

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 Ilija 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 Villems¹

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

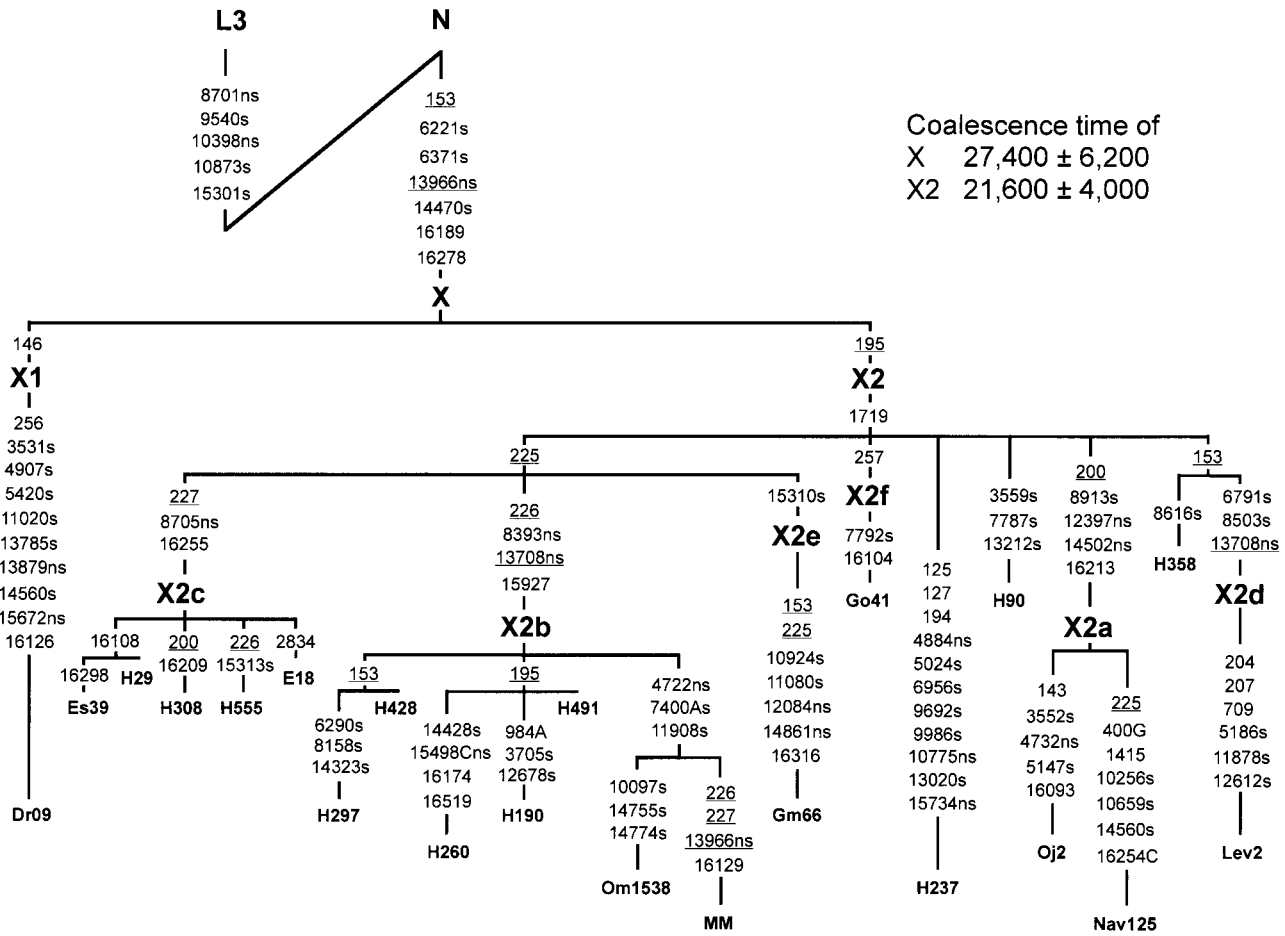
mtDNA and the nonrecombining part of the Y chromosome are widely used in archaeogenetic studies (Renfrew 2000; Cavalli-Sforza and Feldman 2003) that aim

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

© 2003 by The American Society of Human Genetics. All rights reserved. 0002-9297/2003/7305-0020\$15.00



Coalescence time of
X 27,400 ± 6,200
X2 21,600 ± 4,000

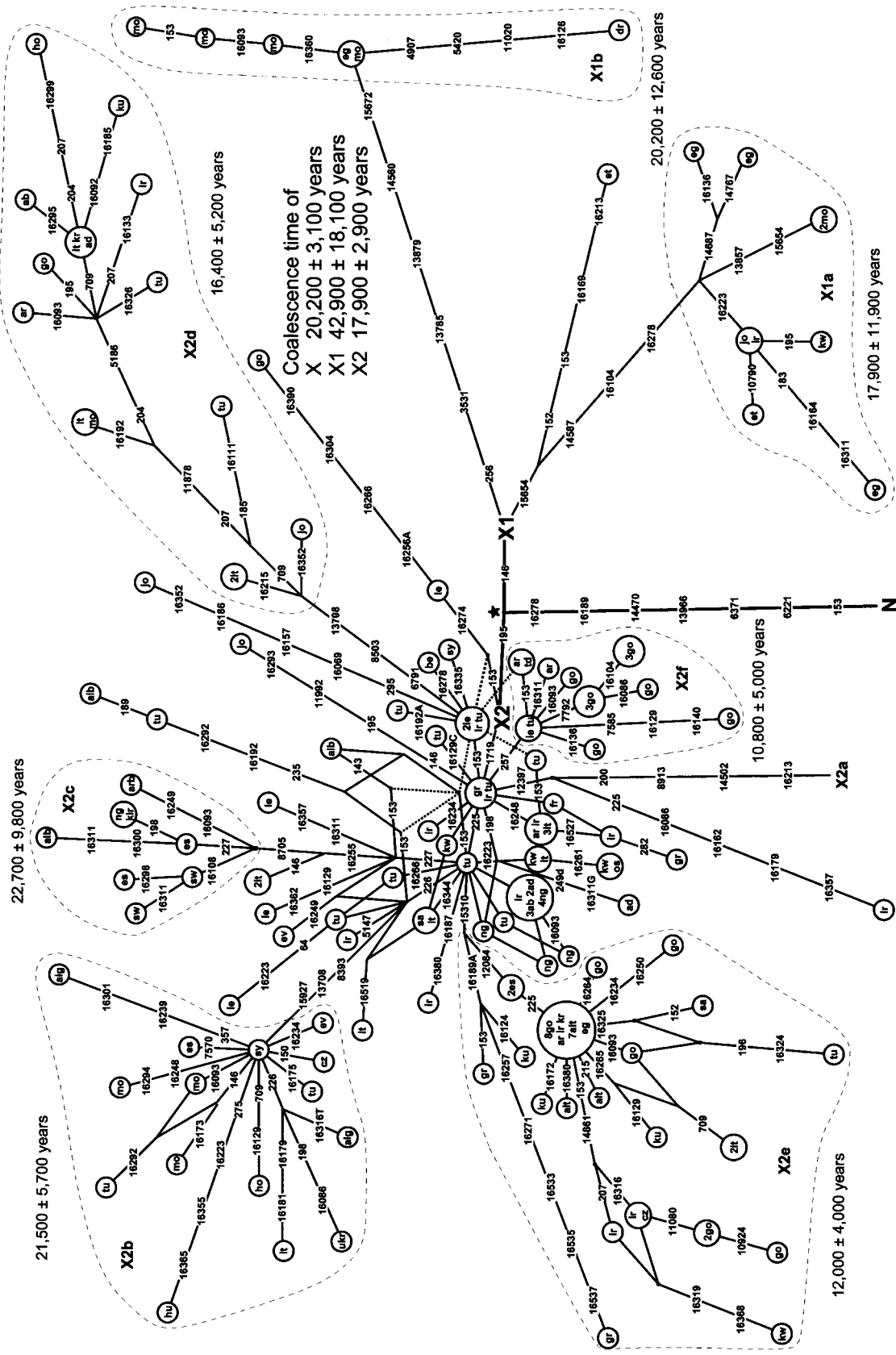
Figure 1 Maximum parsimony phylogenetic tree of complete mtDNA sequences belonging to haplogroup X. Five mtDNAs were selected for complete sequence analysis in the course of the present study (Dr09, Es39, Gm66, Go41, and Om1538), 11 coding region sequences were from the work of Herrnstadt et al. (2002), 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-



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 ρ , 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 ρ 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]	
Portuguese 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 [.0–2.9]	
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		X2	
		Mean Frequency [95% CR] (%)	Diversity	Mean Frequency [95% CR] (%)	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
Adygeys 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: ¹present study, ²Bermisheva et al. (2002), ³Bertranpetit et al. (1995), ⁴Corte-Real et al. (1996), ⁵Crespillo et al. (2000), ⁶Derbeneva et al. (2002), ⁷Derenko et al. (2001), ⁸Dimo-Simonin et al. (2000), ⁹Di Rienzo and Wilson (1991), ¹⁰Finnilä et al. (2001), ¹¹Helgason et al. (2000), ¹²Helgason et al. (2001), ¹³Hofmann et al. (1997), ¹⁴Kivisild et al. (1999), ¹⁵Kivisild et al. (2003), ¹⁶Lutz et al. (1998), ¹⁷Macaulay et al. (1999), ¹⁸Malyarchuk et al. (2002), ¹⁹Mogentale-Profizi et al. (2001), ²⁰Opdal et al. (1998), ²¹Orekhov et al. (1999), ²²Parson et al. (1998), ²³Pereira et al. (2000), ²⁴Pereira et al. (2001), ²⁵Pfeiffer et al. (1999), ²⁶Piercy et al. (1993), ²⁷Rando et al. (1998), ²⁸Richards et al. (1996), ²⁹Richards et al. (2000), ³⁰Rousset and Mangin (1998), ³¹Saillard et al. (2000), ³²Salas et al. (2002), ³³Tagliabracci 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.

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— $17,900 \pm 2,900$ YBP and $21,600 \pm 4,000$ 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 ($20,200 \pm 3,100$ 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 ($12,000 \pm 4,000$ YBP and $10,800 \pm 5,000$ 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 di-

versities 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 ρ 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 $\sim 27\%$ 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 $18,000 \pm 6,800$ 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 Foun-

ation 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, Stefansson 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 ME, 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, Cavall-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, Bonnè-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, Ege-land 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* 88:8720–8724
- Watson E, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691–704