] | 253 | 0 [.0–1.2] | | .4 [.1–2.2] | |
| Mansis[1,6] | 138 | 0 [.0–2.1] | | 0 [.0–2.1] | |

[a] An asterisk (*) denotes frequencies deduced from published HVS-I sequences. Populations are divided into regions basically as in (Richards et al. 2000), with some changes: North African and South Caucasian populations were considered separately from populations of the Near East, and all European populations of the Mediterranean area were aggregated into a single Mediterranean European region. Sources are denoted as follows: [1]present study, [2]Bermisheva et al. (2002), [3]Bertranpetit et al. (1995), [4]Corte-Real et al. (1996), [5]Crespillo et al. (2000), [6]Derbeneva et al. (2002), [7]Derenko et al. (2001), [8]Dimo-Simonin et al. (2000), [9]Di Rienzo and Wilson (1991), [10]Finnilä et al. (2001), [11]Helgason et al. (2000), [12]Helgason et al. (2001), [13]Hofmann et al. (1997), [14]Kivisild et al. (1999), [15]Kivisild et al. (2003), [16]Lutz et al. (1998), [17]Macaulay et al. (1999), [18]Malyarchuk et al. (2002), [19]Mogentale-Profizi et al. (2001), [20]Opdal et al. (1998), [21]Orekhov et al. (1999), [22]Parson et al. (1998), [23]Pereira et al. (2000), [24]Pereira et al. (2001), [25]Pfeiffer et al. (1999), [26]Piercy et al. (1993), [27]Rando et al. (1998), [28]Richards et al. (1996), [29]Richards et al. (2000), [30]Rousselet and Mangin (1998), [31]Saillard et al. (2000), [32]Salas et al. (2002), [33]Tagliabracci et al. (2001).

[b] The 95% credible regions (CR), calculated using software kindly provided by Vincent Macaulay, are shown in brackets.

[c] Haplotype diversities were calculated as in the work of Nei (1987), only for regions where more than six haplogroup X mtDNAs were available considering the sequence variation observed between nps 16090 and 16365.

## Table 2

### Control-Region and Coding-Region Variation of Haplogroup X mtDNAs

| Sequence[c] | Haplogroup | Control Region Sequence[a] HVS-I[d] | Control Region Sequence[a] HVS-II[e] | Coding Region Sequence and RFLP Variation[b] |
|---|---|---|---|---|
| | | | | 1111111111111111111111111111 |
| | | | | 1122333445556666777888889000111122233333344444445555 |
| | | | | 7475855679114233755735579827900890360377889455567836 69 |
| | | | | 0113335230482268978990701359228798903608576606886615 72 |
| | | | | 9357111427760193105213253040400724781645796520777104 25 |
| | | | | lckg w k we e e j u ygmt s i |
| CRS[f] | | | | G++--G+A-TGATT-+AAAC+T+TAC+TCAT+TCA++-+CATA-TGAATGTTT+ |
| et 216 | X1* | 169 189 213 223 278 | 146 152 | G + T TC T A + + GAAT CT |
| eg 96 | X1a | 104 136 189 223 | 146 153 | G + G T ATC+ A T T A + +CATG+ GGGT CT |
| eg 818 | X1a | 104 164 189 311 | 146 153 183 | G + G T T T A +CATG+ GGAT CT |
| et 146 | X1a | 104 189 | 146 153 | G + T T + A T C A + + + G+ GGAT CT |
| ir 143 | X1a | 104 189 | 146 153 | G + G T T T A +CATG+ GGAT CT |
| jo 924 | X1a | 104 189 | 146 153 | G + G T T T A +CATG+ GGAT CT |
| kw 103 | X1a | 104 189 | 146 153 195 | G + G T ATC+ A T T A + +CATG+ GGAT CT |
| eg 971 | X1a | 104 189 223 | 146 153 | G + G T T T A +CATG+ GGGC CT |
| mo 90E | X1a | 104 189 223 | 146 153 | G + G T T T A +CGTG+ GGAT TT |
| mo E99 | X1a | 104 189 223 | 146 153 | G + GTG+ GG TT |
| dr 09 | X1b | 126 189 223 278 | 146 153 256 | G++--A+A-CGACC++AAAC+T+TAC+TCGT+TCA++-+TACG+TAAATGTTC+ |
| eg 443 | X1b | 189 223 278 | 146 153 256 | G + A T ATC+ A T T A + +TACG+ AAAT TC |
| mo C49 | X1b | 189 223 278 | 146 153 256 | + T T C + |
| mo A79 | X1b | 189 223 278 360 | 146 153 256 | G + A T T A TACG+ AA TC |
| mo A17 | X1b | 93 189 223 278 360 | 146 153 256 | G + T T A TACG+ TC |
| mo B64 | X1b | 93 189 223 278 360 | 146 256 | + T T C + |
| le 938 | X2* | 129 189 223 278 362 | 153 195 225 227 | G - - G + +T T |
| tu 47 | X2* | 129C 189 223 278 | 153 195 | G+- + - A + A T + + A+ + G+ T |
| ir 71 | X2* | 187 189 223 278 380 | 153 195 225 | G - + - G A A + +T T |
| alb 53 | X2* | 189 192 223 278 292 | 143 189 195 225 226 235 | G - + - GA C+ A A C T+ A+ + +T T + |
| tu 208 | X2* | 189 192 223 278 292 | 143 195 225 226 235 | G - + - GA + A T A C T+ A+ -+ G+T T + |
| tu 296 | X2* | 189 192A 223 278 | 195 | G -- + - GA + A T A + CA+ + G+T GT |
| be 50 | X2* | 189 223 | 195 | G - A + A T + C + + G+ T |
| ir 161 | X2* | 189 223 234 278 | 195 225 | G - + - G A CA + +T T |
| ir 282 | X2* | 189 223 248 278 | 153 195 | G - + - G A A + +T T |
| it 173 | X2* | 189 223 248 278 | 153 195 | G - + C + + T |
| it 74 | X2* | 189 223 248 278 | 153 195 | G - - G + T A C + +T T |
| it 33 | X2* | 189 223 248 278 | 153 195 | G - + A T + C + + G+ T |
| ar 175 | X2* | 189 223 248 278 | 153 195 | G+- - G + T A + + + + G+T T |
| ng 227 | X2* | 189 223 248 278 | 153 195 198 225 | - - + + T |
| ir 224 | X2* | 189 223 248 278 | 153 195 225 | G - + - G TA A + +T T |
| ng 167 | X2* | 189 223 248 278 | 153 195 225 | G - - G + +T T |
| ab 17 | X2* | 189 223 248 278 | 153 195 225 | G - + - G A CA + +T T |
| ab 31 | X2* | 189 223 248 278 | 153 195 225 | - + - G A A + +T T |
| ab 44 | X2* | 189 223 248 278 | 153 195 225 | G - + + T |
| ad 67 | X2* | 189 223 248 278 | 153 195 225 | G - - G A + +T T |
| ad 79 | X2* | 189 223 248 278 | 153 195 225 | - - G A + +T T |
| ng 2 | X2* | 189 223 248 278 | 153 195 225 | - + A + + T |
| ng 264 | X2* | 189 223 248 278 | 153 195 225 | G - + + T |
| ng 6 | X2* | 189 223 248 278 | 153 195 225 | G - + A A + + T |
| ir 240 | X2* | 189 223 248 278 | 16527 153 195 | G - + - G A CA + +T T |
| gr 34 | X2* | 189 223 248 278 | 16527 153 195 282 | G - + + + |
| tu 196 | X2* | 189 223 248 278 | 195 | G+-- + - GA + A T A + + C + + G+T T |
| ev 4076 | X2* | 189 223 249 278 | 153 195 225 226 227 | G - - G A + +T T |
| it 4202 | X2* | 189 223 255 278 | 146 153 195 225 227 | G - - A + A T T + C ++ + G+ T |
| it 204 | X2* | 189 223 255 278 | 146 153 195 225 227 | G - - - G + A T TA + ++ + G+T T |
| gm 30 | X2* | 189 223 256A 266 274 278 304 (390) | 195 | G +- ATC+ T C T+ C + G+ GT |
| tu 83 | X2* | 189 223 266 278 | 153 195 225 | G+- - + - G + A T A + + A++ + G+T T |
| tu 131 | X2* | 189 223 266 278 | 153 195 225 226 | G - - + - GA + A T TA + CA++ + G+T T + |
| le 376 | X2* | 189 223 274 278 | 195 | G + TC+ A + + G+ G |
| gr 70 | X2* | 189 223 278 | 153 195 | G+- + - G + AC T A + + A+ + G+T T |
| alb 128 | X2* | 189 223 278 | 143 195 225 227 | G+--- - GA C+ A A + + ++-+ +T T + |
| ir 241 | X2* | 189 223 278 | 153 195 | G - + - G A C + +T T |
| tu 265 | X2* | 189 223 278 | 153 195 | G+- + - GA + A AC A + A + +T T |
| ng 144 | X2* | 189 223 278 | 153 195 198 225 | G - - - A + + T |
| tu 212 | X2* | 189 223 278 | 153 195 225 | G+- - + - GA + A T TA + + A++ + G+T T + |
| ir 267 | X2* | 189 223 278 | 153 195 225 226 | G - + - A TAC + +T T |
| fr 53 | X2* | 189 223 278 | 16527 153 195 | G+- + GA + A + A + +T T |
| ir 254 | X2* | 189 223 278 | 195 | G - + - G T A A CA + G+T T |
| le 547 | X2* | 189 223 278 | 195 | G - - G A C + G+T T |
| le 550 | X2* | 189 223 278 | 195 | G - A C + + T |
| tu 264 | X2* | 189 223 278 | 195 | G -- + - GA + A T A C T+ C + G+T GT |
| jo 992 | X2* | 189 223 278 293 | 146 153 | G - G -TG T A A T A CC +CATG+TGAAT TTT |
| le 936 | X2* | 189 223 278 311 357 | 153 195 225 227 | G - - G + +T T |
| ad 40 | X2* | 189 223 278 311G | 153 195 198 249d | - - G A + +T T |
| sy 289 | X2* | 189 223 278 335 | 195 | G -- + - + A T C T+ CA + + GT |
| kw 148 | X2* | 189 223 278 344 | 153 195 | G+- + - + A + + A + + T |
| sa 844 | X2* | 189 223 278 344 | 153 195 225 | G - + - T A + G+ T |
| it 154 | X2* | 189 223 278 344 | 153 195 225 | G+- - + - G + A TA + CA + + +T T |

(*continued*)

## Table 2 (continued)

| Sequence[c] | Haplogroup | Control Region Sequence[a] HVS-I[d] | HVS-II[e] | Coding Region Sequence and RFLP Variation[b] lckg w k we e e j u ygmt s i |
|---|---|---|---|---|
| it 4327 | X2* | 189 223 278 344 | 16519 153 195 225 226 | G − + − G + A T TAC + A+ + G+T T + |
| os 107 | X2* | 189 261 278 | 153 195 225 | G − − G + A T TA + + + +T T |
| kw 172 | X2* | 189 261 278 | 153 195 225 | G − − + − GA + A A + A + + +T T |
| le 1056 | X2* | 189 266 278 | 64 153 195 225 226 | G − − G + +T T |
| kw 237 | X2* | 189 278 | 153 195 225 | G+− − + − G + A T TA + + A + + +T T |
| it 8 | X2* | 189 278 | 153 195 225 | G+− − + − G + TA + C + + + T |
| jo 761 | X2* | 69 157 186 189 223 278 352 | 195 295 | G − − G C + +T T |
| ir 404 | X2* | 86 162 179 189 223 278 357 | 153 195 225 | G+− − + − G + A T TA + + G + + +T T |
| ng 68 | X2* | 93 189 223 248 278 | 153 195 225 | G − − G A + +T T |
| tu 145 | X2* | 93 189 223 278 | 153 195 225 | G+− − + − G + A T TA + + A++ + G+T T |
| ho 78 | X2b | 129 189 223 278 | 153 195 225 226 | A − +AA + + − + |
| mo E26 | X2b | 173 189 223 278 | 146 153 195 225 226 | − − − + |
| tu 42 | X2b | 175 189 223 278 | 153 195 225 226 | G − A ++AA −T+T C + − + − |
| it 13 | X2b | 179 181 189 223 278 | 153 195 225 | G − − A ++ A − +T C + ++ − + − |
| ev 4107 | X2b | 189 223 234 278 | 153 195 225 226 | G − − + |
| alg R34 | X2b | 189 223 239 278 301 | 153 195 225 226 (357) | − − − + |
| mo B45 | X2b | 189 223 248 278 294 | 153 195 225 226 | − − − + |
| cz 63 | X2b | 189 223 278 | 150 153 195 225 226 | G − A A ++AA −T+T C + + − G+ − |
| es 15 | X2b | 189 223 278 | 153 195 225 226 | G − ++AG −T+ + + − G+ − |
| sy 184 | X2b | 189 223 278 | 153 195 225 226 | G − A ++AA −T+T C + − + − |
| alg R47 | X2b | 189 223 278 316T | 153 195 225 | − − − + |
| hu 16 | X2b | 189 278 355 365 | 153 195 225 226 275 | G − A ++ A − +T C − + − |
| ukr | X2b | 86 189 223 278 | 153 195 198 225 | G − − G+ |
| mo D17 | X2b | 93 189 223 278 | 153 195 225 226 | − − − + |
| tu 77 | X2b | 93 189 223 278 292 | 146 153 195 225 226 | G − A ++AA −T+T C + + − G+ − |
| sw 44 | X2c | 108 189 223 255 278 | 153 195 225 227 | G − C + + |
| es 39 | X2c | 108 189 223 255 278 298 | 153 195 225 227 | G+−−−G+A−TGATC++AAAC+T+CAC+TCAT+TCA++−+CATG+TGAATGTTT+ |
| sw 18 | X2c | 108 189 223 255 278 311 | 153 195 225 227 | G − C G+ |
| kir 16m | X2c | 189 223 255 278 | 153 195 198 225 | G − C + + |
| ng 250 | X2c | 189 223 255 278 | 153 195 198 225 | G − C + + |
| es 314 | X2c | 189 223 255 278 | 153 195 225 | G+− − + A T C + + ++ + G+ |
| alb 124 | X2c | 189 223 255 278 300 311 | 153 195 225 | G − − C+ A C + ++ + + |
| arb 60 | X2c | 93 189 223 249 255 278 | 153 195 225 227 | G − C + + |
| tu 144 | X2d | 111 189 223 278 | 185 195 | A − A + G C + + − G+ G |
| ir 347 | X2d | 133 189 223 278 | 195 204 | A − G + G + + − − + |
| mo C16 | X2d | 189 192 223 278 | 195 207 | A − − − + |
| it 350 | X2d | 189 192 223 278 | 195 207 | A − A + G +C+ − − + |
| it 80 | X2d | 189 215 223 278 | 195 | G − A + G C + − + |
| it 82 | X2d | 189 215 223 278 | 195 | G − A + C + − + |
| it 101 | X2d | 189 223 278 | 195 204 207 | G − G C+ G +C+ − + − + |
| ad 37 | X2d | 189 223 278 | 195 204 207 | G − G − − + |
| kr 71 | X2d | 189 223 278 | 195 204 207 | G − G − − + |
| go 57 | X2d | 189 223 278 | 204 207 | A − G + G +C+ − − + |
| ab 21 | X2d | 189 223 278 295 | 195 204 207 | G − G − − + |
| ho 83 | X2d | 189 223 278 299 | 195 | G − G + G C − + − G+ |
| tu 162 | X2d | 189 223 278 326 | 195 204 207 | A − G + G +C+ − + − G+ |
| jo 958 | X2d | 189 223 278 352 | 195 | G − G C + C − + T |
| ku 31 | X2d | 92 185 189 223 278 | 195 204 207 | G − G − − + |
| ar 31 | X2d | 93 189 223 278 | 195 204 207 | A − G C+ G +C+ − + − G+ |
| ku 26 | X2e | 124 189A 223 278 | 153 195 225 | G − C + + GC |
| ku 14 | X2e | 129 189 223 265 278 | 153 195 | G − T + + GC |
| ku 39 | X2e | 172 189 223 278 | 153 195 | G − T + + GC |
| go 3 | X2e | 189 223 234 250 278 | 153 195 | G+− A C+ T + + T + + G+ GC |
| go 78 | X2e | 189 223 264 278 | 153 195 | G+− A C+ A C T + + T + + + GC |
| go 28 | X2e | 189 223 278 | 153 195 | G − + A C T + + GC |
| eg 596 | X2e | 189 223 278 | 153 195 | G − A T + + GC |
| kr 18 | X2e | 189 223 278 | 153 195 | G − T + + GC |
| ir 184 | X2e | 189 223 278 | 153 195 | G − T + + GC |
| ar 118 | X2e | 189 223 278 | 153 195 | G − A C+ AC T + + C |
| alt 81 | X2e | 189 223 278 | 153 195 | G − C T T + GC |
| alt 161 | X2e | 189 223 278 | 153 195 | G − C T T + GC |
| alt 171 | X2e | 189 223 278 | 153 195 | G − C T T + G |
| alt 188 | X2e | 189 223 278 | 153 195 | − C T T + GC |
| alt 208 | X2e | 189 223 278 | 153 195 | G − C T T + |
| gm 168 | X2e | 189 223 278 | 153 195 | G+− A C+ C T + + T + + G+ GC |
| gm 42 | X2e | 189 223 278 | 153 195 | G − + T + + GC |
| gm 6 | X2e | 189 223 278 | 153 195 | G −− C+ AC C T T + + GC |
| go 111 | X2e | 189 223 278 | 153 195 | G − + A C T T + + GC |
| go 60 | X2e | 189 223 278 | 153 195 | G − C+ C T + + GC |
| alt 17 | X2e | 189 223 278 | 153 195 | G − T + + GC |

(*continued*)

**Table 2 (continued)**

| | | Control Region Sequence[a] | | Coding Region Sequence and RFLP Variation[b] |
|---|---|---|---|---|
| | | | | 11111111111111111111111111111 |
| | | | | 11223334455566667778888890001111222333333344444445555 |
| | | | | 74758556791142337557355798279008903603778894555678366 9 |
| | | | | 01133352304822689789907013592287989036085766068866157 2 |
| | | | | 93571114277601931052132530404007247816457965207771042 5 |
| Sequence[c] | Haplogroup | HVS-I[d] | HVS-II[e] | lckg w k we e e j u ygmt s i |
| go 6 | X2e | 189 223 278 | 153 195 | G – C+ A C T + T + + G+ GC |
| gm 120 | X2e | 189 223 278 | 153 195 | G – C+ C T + + GC |
| alt 43 | X2e | 189 223 278 | 153 195 | – C T T + GC |
| alt 16 | X2e | 189 223 278 | 153 195 215 | – C T T + GC |
| es 167 | X2e | 189 223 278 | 153 195 225 | G – + T C T + + GC |
| es 297 | X2e | 189 223 278 | 153 195 225 | G+- – A ++ A + +T + T + + + GC + |
| ir 304 | X2e | 189 223 278 | 195 207 | G – A T + G+ AC |
| cz 55 | X2e | 189 223 278 316 | 195 | G –– A + A T C T+ T + + G+ AC |
| gm 164 | X2e | 189 223 278 316 | 195 | G – C+ A T C C+ T + + G+ AC |
| gm 66 | X2e | 189 223 278 316 | 195 | G+---G+A-TGATC++AAAC+T+TAC+TTAC+TTA++-+CATG+TGAATACTT+ |
| gm 96 | X2e | 189 223 278 316 | 195 | G – + T C C+ T + + AC |
| ir 81 | X2e | 189 223 278 316 | 195 | G – A T + + AC |
| kw 225 | X2e | 189 223 278 316 319 368 | 195 207 | G – A + A C T+ T + + AC |
| sa 1159 | X2e | 189 223 278 325 | 152 153 195 | G – T + + C |
| alt 61 | X2e | 189 223 278 380 | 153 195 | G – T + G+ GC |
| gr 5 | X2e | 189A 223 257 271 278 | 16533 16535 16537 153 195 225 | G+- – A + T T + + C ++ + G+ GC |
| gr 95 | X2e | 189A 223 278 | 195 225 | G+--- A + + C T+ C ++-+ G+ GC |
| it 56 | X2e | 93 189 223 265 278 | 153 195 | A – + A T + T + + G+ C |
| it 57 | X2e | 93 189 223 265 278 | 153 195 | A – + T + + GC |
| gm 156 | X2e | 93 189 223 278 | 153 195 | G+- A C+ A C T + + T + + G+ GC |
| tu 114 | X2e | 93 189 223 278 324 325 | 153 195 196 | G – A + A + T + + GC |
| go 105 | X2f | 104 189 223 278 | 153 195 257 | G – C+ T + + |
| go 110 | X2f | 104 189 223 278 | 153 195 257 | G – A C+ A AT T + + + G+ |
| go 41 | X2f | 104 189 223 278 | 153 195 257 | G+---G+A-TGATC++AAAT+T+TAC+TCAT+TCA++-+CATG+TGAATGTTT+ |
| go 17 | X2f | 129 140 189 223 278 | 153 195 257 | G – C+ A GC T + + + G+ |
| gm 172 | X2f | 136 189 223 278 | 153 195 257 | G – A + AC T + + + G+ |
| gm 140 | X2f | 189 223 278 | 153 195 257 | G – + T + + |
| le 208 | X2f | 189 223 278 | 153 195 257 | G – + A AC T + + + |
| go 133 | X2f | 189 223 278 | 153 195 257 | G – + AT + + |
| go 24 | X2f | 189 223 278 | 153 195 257 | G – + A T + + |
| tu 344 | X2f | 189 223 278 | 153 195 257 | G – A + A AC + + + |
| td 41m | X2f | 189 223 278 | 195 257 | G – C T C + + T |
| ar 101 | X2f | 189 223 278 | 195 257 | G –– C+ A AC T + C + + G+ T |
| ar 140 | X2f | 189 223 278 311 | 153 195 257 | G – + + |
| gm 143 | X2f | 86 189 223 278 | 153 195 257 | G – C+ A AT T + + + G+ |
| go 39 | X2f | 93 189 223 278 | 153 195 257 | G – A C+ A AC T + + + G+ |

[a] Two hypervariable segments of the control region were sequenced: HVS-I (16024-16383) and HVS-II (16518-310). Nucleotide change is specified for transversions; d = deletion. Length variation in the C stretch (16184-16193) is not shown. Mutations that were sequenced outside the specified range are shown in parentheses. Sequence analyses were performed using the Sanger dideoxy chain-termination method with the Amersham DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Pharmacia Biotech) on an ABI PRISM 377 DNA Sequencer. Sequences were alignedand analyzed with the Wisconsin Package (GCG).

[b] Restriction-endonuclease sites are indicated as follows: c = *Dde*I, e = *Hae*III, g = *Hin*fI, i = *Msp*I, j = *Mbo*I, k = *Rsa*I, l = *Taq*I, m = *Bam*HI, s = *Acc*I,t = *Bst*OI, u = *Mse*I, w = *Mbo*II, y = *Bfm*I. Haplogroup assignments are according to figure 2.

[c] Populations are indicated by the following letter code: ab = Abazin, ad = Adygei, alb = Albanian, alg = Algerian, alt = Altaian, ar = Armenian, arb = Arab from Uzbekistan, be = Bengali, cz = Czech, dr = Druze, eg = Egyptian, es = Estonian, et = Ethiopian, ev = Evenk, fr = French, gm and go = Georgian, gr = Greek, ho = Croat, hu = Hungarian, ir = Iranian, it = Italian, jo = Jordanian, kir = Kyrgyz, kr = Karachay, ku = Kumyk, kw = Kuwaiti, le = Lebanese, mo = Moroccan, ng = Nogay, os = Ossetian, sa = Saudi Arabian, sw = Swede, sy = Syrian, td = Tadjik, tu = Turk, ukr = Ukrainian.

[d] HVS-I nucleotide positions are −16000.

[e] All listed HVS-II sequences also harbored the transitions 16519, 73, and 263.

[f] Revised Cambridge reference sequence (Andrews et al. 1999).

Mediterranean Sea and the Red Sea (table 1). This subhaplogroup is subdivided into the two clades X1a and X1b, which are defined by two and five coding region mutations, respectively (fig. 2). Both clades share a recurrent transition at 146 in HVS-II. The coalescence time of the entire X1 subhaplogroup using HVS-I variation is 42,900 ± 18,100 YBP, whereas the coalescence time of the X1a clade is 17,900 ± 11,900 YBP.

Virtually all (97.2%) haplogroup X mtDNAs from the Near East, the South Caucasus, and Europe were found to belong to subhaplogroup X2, as did all (100%) of those from Siberia and Central Asia and some (36.8%) of those from North Africa (table 2). Thus, subhaplogroup X2 is characterized by a very wide geographic range but also by an infrequent occurrence. Indeed, it generally comprises <5% of the mtDNAs in West Eurasian and North African populations (table 1). Three exceptions include the Druze, the Georgians, and the Orkney Islanders, among whom the frequency of X2 reaches 11%, 8%, and 7%, respectively. The high frequencies of X2 in the Druze and the Orkney Islanders are combined with a low haplotype diversity (0.400 and 0.473, respectively), and the relatively high frequency in these populations is most likely due to genetic drift and

founder events. Overall, it appears that the populations of the Near East, the Caucasus, and Mediterranean Europe harbor subhaplogroup X2 at higher frequencies than those of northern and northeastern Europe ($P < .05$) and that X2 is rare in Eastern European as well as Central Asian, Siberian, and Indian populations and is virtually absent in the Finno-Ugric and Turkic-speaking people of the Volga-Ural region. Coalescence time estimates based on HVS-I and coding region variation—17,900 ± 2,900 YBP and 21,600 ± 4,000 YBP, respectively (figs. 1 and 2)—are consistent with the range expansion of X2 around or after the last glacial maximum (LGM). It is intriguing that the estimated coalescence time for X2 alone is very close to that obtained from HVS-I data for the entire haplogroup X (20,200 ± 3,100 YBP) (fig. 2). However, the latter is probably an underestimate due to both the higher proportion (>90%) of X2 mtDNAs included in the calculations and the fact that the HVS-I consensus sequence of X2 is completely identical to that of the overall haplogroup X.

Two-thirds of the subhaplogroup X2 sequences fall into the five clades X2b–X2f (fig. 2). Two sequences—one from Lebanon and one from Georgia—lacked the transition at np 1719, suggesting either the presence of an early X2 branch or a reversion at that nucleotide position. The sister groups X2b and X2c (X1 and X2, respectively, in the work of Herrnstadt et al. 2002) encompass one-third of the European sequences (excluding the samples from the North Caucasus). It is of interest that some North African sequences (from Morocco and Algeria) belong to X2b as well. Subhaplogroup X2b shows a diversity that is consistent with a postglacial population expansion in both West Eurasia and North Africa. Clades X2e and X2f encompass the majority (87.1%) of the sequences from the South Caucasus area and show coalescence times (12,000 ± 4,000 YBP and 10,800 ± 5,000 YBP, respectively) consistent with a Late Upper Paleolithic (LUP) origin and a subsequent spread in the region. We found significant differences between the haplogroup distribution between the North and the South Caucasian samples, a result that indicates a major geographical barrier between the two regions. The South Caucasian sample is enriched in mtDNAs belonging to clades X2e and X2f ($P < .01$), whereas the North Caucasian sample shows a higher proportion of sequences derived at nps 225 and 16248 ($P < .01$).

Clade X2e, defined by the synonymous substitution at 15310, encompasses all haplogroup X sequences in the Altaians (fig. 2). Among the nine Altaian X sequences, eight harbor the founder HVS-I motif of X2e, and seven of these eight also carry the HVS-II founder motif. As a result, a very low haplotype diversity of haplogroup X (0) in the Altaian region was obtained, making it significantly different from the haplotype diversities for haplogroups C and D (0.835 and 0.943, respectively) in the same region. Moreover, the nine Altaian mtDNAs do not harbor any nucleotide difference between nps 16090 and 16365. Therefore, under the assumption that these sequences are a random sample of the Altaian haplogroup X, an estimated $\rho$ value <0.33 ($P < .05$) was obtained. This value corresponds to a time depth of <6,700 years (Forster et al. 1996), and it would suggest that Altaians have acquired haplogroup X2 only relatively recently. This scenario is supported by the concomitant presence in the Altaians of a wide range of other West Eurasian haplogroups (H, J, I, T, U1, U4, and U5) that comprise ~27% of their mtDNAs. Indeed, any recent migration (for example, from the [southern] Caucasus region) that might have carried X2e mtDNA sequences to the Altai region would also explain the presence of the other West Eurasian mtDNA haplogroups in modern Altaians.

Further northeast of the Altai area, haplogroup X sequences were detected in the Tungusic-speaking Evenks, of the Podkamennaya Tunguska basin (Central Siberia). In contrast to the Altaians, the Evenks did not harbor any West Eurasian mtDNA haplogroups other than X. However, neither of the two Evenk X haplotypes showed mutations characteristic of the Native American clade X2a. Instead, one sequence was a member of X2b and the other of X2* (fig. 2). Thus, one possible scenario is that several X haplotypes arrived in Siberia from western Asia during the Palaeolithic, but only X2a crossed Beringia and survived in modern Native Americans. Alternatively, the presence of two phylogenetically different haplogroup X mtDNA sequences in this particular subpopulation of Evenks might be due to recent gene flow.

The Native American–specific clade X2a appears to be defined by five mutations, three in the coding region (8913, 12397, and 14502) and two in the control region (200 and 16213) (fig. 1). The transition at np 200 was seen in virtually all previously analyzed Native American haplogroup X mtDNAs, whereas the transition at np 16213 was absent in some of the Ojibwa described by Brown et al. (1998). We surveyed our Old World haplogroup X mtDNAs for the five diagnostic X2a mutations (table 2) and found a match only for the transition at np 12397 in a single X2* sequence from Iran. In a parsimony tree, this Iranian mtDNA would share a common ancestor with the Native American clade (fig. 2). Yet, the nonsynonymous substitution at np 12397 converting threonine to alanine cannot be regarded a conservative marker, as it has also been observed in two different phylogenetic contexts—in haplogroups J1 and L3e—among 794 complete mtDNA sequences (Finnilä et al. 2001; Maca-Meyer et al. 2001; Herrnstadt et al. 2002). Therefore, the scenario that the threonine to alanine change in the haplogroup X background is indeed due to recurrence appears most plausible.

These findings leave unanswered the question of the geographic source of Native American X2a in the Old World, although our analysis provides new clues about the time of the arrival of haplogroup X in the Americas. Indeed, if we assume that the two complete Native American X sequences (from one Navajo and one Ojibwa) began to diverge while their common ancestor was already in the Americas, we obtain a coalescence time of 18,000 ± 6,800 YBP, implying an arrival time not later than 11,000 YBP.

The results of this study point to the following conclusions. First, haplogroup X variation is completely captured by two ancient clades that display distinctive phylogeographic patterns—X1 is largely restricted to North and East Africa, whereas X2 is spread widely throughout West Eurasia. Second, it is apparent that the Native American haplogroup X mtDNAs derive from X2 by a unique combination of five mutations. Third, the few Altaian (Derenko et al. 2001) and Siberian haplogroup X lineages are not related to the Native American cluster, and they are more likely explained by recent gene flow from Europe or from West Asia. Fourth, the split between "African" X1 and "Eurasian" X2 subhaplogroups of X is phylogenetically as deep as that within the branches of haplogroup U that also differ profoundly in their phylogeography. Thus, subhaplogroup U6 is largely restricted to North Africa (as X1), whereas subhaplogroup U5 is widespread in West Eurasia (as X2). The phylogeographic patterns and the coalescence times that we obtained here suggest that the basic phylogenetic structures of the mtDNA haplogroups in West Eurasia and North Africa are as ancient as the beginning of the spread of anatomically modern humans in this region. Finally, phylogeography of the subclades of haplogroup X suggests that the Near East is the likely geographical source for the spread of subhaplogroup X2, and the associated population dispersal occurred around, or after, the LGM when the climate ameliorated. The presence of a daughter clade in northern Native Americans testifies to the range of this population expansion.

## Acknowledgments

We thank Peter Underhill and Ripan Malhi, for advice, and Jaan Lind and Ille Hilpus, for technical assistance. A.T. received support from Telethon Italy (grant E.0890), the Italian Ministry of the University (Progetti Ricerca Interesse Nazionale 2002), and Progetto CNR-MIUR Genomica Funzionale-Legge 449/97. The research of R.V. was supported by Estonian basic research grant 514 and EC DG Research grant ICA1CT20070006, and T.K. was supported by Estonian basic research grant 5574. The work of L.O. was supported by the Russian Foundation Basic Research (RFBR) PTTH98-04-49626-YU, and E.K. was supported by the Russian Foundation Basic Research (RFBR) 01-0448487. N.H. acknowledges research support from the National Science Foundation.

## Electronic-Database Information

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

Estonian Biocentre, http://www.ebc.ee/EVOLUTSIOON/mtDNA-X/ (for mtDNA sequence data)
Shareware Phylogenetic Network Software, http://www.fluxus-engineering.com/sharenet.htm (for Network 3111)

## References

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

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

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Bandelt H-J, Herrnstadt C, Yao Y-G, Kong Q-P, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang Y-P. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet (in press)

Bermisheva M, Tambets K, Villems R, Khusnutdinova E (2002) Diversity of mitochondrial DNA haplotypes in ethnic populations of the Volga-Ural region of Russia. Mol Biol (Mosk) 36:990–1001

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

Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC (1998) mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? Am J Hum Genet 63:1852–1861

Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. Nat Genet 33:266–275

Chen YS, 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

Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC (1996) Genetic diversity in the Iberian Peninsula deter-

mined from mitochondrial sequence analysis. Ann Hum Genet 60:331–350

Crespillo M, Luque JA, Paredes M, Fernandez R, Ramirez E, Valverde JL (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. Int J Legal Med 114: 130–132

Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI (2002) Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am J Hum Genet 70:1009–1014

Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwka D, Zakharov IA (2001) The presence of mitochondrial haplogroup X in Altaians from South Siberia. Am J Hum Genet 69:237–241

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

Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. Int J Legal Med 113: 89–97

Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475–1484

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

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

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

Helgason A, Sigurdadottir S, Gulcher J, Ward R, Stefanson K (2000) mtDNA and the origins of the Icelanders: deciphering signals of recent population history. Am J Hum Genet 66: 999–1016

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, 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

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz KD (1997) Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. Hum Mol Genet 6: 1835–1846

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331–1334

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H-V, Stepanov V, Gölge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet 72:313–332

Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737–1751

Levin BC, Cheng H, Reeder DJ (1999) A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection. Genomics 55:135–146

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

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13

Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-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

Malhi RS, Schultz BA, Smith DG (2001) Distribution of mitochondrial DNA lineages among Native American tribes of Northeastern North America. Hum Biol 73:17–55

Malhi RS, Smith DG (2002) Brief communication: Haplogroup X confirmed in prehistoric North America. Am J Phys Anthropol 119:84–86

Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miscicka-Sliwka D (2002) Mitochondrial DNA variability in Poles and Russians. Ann Hum Genet 66:261–283

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

Mogentale-Profizi N, Chollet L, Stevanovitch A, Dubut V, Poggi C, Pradie MP, Spadoni JL, Gilles A, Beraud-Colomb E (2001) Mitochondrial DNA sequence diversity in two groups of Italian Veneto speakers from Veneto. Ann Hum Genet 65:153–166

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, pp 145–163

Opdal SH, Rognum TO, Vege A, Stave AK, Dupuy BM, Egeland T (1998) Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. Acta Paediatr 87:1039–1044

Orekhov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov P, Yankovsky N (1999) Mitochondrial DNA sequence diversity in Russians. FEBS Letters 445:197–201

Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111:124–132

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A (2001) Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Ann Hum Genet 65:439–458

Pereira L, Prata MJ, Amorim A (2000) Diversity of mtDNA

lineages in Portugal: not a genetic edge of European variation. Ann Hum Genet 64:491–506

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

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

Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. Ann Hum Genet 62:531–550

Renfrew C (2000) Archaeogenetics: towards a population prehistory of Europe. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research, Cambridge, pp 3–11

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203

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

Rousselet F, Mangin P (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. Int J Legal Med 111:292–298

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

Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sanchez-Diz P, Macaulay V, Carracedo A (2002) The making of the African mtDNA landscape. Am J Hum Genet 71: 1082–1111

Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea–Bering Sea region during the Neolithic. Am J Phys Anthropol 108:1–39

Scozzari R, Cruciani F, Santolamazza P, Sellitto D, Cole DEC, Rubin LA, Labuda D, Marini E, Succa V, Vona G, Torroni A (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

Smith DG, Malhi RS, Eshleman J, Lorenz JG, Kaestle FA (1999) Distribution of mtDNA haplogroup X among Native North Americans. Am J Phys Anthropol 110:271–284

Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. Am J Hum Genet 63:1473–1491

Stone AC, Stoneking M (1999) Analysis of ancient DNA from a prehistoric Amerindian cemetery. Philos Trans R Soc Lond B Biol Sci 354:153–159

Tagliabracci A, Turchi C, Buscemi L, Sassaroli C (2001) Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 114:224–228

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, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994) 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, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Ward RH, Frazier B, Dew-Jager K, Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720–8274

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