Phylogeographic Differentiation of Mitochondrial DNA in Han Chinese

Yong-Gang Yao,[1] Qing-Peng Kong,[1] Hans-Jürgen Bandelt,[2] Toomas Kivisild,[3] and Ya-Ping Zhang[1]

[1]Laboratory of Molecular Evolution and Genome Diversity, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; [2]Fachbereich Mathematik, Universität Hamburg, Hamburg; [3]Department of Evolutional Biology, Institute of Molecular and Cell Biology, Tartu University and Estonian Biocentre, Tartu, Estonia

To characterize the mitochondrial DNA (mtDNA) variation in Han Chinese from several provinces of China, we have sequenced the two hypervariable segments of the control region and the segment spanning nucleotide positions 10171–10659 of the coding region, and we have identified a number of specific coding-region mutations by direct sequencing or restriction-fragment–length–polymorphism tests. This allows us to define new haplogroups (clades of the mtDNA phylogeny) and to dissect the Han mtDNA pool on a phylogenetic basis, which is a prerequisite for any fine-grained phylogeographic analysis, the interpretation of ancient mtDNA, or future complete mtDNA sequencing efforts. Some of the haplogroups under study differ considerably in frequencies across different provinces. The southernmost provinces show more pronounced contrasts in their regional Han mtDNA pools than the central and northern provinces. These and other features of the geographical distribution of the mtDNA haplogroups observed in the Han Chinese make an initial Paleolithic colonization from south to north plausible but would suggest subsequent migration events in China that mainly proceeded from north to south and east to west. Lumping together all regional Han mtDNA pools into one fictive general mtDNA pool or choosing one or two regional Han populations to represent all Han Chinese is inappropriate for prehistoric considerations as well as for forensic purposes or medical disease studies.

Introduction

The Han people constitute China's and the world's largest ethnic group, making up ~93% of the country's population and nearly 20% of all humankind. The formation of the Han people was a process of continuous expansion by integration of numerous tribes or ethnic groups; it began with the ancient Huaxia tribe, which was formed during the 21st–8th centuries B.C. Although the Han people are now spread all over the country, the highest population concentrations are in the basins of the Yellow River, the Yangtze River, and the Zhujiang River and on the Songhuajiang-Liaohe plain in northeast China, as well as on the islands of Taiwan and Hainan (Du and Yip 1993; Ge et al. 1997). The migration of Han people to provinces such as Xinjiang and Yunnan occurred relatively recently, having started mainly ~100–600 years ago, and was caused by war, plague, and other reasons (Ge et al. 1997). Do these populations bear some genetic differences from those from the his-

torical Han regions, such as Wuhan and Qingdao? To what extent can the genetic data reflect those recent migration events? A prerequisite for answering these and more-specific questions with genetic data is a thorough screening of mtDNA and Y-chromosome variation across China.

Hitherto, mtDNA from Han Chinese has been poorly sampled and understood in its variation, with only limited data available from Guangdong (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data), Hong Kong (Betty et al. 1996), Shanghai (Nishimaki et al. 1999), Shandong (Wang et al. 2000), and Taiwan (Horai et al. 1996; Tsai et al. 2001). Moreover, previous genetic studies of the Chinese populations either grouped the various regional Han populations into "Southern Han" and "Northern Han" (Su et al. 1999, 2000) or simply used Han samples from only one or two regions to stand for all Han Chinese (Horai et al. 1996; Hou et al. 2001; Karafet et al. 2001), thereby neglecting potential geographic differences between different Han populations, as well as migrations between north and south. Although genetic contrast between southern and northern populations has been claimed in classical genetic markers (e.g., Zhao and Lee 1989; Chen et al. 1993; Du et al. 1998), dermatoglyphic data (Zhang et al. 1998), archaeological assemblages (Wu et al. 1989), as well as in nuclear microsatellites (Chu et al. 1998) and Y-chromosome sin-

gle-nucleotide polymorphism (SNP) data (Su et al. 1999; Karafet et al. 2001), no detailed mtDNA study has been performed to substantiate this claim. Chu et al. (1998) and Su et al. (1999) also argued for a southern origin of northern populations, whereas Ding et al. (2000) emphasized that the regional genetic difference observed in the principal-component (PC) maps of mtDNA, nuclear short tandem repeats (STRs), and Y-chromosome SNPs might be more properly explained by a simple model of isolation by distance (IBD). Given the large census size of the Han people, the complexity of the migration events, and these hotly debated issues, it is necessary to gather detailed information about the regional Han populations.

To take full advantage of a uniparental marker system, such as mtDNA, one needs a sufficiently resolved phylogeny that is not overly blurred by recurrent mutations. Because the two hypervariable segments (HVS-I and HVS-II) alone—although useful for forensic purposes—cannot support a very reliable estimate of the mtDNA phylogeny (Bandelt et al. 2000), we opted for sequencing one stretch of the coding region (10171–10659) as well, which turned out to be highly informative for East Asian mtDNAs. Another segment (14055–14590) was sequenced in a few samples, helping to define four haplogroups. In addition, a number of further sites relevant for Eurasian mtDNAs (Macaulay et al. 1999; Schurr et al. 1999; T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data) were checked either by direct sequencing or through RFLP testing in specific mtDNAs.

Material and Methods

Sampling

From six provinces in China, 263 unrelated Han individuals were analyzed: 43 from Kunming, Yunnan; 42 from Wuhan, Hubei; 50 from Qingdao, Shandong; 47 from Yili, Xinjiang; 51 from Fengcheng, Liaoning; and 30 from Zhanjiang, Guangdong (see fig. 1 for sample locations). The maternal pedigrees (unrelated through at least three generations) of all individuals were ascertained before sampling. Except for 17 samples from Xinjiang, all subjects were able to confirm that the birthplace of their maternal grandmothers was in the same province.

Previously published Han mtDNA data used here for comparison include 69 mtDNAs from Guangzhou, Guangdong (with HVS-I, HVS-II, and additional coding-region information; T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data), 20 mtDNAs from Hong Kong (HVS-I; Betty et al. 1996), 120 mtDNAs from Shanghai (HVS-

I; Nishimaki et al. 1999—however, these data are not fully reliable; see Bandelt et al. 2001), 155 Taiwanese mtDNAs (HVS-I and HVS-II; Tsai et al. 2001), and another 66 Taiwanese mtDNAs (HVS-I; Horai et al. 1996). Further, mtDNAs (HVS-I) from 78 patients with type 2 diabetes mellitus (Y.-G. Yao, P.-L. Geng, Q.-P. Kong, and Y.-P. Zhang, unpublished data) from Xining, Qinghai, who do not bear the 3243 A→G transition (a well-known pathogenic mutation), were included here. Fifty mtDNAs from Zibo, Shandong, represented by a 185-bp fragment of HVS-I (16194–16378; Wang et al. 2000), were tentatively taken into consideration.

Amplification and Sequencing of HVS-I, HVS-II, and Region 10171–10659

Genomic DNA was extracted from whole blood by standard phenol/chloroform methods. The sequences of HVS-I from position 16001 to 16497 (relative to the revised Cambridge reference sequence [CRS]; Andrews et al. 1999) were amplified and sequenced as described elsewhere (Yao et al. 2000a). For HVS-II, the primer pair L29 and H408 was used in amplification and sequencing. For the segment 10148–10659, which covers the tRNA^{Arg} gene (10405–10469) and parts of the ND3 (10059–10406) and ND4L (10470–10766) genes, we used primers L10170 and H10660 for amplification and sequencing (table 1). Since several segments of the same mtDNA had to be screened, care was taken to avoid artificial recombination caused by potential sample crossover; therefore, doubtful segments were resequenced.

Typing of Other Polymorphisms

First, those Han individuals who had not yet been screened for the mtDNA 9-bp deletion in the COII/tRNA^{Lys} intergenic region (Yao et al. 2000b) were analyzed as described in that study. Then, as for the typing of further coding-region polymorphisms in specific lineages, we took advantage of the phylogenetic analyses of Eurasian mtDNAs provided by Macaulay et al. (1999) and Kivisild and colleagues (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data), which employed coding-region information (mainly derived from Ozawa et al. 1991, 1995; Ikebe et al. 1995; Ingman et al. 2000). In each run, a few (random) controls were tested. Some (unexpected) mutations observed in the controls were then systematically screened in related mtDNAs, which eventually led to the identification of novel characteristic markers for some haplogroups. In total, 13 pairs of primers were designed for RFLP typing and coding-region sequencing, as listed (along with the PCR conditions) in table 1.

Figure 1   Geographic locations of the Han samples under study

Data Analyses

   The sequences were edited and aligned by the DNA-STAR software (DNASTAR, Inc.) and were compared with the revised CRS (Andrews et al. 1999). The length polymorphisms of the A and C stretches in 16180–16188 (triggered by the 16189 T→C substitution) were disregarded in the analyses. We adopted the classification tree proposed by Kivisild and colleagues (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data), but without high-lighting haplogroup E (which is still poorly described and apparently very rare in China) and subhaplogroups of A and Y. We then assigned the mtDNAs to the (nested) haplogroups according to HVS-I, HVS-II, and coding-region information, in such a way that each mtDNA was allocated to the most-derived (i.e., smallest) named haplogroup it belongs to. If the haplogroup has further named subhaplogroups, then (following Richards et al. 1998) a star is attached to the haplogroup name that refers to the mtDNA under consideration, to emphasize

that the haplogroup status of the mtDNA cannot be specified further (relative to the classification tree). Co-alescence times, along with standard deviations, were estimated according to the methods of Forster et al. (1996) and Saillard et al. (2000) for the major haplo-groups detected in the 332 mtDNAs (263 from this study and 69 from T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems [unpublished data]).

   Haplogroup frequencies were then computed for the regional Han mtDNA samples. To compare these hap-logroup profiles with those from the previously pub-lished Han HVS-I data sets (lacking coding-region in-formation), we classified the published mtDNAs in another, coarser scheme guided by HVS-I and HVS-II motifs and (near-)matching with the 332 Han mtDNAs. This necessarily precluded the finer subdivision of hap-logroup D4, the recognition of F2, and the distinction between M* and N*. The frequency vectors of the basal mtDNA profiles (which only record the frequencies of

Table 1

Primers for Amplification, Sequencing, and RFLP Analyses

Primer Pair	Locations in CRS	Annealing Temperature (°C)	Polymorphisms at/in
L29/H408	8–29/429–408	54	HVS-II
L394/H902	375–394/922–902	60	+663*Hae*III (663)
L2796/H3274	2777–2796/3293–3274	57	3010, 3206
L3179/H3674	3160–3179/3693–3674	59	+3391*Hae*III (3394)
L4499/H5099	4480–4499/5118–5099	60	+4831*Hha*I (4833), 4715
L4887/H5442	4866–4887/5461–5442	56	−5176 *Alu*I (5178A), 5231, 5417
L7356/H7805	7337–7356/7824–7805	57	−7598*Hha*I (7598, 7600)
L8215/H8297	8196–8215/8316–8297	57	9-bp deletion
L9794/H10164	9774–9794/10181–10164	60	+9824 *Hin*fI (9824)
L10170/H10660	10147–10170/10679–10660	59	10171–10659
L11338/H11944	11319–11338/11963–11944	53	11719
L12334/H12878	12315–12334/12897–12878	57	12705, 12358, 12372
L14054/H14591	14035–14054/14610–14591	57	14178, 14308, 14318, 14470
L14575/H15086	14556–14575/15105–15086	57	14766
L15391/H16048	15372–15391/16067–16048	58	15487T, 15784
L15996/H16498	15975–15996/16517–16498	60	HVS-I

NOTE.—PCR conditions were 94°C for 2 min, for denaturation; 94°C for 40 s; annealing temperature shown for 1 min, for amplification; and 72 °C for 1 min, for 35–40 cycles; incubation at 72°C for 5 min.

the 10 basal haplogroups M7, M8, M9, M10, G2, D, A, N9, B, and R9 and the R* and M*/N* haplotypes in 13 Han samples) and the coarse mtDNA profiles were then subjected to PC analysis by the POPSTR program.

Results

Classification Tree

The sequence variation in HVS-I, HVS-II, region 10171–10659, and at further polymorphic sites detected in the 263 Han individuals is shown in table 2. The present data suggest two new subhaplogroups of M, which we name "M9" and "M10," as well as subhaplogroups of D4 (D4a and D4b), D5 (D5a), and F1 (F1c). Except for M10 and F1c, these new haplogroups each have at least one representative in the complete sequence database (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data). Altogether we distinguish 44 named nested haplogroups in the Han mtDNA classification tree. Figure 2 displays these haplogroups, along with the defining sites considered in this study. Almost all samples can be affiliated with proper haplogroups of macrohaplogroups M and N, with the exception of a few M* haplotypes and one N* haplotype that could not be specified further. Evidently, some of the M* haplotypes belong to specific clades (one with motif 16234-16290-125-127 and another with 318-326), the mutual relationships of which are not yet clear. Among the three R* haplotypes that could not be classified as B or R9, two bear a mutation

motif of 185-189-10398-16189-16311, similar to the motif of B5, but were found to lack the 9-bp deletion.

Two mtDNAs, one sampled in Yunnan and the other in Liaoning, are regarded as resulting from admixture from western Eurasia (via central Asia), as they belong to the west Eurasian haplogroups HV and T1 (Macaulay et al. 1999). Note that the sample from Guangzhou contains one W haplotype (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data).

The region 10171–10659 harbors numerous sites that support basal branches in the Asian mtDNA phylogeny. To begin with, site 10400 is one of the defining sites for macrohaplogroup M, whereas 10398 is one of the characteristic sites for macrohaplogroup N (Quintana-Murci et al. 1999). Back mutations at 10398—which occur occasionally (Macaulay et al. 1999)—are then characteristic of haplogroups Y and B5. The transition at 10397, which defines haplogroup D5, leads to the simultaneous loss of two prominent RFLP sites (10394 *Dde*I and 10397 *Alu*I; Bandelt et al. 1999). Site 10181 defines haplogroup D4b, and site 10410 defines a subclade of D4a that seems to be frequent in Japan (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data) but occurs only once in our Han data (from Liaoning). Subhaplogroup M7b2 of M7b can also be recognized by 10345. We define the new haplogroup M10 by sites 10646 (+10646 *Rsa*I) and 16311, although one should bear in mind that both sites are prone to recurrent mutations. Haplogroup R9, as defined by Kivisild and colleagues

Table 2

Sequence Variation in the 263 Chinese Han Individuals Analyzed in the Present Study

Haplogroup	Sample Number[a]	16001–16497 HVS-I (16000+)	30–407 HVS-II (73 and 263)	10171–10659 (10000+)	14055–14590 (14000+)	663e	3010	3206	3391e	4831f	5176a	7598f	9-bp	9820g	Other Polymorphisms[d]
M7b1	4YN285	*129 192* 223 *297*	*150* 152 *199* 315+C	398 *400*									2	+	
M7b1	YN156	051 *129 192* 223 *297*	*150 199* 315+C	398 *400*									2	+	*9824*
M7b1	YN288	*129 192* 223 271 *297*	*150 199* 309+C 315+C	398 *400*									2	+	
M7b1	XJ8415	*129 192* 223 *297* 301G 391	*150 199* 315+C	398 *400*									2	+	
M7b1	QD8160	*129* 188G *192* 223 *297*	*150 199* 315+C	ND									2	+	
M7b1	LN7717	093 *129 192 297*	*150 199* 309+C 315+C	398 *400*									2	+	
M7b2	QD8142	*129 189* 223 *297 298*	*150* 195A *199* 309+CC 315+C	345 398 *400*									2	+	
M7b2	XJ8450	*129* 183C *189* 223 *297 298*	*150 199* 204 228 309+CC 315+C	345 398 *400*	CRS								2	+	
M7b2	WH6953	*129* 183C *189* 223 *297 298* 325	*150 199* 204 309+C 315+C	345 398 *400*									2	+	
M7b2	XJ8422	*129* 183C *189* 223 *297 298* 325	*150 199* 309+C 315+C	345 398 *400*									2	+	
M7b	WH6242	129 189 223 293 *297*	*150* 195 198 *199* 315+C	398 *400*									2	+	
M7b	YN173	129 223 *297*	*150* 159 *199* 309+C 315+C	398 *400*							+		2	+	*9824*
M7c	GD7823	223 *295*	*146 199* 262 315+C	398 *400*						−	+	+	2	+	
M7c	2LN7711	223 *295*	*146 199* 315+C	398 *400*						−	+	+	2	+	
M7c	XJ8439n	223 *295*	*146 199* 309+C 315+C	398 *400*								+	2	+	
M7c	WH6939	*295* 319	*146 199* 315+C	398 *400*		−				−	+	+	2	+	
M7c	LN7605	223 *295* 296 319	*146 199* 309+C 315+C	398 *400*							+	+	2	+	
M7	YN250	172 223 311	*146 199* 315+C	398 *400*						−	+	+	2	+	
M7	XJ8438	172 223 287 311	093 *146* 315+C	398 *400*									2	+	*9824*
M7	WH6955	230 304	146A *199* 309+C 315+C	398 *400*									2	+	*9824* 9861
M8a	WH6952	223 *298 319*	309+C 315+C	398 *400*	470						+		2		
M8a	QD8159	*184* 223 278 *298 319*	234 309+C 315+C	398 *400* 646	470						+		2		
M8a	2XJ8417	*184* 223 293C *298 319*	152 309+C 315+C	398 *400*	470								2		*4715* 4769; 7357–7804=CRS; *15487T*
M8a	QD8150	*184* 223 293 *298 319*	152 200 315+C	398 *400*	470						+		2		
M8a	LN7715	*184* 209 223 293 *298* 311 *319*	207 309+CC 315+C	398 *400*	470						+		2		
M8a	WH6958	*184* 189 223 *298* 311 *319* 390 468 470 471 473	152 309+C 315+C	398 *400*	470						+		3		
M8a	WH6981	134 *184* 223 *298 319*	315+C	398 *400*	470						+		2		
M8a	2QD8120	134 *184* 223 *298 319*	309+C 315+C	398 *400*	470						+		2		7357–7804=CRS; *15487T*
M8a	2LN7597	*184* 223 *298 319* 400	152 315+C	398 *400*	470						+		2		
M8a	LN7590	*184* 223 *298 319*	315+C	398 *400*	94 *470*						+		2		
C	XJ8453	223 *327*	*249d* 309+C 315+C	398 *400*	318								2		
C	YN157	093 129 223 *298 327*	146 194 *249d* 315+C	398 *400*	*318*								2		
C	XJ8435	129 223 *298 327*	195 *249d* 309+C 315+C	398 *400*	*318*								2		*4715* 4769; 7357–7804=CRS; *15487T* 15968
C	XJ8418	223 243 297 *298* 324 *327*	146 *249d* 309+C 315+C	398 *400*	*318*						+		2	−	*4715* 4769; 7382; *15487T* 15515
C	YN177	092 183C 189 223 *298 327* 355	*249d* 309+C 315+C	398 *400*	318								2		
C	WH6938	189 *298 327*	234 *249d* 309+C 315+C	398 *400*	318								2		
C	GD7839n	183C 189 223 *298 327*	*249d* 309+C 315+C	398 *400*									2		
C	LN7710	217 223 *298* 311 *327*	146 *249d* 309+CC 315+C	398 *400*	318								2		*4715* 4769; 7357–7804=CRS; *15487T* 15930
Z	WH6249	*185* 223 *260 298* 302	152 *249d* 309+C 315+C	398 *400*									2		
Z	XJ8419	*185* 223 *260 298*	152 *249d* 309+C 315+C	208 398 *400*	CRS								2		*4715* 4769; 7357–7804=CRS; *15487T* 15784
Z	LN7620	*185* 223 *260 298*	152 *249d* 315+C	398 *400*									2		
Z	WH6943	136 *185* 223 *260 298*	152 *249d* 309+C 315+C	398 *400*									2		

Table 2 (continued)

Haplogroup	Sample Number[a]	Mutations in Region[b]				RFLP Polymorphisms[c]									Other Polymorphisms[d]
		16001–16497 HVS-I (16000+)	30–407 HVS-II (73 and 263)	10171–10659 (10000+)	14055–14590 (14000+)	663e	3010	3206	3391e	4831f	5176a	7598f	9-bp	9820g	
Z	WH6979	*185* 189 223 224 *260* 261 *298* 302	152 185 *249d* 309+C 315+C	398 *400*									2		*4715* 4769; 7357–7804=CRS; 15475 *15487T 15784* 15944d
M9	QD8125	223 *234* 291 *316*	*153* 309+C 315+C	398 *400*	*308* 417				+		+	+	2	−	
M9	XJ8452	145 223 *234 316*	*153* 315+C	398 *400*	*308*				+	−	+		2		
M9	LN7584	223 *234 316* 362	152 *153* 315+C	398 *400*	*308*				+	−	+		2		
M9	XJ8420	223 *234* 291 *316* 362	*153* 315+C	398 *400*	*308*				+	−	+	+	2		7357–7804=CRS
M9	LN7606	209 223 *234* 291 *316* 362	*153* 309+C 315+C	398 *400*	*308* 417				+	−	+	+	2		7719A
M9	QD8155	223 *234* 255 271 362	*153* 315+C	398 *400*	*308*				+	−	+		2		
M9	GD7822	223 *234* 311 362	146 217 309+C 315+C	398 *400*	*308*				+	−	+		2		
M10	LN7720	223 *311*	195 315+C 331	398 *400* 646							+	+	2	−	
M10	LN7596	066 086 092 223 *311*	152 315+C	398 *400* 646							+	+	2	−	
M10	YN163	093 129 223 *311* 357 497	309+CC 315+C	398 *400* 646					−	−	+	+	2	−	
M10	LN7593	093 129 193 223 311 357 497	146 152 309+C 315+C	398 *400* 646							+	+	2	−	
M10	QD8122	129 223 *311*	315+C	398 *400* 646							+	+	2	−	
M	GD7825	172 223 234 290 311	125 127 309+C 315+C	398 *400*					−	−	+	+	2	−	
M	GD7835n	223 234 287 290 362	125 127 128 309+C 315+C 318	398 *400*					−	−	+	+	2	−	5437
M	QD8130	223	189 198 200 215 315+C 318 326	398 *400*					−		+	+	2	−	12351 12705
M	XJ8436	223 294 295	200 215 309+C 315+C 318 326	398 *400*						+	+	2	−	9950
M	GD7817	172 173 223	146 151 200 215 309+C 315+C 318 326	398 *400*					−	−	+	+	2	−	
M	YN149	183C 189 293C 325 362	146 234 315+C	325 398 *400*					−	−	+	+	2	−	
M	2GD7819	129 223 270 362	103 204 309+C 315+C	398 *400*					−	−	+	+	1	−	
M	LN7594	172 174 223 362	315+C	398 *400*					−	−	+	+	2	−	
M	GD7815	093 104 111 223 362	146 309+C 315+C	398 *400*			G	C	−	−	+	+	2	−	
M	GD7821	093 104 111 223 235 362	309+C 315+C	398 *400*	CRS				−	−	+	+	2	−	
G2	LN7709	166 223 *278* 335 *362*	152 315+C	398 *400*						+	+	−	2	−	4769, *4833; 7600;* 15392–16047=CRS
G2a	QD8136	223 *227 278* 311 362	152 315+C	398 *400*						+	+	−	2	−	
G2a	LN7719	189 194 223 *227 278* 311 *362*	309+C 315+C	398 *400*						+	+	−	2	−	
G2a	XJ8416	189 223 *227* 256 *278 362*	153 195 309+C 315+C	398 *400*						+	+	−	2	−	4769, *4833;* 7533 *7600;* 15392–16047=CRS
G2a	WH6251	223 *227* 262 *278 362*	152 309+C 315+C	398 *400*						+	+	−	2	−	
G2a	LN7551	111 223 *227 278 362*	309+CC 315+C	398 *400*						+	+	−	2	−	
G2a	LN7587	111 209 223 *227* 274 *278* 326 *362*	309+C 315+C	398 *400*						+	+	−	2	−	
G2a	QD8117	209 223 *227* 234 *278* 309 *362*	152 315+C	398 *400*						+	+	−	2	−	
G2a	QD8152	223 *227* 272 *278* 319 *362* 365	152 198 282 309+C 315+C	398 *400*						+	+	−	2	−	
D5a	GD7837	164 172 182C 183C *189* 223 235 *266* 291 491G	*150* 315+C	364 *397* 398 *400*							−		2		
D5a	WH6250	164 172 182C 183C *189* 223 *266* 300 *362*	309+C 315+C	397 398 *400*							−		2		
D5a	QD8126	164 172 182C 183C *189* 223 *266 362*	*150* 315+C	397 398 *400*							−		2		
D5a	XJ8437	092 145 164 182d 183C *189* 223 *266 362*	*150* 315+C	397 398 *400*							−		2		
D5a	QD8124	092 164 167 182C 183C *189* *266 362*	*150* 309+CC 315+C	397 398 *400*							−		2		

D5a	QD8144	092 172 182C 183C 189 223 362	150 315+C	397 398 400					−		2		
D5a	XJ8423	092 172 182C 183C 189 223 266 362	150 309+C 315+C	397 398 400				−	−		2		
D5a	WH6984	172 182d 183C 189 223 266 362	150 309+CC 315+C	397 398 400					−		2		
D5a	YN167	172 182C 183C 189 223 266 299 319 362	150 309+C 315+C	397398 400					−		2		
D5a	LN7577	169 172 182C 183C 189 223 266 362	150 309+C 315+C	397 398 400					−		2		
D5	LN7713	092 148 183C 189 223 256 362	150 152 185 309+C 315+C	397398 400 654					−		2		
D5	QD8149n	189 223 319 362	150 185 237 309+CC 315+C	397 398 400					−		2		
D5	GD7820	148 182C 183C 189 223 362	150 152 309+C 315+C	397 398 400					−		2		
D5	LN7578	189 210 223 311 316 362	150 151 152 309 C 315+C	397398 400					−		2		
D5	YN289	189 223 362	150 315+C	397 398 400					−		2		
D5	XJ8412	183C 189 223 362	150 152 309+C 315+C	397 398 400	CRS				−		2		
D5	QD8162	ND	146 150 247 309+CC 315+C	397 398 400					−		2		9775–10163 = CRS
D4a	XJ8441	129 223 249 278 311 362	152 200 309+CC 315+C	398 400		A	T	−	−	+	2	−	
D4a	QD8166	129 223 234 249 311 362	152 315+C	398 400	CRS	A	T		−		2		5178A
D4a	LN7581	129 193 223 256 362	152 315+C	398 400 410		A	T		−		2		
D4a	LN7600	111G 129 223 362	152 315+C	398 400		A	T		−		2	−	
D4a	2QD8127	129 223 362	152 309+C 315+C	398 400		A	T		−		2		
D4a	QD8140	129 223 362	152 217 315+C	398 400		A	T		−		2		
D4a	GD7841	129 162 223 362	152 282 309+C 315+C	398 400		A	T		−		2		
D4b	YN171	184d 186 189 223 319 362	185 189 315+C	181 398 400		A	C		−		2		
D4b	GD7830	223 287 319 362 380	315+C	181 398 400		A	C		−		1		
D4	WH6245	111 223 261 362	194 315+C	398 400		A	C		−		2		
D4	LN7599	111 187 223 362	194 309+C 315+C	398 400		A	C		−		2		
D4	LN7575	176 223 291A 362	94 194 315+C	398 400		A	C		−		2		
D4	QD8137	093 176 223 362	94 194 309+C 315+C	398 400		A	C		−		2		
D4	YN283	093 223 232 290 362	195 198 315+C	398 400 646		A	C		−		2		
D4	XJ8425	093 223 362	94 315+C 325	398 400		A	C		−		2		
D4	XJ8434	093 362	109 153 194 315+C	398 400		A	C		−		2		
D4	QD8146	051 069 223 362	194 315+C	274 398 400		A	C		−		2		
D4	YN286	223 291 362	150 194 310 315+C	398 400		A	C		−		2		
D4	LN7603	223 245 269 362 368	315+C	398 400		A	C		−		2		
D4	QD8131	223 224 245 292 362	146 315+C	398 400		A	C		−		2		
D4	QD8115	184 223 311 362	152 309+C 315+C	398 400		A	C		−		2		
D4	WH6253	192 223 278 316 362	143 184 309+C 315+C	398 400		A	C	−	−	+	2	−	
D4	XJ8433	223 249 261 278 362	152 204 309+C 315+C	398 400		A	C	−	−	+	2	−	
D4	QD8121	223 249 362	309+C 315+C	398 400		A	C		−		2		
D4	QD8153	148 223 249 301 342 362	152 309+CC 315+C	398 400		A	C		−		2		
D4	LN7582	223 274 362	151 298 309+C 315+C	398 400		A	C		−		2		
D4	LN7568	095 209 223 294 362	195 315+C	398 400		A	C		−		3		
D4	LN7553	174 223 362	182 309+C 315+C	400		A	C		−		2		
D4	XJ8444	174 362	309+CC 315+C	398 400		A	C		−		2		
D4	XJ8432	362	194 315+C	398 400		A	C		−		2		
D4	3XJ8411	223 362	194 315+C	398 400		A	C		−		2		
D4	LN7550	223 362	309+C 315+C	398 400		A	C		−		2		
D4	QD8129	223 362	184 199 204 309+C 315+C	398 400		A	C		−		2		
D4	QD8157	223 362	(263) 315+C	398 400					−		2		
D4	QD8139	174 223 311 343 362	152 309+C 315+C	398 400					−		2		5048 5147 5178A
D4	YN159	223 362 497	94 315+C	398 400		A	C		−		2		
D	GD7829	183C 189 223 311 362	152 204 309+C 315+C	398 400		G	C		−		2		

Table 2 (continued)

Haplogroup	Sample Number[a]	16001–16497 HVS-I (16000+)	30–407 HVS-II (73 and 263)	10171–10659 (10000+)	14055–14590 (14000+)	663e	3010	3206	3391e	4831f	5176a	7598f	9-bp	9820g	Other Polymorphisms[d]
			Mutations in Region[b]							**RFLP Polymorphisms**[c]					
A	XJ8409	223 *290 319* 362	*152 235* 315+C	CRS	CRS	+							2		(522–523)d *663*750
A	WH6247	223 *290 319* 362	*152* 207 *235* (263) 309+C 315+C 372	CRS		+							2		
A	QD8164	223 *290 319* 362	*152* 156 159 182 *235* 309+CC 315+C	CRS		+							2		
A	XJ8446	223 289 *290 319* 362	151 *152 235* 315+C	CRS		+							2		
A	2XJ8430	223 274 *290 319* 362	200 *235* 309+C 315+C	CRS		+							2		
A	WH6243	223 274 *290 319* 362	151 *152 235* 309+C 315+C	CRS		+							2		
A	WH6959	126 223 *290 319* 362	*152 235* 309+C 315+C	320 335		+							2		
A	WH6980	223 *290* 294 *319* 362	*152 235* 315+C	CRS		+							2		
A	QD8148	131 222A 223 *290 319* 362	151 *152* 200 *235* 309+CC 315+C	CRS		+							2		
A	LN7604	037 086 223 *290 319* 356 362	*152 235* 315+C	CRS		+							2		
A	YN178	086 223 *290 319* 362	150 *235* 249d 315+C	364		+							2		
A	YN271	051 223 *290 319*	*152 235* 315+C	646		+	G	C					2		
A	WH6965	189 223 *290 319*	*152 235* 292 309+C 315+C	CRS		+							1		
A	WH6956	051 129 182C 183C 189 223 *290 319*	*152* 200 *235* 309+C 315+C	CRS		+							2		
A	XJ8445	223 *290* 293C *319*	*152* 309+C 315+C	CRS		+							2		
A	LN7588	093 223 263 *290* 293C *319*	*152 235* 315+CC	CRS		+							2		
A	WH6954	223 *290 319*	*152 235* 309+C 315+C	CRS		+							2		
A	LN7580	129 213 223 *290 319*	*152 235* 309+CC 315+C 392	CRS		+							2		
N	WH6976	189 223 355	195 198 315+C	289		−						+	2		4888–5441=CRS; 12634 12705
N9a	WH6254n	086 111 129 192A 223 *257A 261*	150 309+CC 315+C	CRS									2		*12358 12372* 12705
N9a	GD7834	111 129 223 *257A 261*	146 *150* 309+C 315+C	CRS									2		*5231 5417*
N9a	YN284n	129 162 223 250 *257A 261*	*150* 309+C 315+C	CRS									2		
N9a	WH6972	166C 173 223 250 *257A* 324	*150* 315+C	CRS								+	2	−	*5231 5417*
N9a	WH6977	129 189 223 *257A 261*	*150* 183 194 195 309+CC 315+C	CRS									2		
N9a	QD8145	066 092 145 172 223 245 *257A 261*	*150* 315+C	CRS									2		
N9a	GD7828	145 172 223 245 *257A 261*	*150* 309+C 315+C	CRS									2		
N9a	YN175	172 223 *257A 261* 311	*150* 195 309+CC 315+C	CRS									2		
N9a	YN176	223 *257A 261* 362	*150* 309+C 315+C	CRS									2		
N9a	LN7591	223 *257A 261*	*150* 309+CC 315+C	CRS									2		
N9a	QD8123	223 *257A 261*	*150* 309+C 315+C	CRS									2		
N9a	QD8156	223 *257A 261*	*150* 309+C 315+C	CRS									2		*5231 5417*
Y	XJ8426	*126 231 266*	146 309+CC 315+C	*398*	178								2		*5417*
Y	LN7579	*126 231 266* 293	*146* 315+C	*398*	178								2		*5417*
Y	QD8151	*126* 193 *231 266*	*146* 245 315+C	205 *398*	178								2		*5417*
B4a	LN7565	182C 183C *189 217 261*	71+G 73C 75 89 315+C	238									1		
B4a	QD8118	182C 183C *189 217 261*	146 150 152 195 309+CCC 315+C	238									1		
B4a	LN7585	153 182C 183C *189 217 261*	146 (306–309)d 315+C	238									1		
B4a	QD8128	129 182C 183C *189* 223 *261* 311	151 152 310 (314–315)d	238								+	1	−	
B4a	YN155	182C 183C *189 217 261* 299 355 390	35 36 152 309+CC 315+C	495									1		
B4a	YN158	092 182C 183C *189 217 261* 299	193 309+C 315+C	CRS									1		
B4a	LN7602	182C 183C 184 *189 217* 247 *261* 299	193 315+C	CRS									1		

B4a	GD7840	093 153 (181–183)C *189 217 261* 292 311 362	309+CC 315+C	CRS			1	
B4a	GD7812	181d 182C 183C *189 217 261* 292	309d 315+C	CRS			1	
B4a	YN174	129 182C 183C *189 217 261*	146 195 257 309+C 315+C	CRS			1	
B4a	QD8170	129 182C 183C *189* 223 *261*	309+CC 315+C	CRS		+	1	
B4a	WH6248	182C 183C *189* 257 *261* 360	152 309+CC 315+C	CRS	CRS		1	
B4b	GD7838	*136* 183C *189 217* 218	93 309+CC 315+C	CRS			1	
B4b	WH6961	*136* 182C 183C *189 217* 218	309+CC 315+C	ND			1	
B4b	GD7813	*136* 183C *189 217* 309 354	207 309+C 315+C	CRS			1	
B4b	GD7814	*136* 183C *189 217* 309 354	146 207 315+C	CRS			1	
B4b	QD8119	092 *136* 183C *189* 309 354	207 315+C	CRS			1	
B4b	QD8169	*136* 183T *189 217* 218 239 248	309+C 315+C	CRS			1	
B4b	WH6978	*136* 183C *189* 257	309+C 315+C	CRS			1	
B4b	XJ8428	*136* 183C *189*	114 309+CC 315+C	CRS			1	
B4b	LN7716	*136* 183C *189* 284	199 202 207 309+CCC 315+C	CRS			1	
B4	LN7589	183C *189 217*	309+CC 315+C 316	CRS			1	
B4	WH6945	182C 183C *189 217* 223 311 320	315+C	CRS			1	
B4	LN7572	182d 183C *189 217* 223 235 291 316	146 185 189 195 196 309+C 315+C	CRS			1	
B4	YN169	182C 183C *189 217* 234	309+C 315+C	CRS			1	
B4	LN7552	140 182d 183C *189 217* 274 311	146 150 315+C	CRS			1	9775–10163=CRS
B4	YN154	140 183C *189 217* 274	150 152 309+C 315+C	CRS	CRS		1	9775–10163=CRS; 11440 11719 11887; 12335–12877=CRS; 14687 14766
B5a	XJ8424	*140* 182C 183C *189 266A*	210 309+CC 315+C 391	*398*			1	
B5a	WH6967	*140* 183C *189 266A*	210 315+C	*398*			1	
B5a	XJ8454	*140* 187 *189* 256 *266G*	93 210 315+C	*398*			1	
B5a	YN150	*140* 145 183C *189* 217 *266A*	93 146 315+C	*398*			1	
B5a	LN7564	*140* 187 *189 266A*	93 146 210 315+C	*398*			1	
B5a	YN168	*140* 145 183C *189 266A*	210 309+C 315+C	*398*			1	
B5b	YN284	111 129 *140* 182C 183C *189* 234 *243* 249 250 463	131 199 204 292 315+C	ND			1	
B5b	XJ8413	111 *140* 182C 183C *189* 234 *243* 344 463	103 315+C	*398*			1	
B5b	WH6973	111 *140* 182C 183C *189* 234 *243* 291 463	131 204 309+C 315+C	*398*			1	
B5b	LN7714	111 *140* 182C 183C *189* 234 *243* 463	131 204 207 309+C 315+C	*398*			1	
B5	WH6246	183C *189* 311	150 195 214 279 309+CC 315+C	*398*			1	
B	QD8141	183C *189*	309+C 315+C	CRS			1	
B	WH6982	183C *189* 234	315+C	CRS			1	
B	GD7832	129 183C *189* 352 355	150 152 185 189 309+C 315+C	589 *595*			1	
B	YN172	093 179 182C 183C *189*	150 185 309+CC 315+C	CRS			1	
R9a	XJ8451	209 *298* 311 *355 362*	195 *249d* 309+C 315+C	*310 320*			2	
R9a	XJ8408	093 260 *298 355 362*	152 207 *249d* 309+C 315+C	*310 320*			2	
R9a	YN153	093 111 126 192 249 263 *298 355 362* 390	207 *249d* 309+C 315+C	*310 320* 496			2	
F1a	YN160	108 *129* 162 *172* 293 *304*	*249d* 315+C	*310 609* 653			2	
F1a	LN7721	108 *129* 162 *172 304*	150 *249d* 309+C 315+C	*310 609*			2	
F1a	GD7824	108 *129* 162 *172 304*	195 *249d* 309+C 315+C	*310 609*	CRS		2	
F1a	YN161	*129* 162 *172* 260 *304*	*249d* 315+C	310 609			2	
F1a	XJ8427	*129* 162 *172 304*	151 153 *249d* 315+C	*310 609*	170		2	
F1a	WH6252	*129* 162 *172 304* 311	*249d* 315+C	*310 609*			2	

(continued)

Table 2 (continued)

Haplogroup	Sample Number[a]	Mutations in Region[b] 16001–16497 HVS-I (16000+)	30–407 HVS-II (73 and 263)	10171–10659 (10000+)	14055–14590 (14000+)	663e	3010	3206	3391e	4831f	5176a	7598f	9-bp	9820g	Other Polymorphisms[d]
F1a	QD8116	*129* 162 *172 304* 399	152 *249d* 315+C	*310 609*									2		
F1a	QD8161	*129* 162 *172 304* 497	*249d* 315+C	*310 609*									2		
F1a	WH6985	*129 172* 295 *304* 311	*249d* 315+C	*310 609*									2		15930
F1a	WH6975	*129 172* 218 *304* 354	195 *249d* 315+C	310 604 609									2		
F1a	GD7816	*129 172 304* 362	151 *249d* 315+C	310 604 609									2		
F1a	YN281n	*129 172 304*	152 *249d* 315+C	310 604 609									2		
F1a	YN151	*129 172* 184 *304*	*249d* 309+CC 315+C	*310 609*	CRS								2		
F1a	XJ8431	*129 172 304* 399C	152 *249d* 315+C	*310 609*									2		
F1a	YN165	093 *129 172* 294 *304* 362	152 *249d* 315+C	*310 609*									2		
F1c	XJ8421	*111 129* 294 *304*	*152* 234 *249d* 315+C	310 454 609									2		
F1c	WH6971	*111 129* 266 *304*	*152 249d* 309+C 315+C	310 454 609	CRS								2		
F1c	QD8167	*111 129* 266 *304*	*152 249d* 309+CC 315+C	310 454 609									2		
F1b	YN166	183C *189* 232A 249 *304* 311	146 204 207 *249d* 309+C 315+C	*310 609*									2		
F1b	LN7586	183C *189* 232A 249 *304* 311	143 204 *249d* 309+C 315+C	*310 609*									2		
F1b	XJ8447	183C *189* 232A 249 264 *304* 311	199 204 *249d* 309+C 315+C	*310 609*							+		2	−	
F1b	YN290	129 145 182C 183C *189* 232A 249 *304* 311 344	152 *249d* 315+C	ND									2		
F1b	QD8154	183C *189 304* 311	195 *249d* 309+C 315+C	*310 609*									2		
F1b	GD7811	172 180 *189 304* 465	217 *249d* 309+CC 315+C	*310 609*									2		
F1b	WH6949n	051 *129* 183C *189 304*	150 238 *249d* 315+C	*310 609*									2		
F1b	QD8143	*189 304*	150 195 *249d* 315+C	*609*									2		
F2a	XJ8414	092A *291 304*	*249d* 309+CC 315+C	*310 535 586*									2		
F2a	GD7836n	092A *291 304* 359	*249d* 309+C 315+C	*310 535 586*									2		
F2a	YN281	051 *291 304*	195 *249d* 315+C	ND									2		
F2a	QD8147	266 *291 304*	146 *249d* 315+C	*310 535 586*									2		
F2a	XJ8407	203 239 *291 304*	*249d* 309+C 315+C	*310 535 586*									2		
F2	GD7809	086 203 *304*	*249d* 315+C	*310 535 586*									2		
F2	LN7601	129 203 *304*	195 *249d* 315+C	*310 535 586*									2		
F2	WH6974	192 *304*	*249d* 315+C	265 *310 535 586*									2		
F2	WH6948	299 *304*	*249d* 309+C 315+C	*310 535 586*									2		
F2	GD7810	261	194 235 *249d* 309+C 315+C	*310 535 586*			−						2		11719; 12338; 14766
F2	GD7842	129 189 *304*	207 *249d* 315+C	310 535									2		
F	XJ8440	207 *304* 362 399	146 152 *249d* 315+C	*310*									2		
F	YN170	157 256 *304* 335	236 *249d* 315+C	CRS									2		
R	XJ8448	093 304 309 390	152 309+C 315+C	CRS									2		
R	LN7595	182C 183C 189 304 311	185 189 309+CCC 315+C	398							+		2		
R	QD8168	182C 183C 189 259 311 390	150 185 189 234 309+CC 315+C	398							+		2		
T1	LN7592	126 *163 186 189* 294	152 195 309+C 315+C	*463*	CRS						+		2		11719; 14766 *14905*
HV	YN287	217 240	152 309+CC 315+C	CRS									2		11339–11943=CRS; 12681; 14576–15085=CRS

NOTE.—The mtDNAs that had no mutation in a sequenced region compared with the reference sequence are labeled by CRS. ND = not determined.

[a] The Han populations from Yunnan, Wuhan, Gongdong, Qingdao, Liaoning, and Xinjiang are abbreviated as YN, WH, GD, QD, LN, and XJ, respectively. Numbers prefixing sample identification codes indicate sample frequencies >1 in the same population; for example, 4YN285 means that four Yunnan Han individuals share the same haplotype.

[b] Sites are numbered according to the revised CRS of Andrews et al. (1999). The suffixes A, G, C, and T indicate transversions, d indicates a deletion, and a plus sign (+) indicates an insertion. Insertions and deletions are recorded at the last possible site (as is usual in forensics); thus, insertions and deletions in HVS-II are scored at 249, 309, 315, and 522–523, instead of, for example, at 248, 303, 311, and 514–515. For each haplogroup, characteristic mutations are shown in boldface, and diagnostic restriction sites are boxed.

[c] The restriction enzymes used in the analyses are designated by the following single-letter codes: a = *Alu*I ; e = *Hae*III ; f = *Hha*I ; g = *Hin*fI ; − and + denote the absence and presence of the restriction site, respectively. "1" denotes the presence of the 9-bp (CCCCCTCTA) deletion, "2" denotes nondeletion (i.e., two repeats of the 9-bp fragment), and "3" denotes triplication of the 9-bp fragment.

[d] Additional polymorphisms in the coding region refer to the segments listed in table 1.
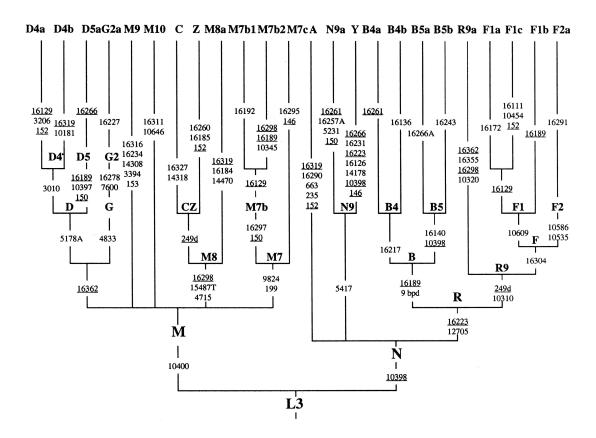
Figure 2 Classification tree of the mtDNA haplogroups observed in Han Chinese. The diagnostic mutations considered here (relative to the revised CRS; Andrews et al. 1999) are indicated on the branches. Nucleotide changes are specified for transversions by suffixes, and "d" indicates deletion; mutations recurrent in this tree are underlined. The revised CRS differs from the root of haplogroup R by mutations at 73, 2706, 7028, 11719, and 14766 that are characteristic of HV or H and by seven private mutations at 263, 315+C, 750, 1438, 4769, 8860, and 15326 (Andrews et al. 1999).

(T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data), is identified by 10310. A branch of R9, haplogroup R9a, is further characterized by 10320 in addition to its HVS-I motif. Haplogroup F1 (F *sensu stricto,* as originally introduced by Torroni et al. [1994]) may be characterized by 10609 as well, whereas its sister group F2 (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data) likely has the defining sites 10535 and 10586. The complete mtDNA sequence (XLIND) from China, reported by Ingman et al. (2000), is thus identified as an F type that does not belong to F1 or F2 (fig. 2). The newly defined subhaplogroup F1c of F1 has the characteristic site at 10454.

The region 14055–14590 is also quite informative for the Asian mtDNA phylogeny. It harbors one marker each for haplogroups C (14318), Y (14178), and M8a (14470, also recognizable by +14465 *Acc*I). Haplogroup M9, introduced here, has the two characteristic sites 14308 and 3394 (identifiable by +3391 *Hae*III).

In the recently published complete sequence data (Ing-

man et al. 2000; Finnilä et al. 2001), haplogroups C and Z were found to share the transition at 4715 and the A→T transversion at 15487 (among other mutations). Our typing of an M8a mtDNA confirms that the former two mutations are also shared by haplogroup M8a, thus supporting the phylogenetic position of M8, with CZ and M8a forming sister clades (T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data). The 9-bp deletion in the COII/tRNA^Lys intergenic region, which is a diagnostic marker for haplogroup B, was found sporadically in lineages from A, D, and M*, thus confirming our previous results about the multiple origin of the deletion in these individuals (Yao et al. 2000b).

As to the dating of the nodes in the classification tree, table 3 lists the age estimates of the major haplogroups. Haplogroups M7, CZ, M8, G2, N9, B, B4, B5, F, F1, and R9 are all rather ancient, with ages >50,000 years. The ages of the other haplogroups seem to fall into the range 30,000–50,000 years, except for that of M8a, which may be <20,000 years.

Table 3

Estimated Haplogroup Coalescent Times

Haplogroup[a]	Size	$\rho \pm \sigma$[b]	Age (years)[b]
M7b	21	2.24 ± 1.06	45,200 ± 21,400
M7	32	2.78 ± .99	56,100 ± 20,000
M8a	15	.87 ± .33	17,500 ± 6,700
CZ	13	3.00 ± .93	60,500 ± 18,700
M8	28	2.93 ± .89	59,100 ± 17,900
G2	10	3.00 ± .95	60,500 ± 19,100
D5	20	2.55 ± .81	51,500 ± 16,300
D4	44	1.73 ± .30	34,900 ± 6,000
D	66	2.30 ± .44	46,500 ± 8,900
A	19	1.42 ± .68	28,700 ± 13,700
N9a	13	1.85 ± .51	37,300 ± 10,300
N9	16	3.19 ± .99	64,300 ± 20,000
B4a	22	2.00 ± .48	40,400 ± 9,600
B4b	13	1.85 ± .60	37,300 ± 12,000
B4	47	2.94 ± .65	59,300 ± 13,200
B5	12	2.50 ± .81	50,500 ± 16,300
B	63	3.70 ± .92	74,600 ± 18,700
F1a	27	1.48 ± .63	29,900 ± 12,600
F1	40	3.35 ± 1.15	67,600 ± 23,300
F2	14	1.50 ± .53	30,300 ± 10,700
F	57	2.86 ± .82	57,700 ± 16,600
R9	61	4.03 ± 1.22	81,400 ± 24,600

[a] Based on 332 Han mtDNAs (present study; T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data).

[b] ρ and σ are as defined by Forster et al. (1996) and Saillard et al. (2000), scoring transitions within 16090–16365.

From Coding Region to Control Region

The present Han mtDNA data (including those of T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data) with coding-region information can serve as a starting point for provisional haplogroup assignment of those east Asian mtDNAs for which only a segment of the control region is available (see GenBank). Potential haplogroup status can then be inferred through a motif search and (near-)matching with the 332 Han mtDNAs. For illustration, we take ancient mtDNA data, which usually offer only short fragments of HVS-I (and HVS-II). The mtDNAs from the 2,000-year-old Yixi site from Shandong Province (Oota et al. 1999), with polymorphic sites reported from 16203 to 16362 and from 146 to 263, can all be assigned to specific haplogroups, albeit at different levels of certainty. For example, sequence 01 (16203-16291-16304, 249d-263) does not match any of the 332 Han mtDNAs but has three one-step neighbors (XJ8414, XJ8407, and GD7809), all in F2; since it bears the full motif 16291-16304-249d for F2a, we can quite safely conclude that the sequence belongs to F2a. In contrast, sequence 19 (16223, 146-263) has no close companion (at distance two or fewer mutational steps) in the Han

data and lacks any salient motif of the haplogroups considered here; therefore, if it can be assigned at all, we could at best assign it to M*.

An interesting case is constituted by the 29 mtDNAs from the 4,500-year-old Nakazuma Jomon site that were sequenced for the region 16209–16402 (Shinoda and Kanai 1999). The haplogroup affiliations of the resulting nine haplotypes, except for type 9 (16256-16278-16295), can be recognized by following our classification strategy. Type 1 (16223-16311-16357) matches haplotypes from M10 (one sampled in Liaoning and another one in Yunnan), and type 7 (16284) matches a B4b haplotype from Liaoning. The other six types have one-step neighbors in the Han mtDNA database: type 2 (16223-16234-16290-16319) is thus related to A haplotypes from Wuhan and Yunnan; type 3 (16223-16298-16319-16355) to M8a haplotypes from Qingdao and Wuhan; type 4 (16223-16266-16274-16362) to a D4 haplotype from Liaoning and to D5a haplotypes from Liaoning, Wuhan, Xinjiang, and Qingdao; type 6 (16223-16278-16362) to two G2 haplotypes and type 8 (16223-16245-16362-16368) to one D4 haplotype, all from Liaoning; finally, type 5 (16223-16357) is a one-step descendant of the matched M10 type 1 (but, alternatively, it would also be a one-step neighbor of an M* haplotype from Qingdao). It is conspicuous that the Jomon mtDNAs find their near-matches within the Han mtDNA database mainly in the northern and central pools, especially in the Liaoning sample.

Haplogroup Profiles

Haplogroup frequencies varied among the regional Han populations (table 4). Five main features can be discerned. (1) Haplogroups A, Z, and Y are absent in the two Guangdong samples. These two samples differ significantly in the number of M* mtDNAs. Haplogroup M7b (including M7b1, M7b2, and M7b*) is absent in the Zhanjiang sample but is present, with a frequency of 8.7%, in the Guangzhou sample. The frequency of F1a in the Guangzhou sample (17.4%) is higher than that in the Zhanjiang sample (6.7%). (2) Haplogroup M7b1 has by far the highest frequency (14.0%) in the Yunnan sample, whereas, in central and northeast China, it only occurs at low frequencies (<5.0%). (3) The Wuhan sample shows a relatively high frequency of haplogroup A (16.7%), followed by the Shanghai (11.7%) and Xinjiang (10.6%) samples. These three samples and the Zibo sample have relatively high frequencies (> 7.5%) of CZ. (4) Most of the mtDNAs that belong to haplogroups M9, M8a, Y, and G2 are restricted to the northern and northwestern populations of Liaoning, Qingdao, Xinjiang, and Qinghai, although the Taiwanese samples also include a good number of M9, Y, and G2 mtDNAs. The newly

Table 4

Estimated Frequencies (%) of mtDNA Haplogroups in Regional Han Populations

mtDNA Haplogroup	ESTIMATED FREQUENCY (%) IN REGION[a]												
	YN (n = 43)	WH (n = 42)	QD (n = 50)	LN (n = 51)	XJ (n = 47)	GD-ZJ (n = 30)[b]	GD-GZ (n = 69)[c]	HK (n = 20)[d]	TW1 (n = 66)[e]	TW2 (n = 155)[f]	QH (n = 78)[g]	SH (n = 120)[h]	ZB (n = 50)[i]
M7b1	14.0		2.0	2.0	2.1		2.9	5.0	9.1	2.6	5.1	2.5	ND
M7b2		2.4	2.0		4.3					.6		.8	2.0
M7b*	2.3	2.4					5.8	5.0		3.9	1.3	2.5	2.0
M7c		2.4		5.9	2.1	3.3	1.4	5.0		4.5	1.3		2.0
M7*	2.3	2.4			2.1		1.4	ND	ND	.6	ND	ND	ND
M8a		7.1	8.0	7.8	4.3		2.9		1.5	3.9	6.4	.8	2.0
C	4.7	2.4		2.0	6.4	3.3		5.0	3.0	3.2	2.6	7.5	8.0
Z		7.1		2.0	2.1					1.3	5.1	2.5	2.0
M9			4.0	3.9	4.3	3.3			1.5	1.3	3.8		2.0
M10	2.3		2.0	5.9			2.9	5.0	1.5	2.6	1.3		2.0
M*	2.3		2.0	2.0	2.1	23.3	2.9	ND	ND	ND	ND	ND	ND
N*		2.4					1.4	ND	ND	ND	ND	ND	ND
M*/N*[j]	2.3	2.4	2.0	2.0	2.1	23.3	4.3	10.0	3.0	3.2	2.6	3.3	2.0
G2		2.4	6.0	7.8	2.1		1.4		3.0	2.6	3.8		6.0
D*						3.3	1.4	ND	ND	ND	ND	ND	ND
D4a			8.0	3.9	2.1	3.3		ND	ND	ND	ND	ND	ND
D4b	2.3					3.3		ND	ND	ND	ND	ND	ND
D4*	7.0	4.8	18.0	13.7	17.0		7.2	ND	ND	ND	ND	ND	ND
D4[k]	9.3	4.8	26.0	17.6	19.1	10.0	8.7	10.0	18.2	18.7	17.9	25.0	26.0
D5a	2.3	4.8	6.0	2.0	4.3	3.3			1.5	2.6	2.6	3.3	4.0
D5*	2.3		4.0	3.9	2.1	3.3	5.8		3.0	5.8	2.6	5.0	2.0
A	4.7	16.7	4.0	5.9	10.6			5.0	6.1	6.5	5.1	11.7	6.0
N9a	7.0	7.1	6.0	2.0		6.7	1.4		3.0	2.6	7.7		6.0
Y			2.0	2.0	2.1				1.5	1.3	3.8	2.5	2.0
B4a	7.0	2.4	6.0	5.9		6.7	14.5	10.0	6.1	7.7	2.6	1.7	2.0
B4b		4.8	4.0	2.0	2.1	10.0	5.8		4.5	3.2	2.6		2.0
B4*	4.7	2.4		5.9			8.7		3.0	1.3	3.8	5.0	4.0
B5a	4.7	2.4		2.0	4.3			10.0	4.5	2.6	1.3	5.0	
B5b	2.3	2.4		2.0	2.1		1.4		1.5	.6	2.6		4.0
B5*		2.4								.6		1.7	
B*	2.3	2.4	2.0			3.3				.6		2.5	
R9a	2.3				4.3		1.4	10.0		1.9		.8	
R*			2.0	2.0	2.1		1.4	5.0	1.5	1.9	1.3	3.3	2.0
F1a	11.6	7.1	4.0	2.0	4.3	6.7	17.4	15.0	13.6	5.8	3.8	5.0	ND
F1b	4.7	2.4	4.0	2.0	2.1	3.3	1.4		1.5	1.3	2.6	3.3	ND
F1c		2.4	2.0		2.1		1.4			.6		.8	2.0
F2a	2.3		2.0		4.3	3.3	1.4		3.0		1.3	0.8	2.0
F2*		4.8		2.0		10.0	2.9		1.5	1.3		ND	ND
F*	2.3				2.1		1.4		3.0	1.9		2.5	6.0
Other[l]	2.3		2.0				1.4				5.1		

NOTE.—Reported samples lacking coding-region information were classified within the coarser haplogroup scheme. Since only 185-bp fragments of HVS-I were available for the Zibo sample, the entries in this column are only approximate; in particular, one cannot exclude that some default F* haplotypes actually belong to F1a or F2*.

[a] YN = Yunnan; WH = Wuhan; GD = Gongdong; QD = Qingdao; LN = Liaoning; XJ = Xinjiang; GD-ZJ = Guangdong, Zhanjiang; GD-GZ = Guangdong, Guangzhou; HK = Hong Kong; TW-1 = Taiwan-1; TW-2 = Taiwan-2; QH = Qinghai; SH = Shanghai; ZB = Zibo, Shandong. ND = not determined.

[b] Present study.

[c] T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data.

[d] Betty et al. (1996).

[e] Horai et al. (1996).

[f] Tsai et al. (2001).

[g] Y.-G. Yao, P.-L. Geng, Q.-P. Kong, and Y.-P. Zhang, unpublished data.

[h] Nishimaki et al. (1999).

[i] Wang et al. (2000).

[j] Compound M* and N* frequency in the first seven columns and unassigned haplotypes in the last six columns.

[k] D minus D5 is taken as a proxy for D4.

[l] West Eurasian haplotypes. This row was not included as a coordinate in the PC analysis.

defined haplogroup, M10, has the highest frequency in the Liaoning sample (5.9%). (5) Generally, the frequencies of haplogroups F1 and B tend to decrease from south to north, whereas the D4 frequency increases.

PC Maps for mtDNA Data

The basal mtDNA haplogroup profiles of the 13 Han samples were treated as input vectors for the PC analy-

sis. Figure 3 displays the PC map for the first two principal components, which together account for 63% of the total variation. A geographic patterning of the samples is evident in the map, as mainly expressed by the first PC. The second PC, however, also contributes to the south-to-north cline (leaving aside the outlier—the Zhanjiang sample from southernmost mainland China). The two populations from Guangdong, Guangzhou and Zhanjiang, are distant from each other in the PC map, although they are geographically proximate. In contrast, the four northern populations (Qinghai, Liaoning, Qingdao, and Zibo) are close together. Although the Zibo data were extremely meager (185-bp fragments of HVS-I), the haplogroup classification, by and large, seems to be correct, since Zibo comes next to Qingdao (from the same province, Shandong) in the map. The populations with recent migration history, Taiwanese and Xinjiang Han, take intermediate positions in the PC map, in the vicinity of the populations from central and east China.

In the PC map, with respect to the coarse profiles (with 33 entries; see table 4), the south-to-north cline of the populations observed in the basal PC map does not change considerably (map not shown). Since the basal haplogroups are probably as old as ≥50,000 years, one could expect that the ancient imprints of the earliest settlement processes on regional mtDNA pools are slightly more pronounced in the basal PC map.

Discussion

The phylogenetic analysis of the Han HVS-I and HVS-II sequences is greatly enhanced by the information provided by the region 10171–10659 and other specific polymorphisms, which enables us to distinguish between the two macrohaplogroups M and N and to identify several new haplogroups. The region 10171–10659, which had not been studied before (unless complete sequencing was carried out), overlaps with the ND3 gene that was sequenced in a small worldwide sample by Nachman et al. (1996); with respect to our classification scheme, we can immediately infer that their types, 11 and 13, belong to haplogroup D5, type 6 to B4a, and type 3 to R9. The now-emerging tree of East Asian mtDNAs (present study; T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S. S. Papiha, H.-J. Bandelt, and R. Villems, unpublished data) can help to direct complete sequencing efforts in that lineages would be selected from those deep branches that are not yet represented by complete sequences, thus filling the lacunae. Another benefit is the tracing of pathogenic mtDNA lineages: if a certain new mutation was found in the coding region of the patient's mtDNA, one could speed up the diagnosis by first typing this mutation in normal individuals from the same haplogroup, to see whether it is haplogroup-specific or pathogenic. The type 2 diabetes mellitus sample from Qing-
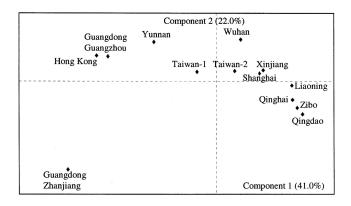

Figure 3 PC map of the mtDNA data (with respect to the basal haplogroup profiles) of 13 regional Han samples.

hai Province included here can serve as a good example in this respect. Although no normal controls from the same province have yet been analyzed for mtDNA, it is reasonable to expect that slight fluctuations in haplogroup frequencies, compared with neighboring regions (as shown in table 4 and fig. 3) reflect regional differences, rather than association with type 2 diabetes mellitus.

Coding-region information is indispensable for phylogenetic analysis of mtDNA. In cases where direct information from the coding region is not available, one can at least link combinations of HVS-I mutations with certain mutations in the coding region. Specifically, we can anticipate the haplogroup status of most East Asian HVS-I sequences via the Han database through (near-)matching and motif recognition. This classification strategy can be very useful for ancient DNA analysis, as demonstrated above. Attempts at estimating a phylogeny solely from HVS-I without any reference to coding-region sites would go astray, in particular, if neighbor joining (NJ) with midpoint rooting comes into action (see the appendix of the article by Richards et al. [1996]). For instance, this approach applied to the large Thai HVS-I data set (see fig. 3 of Fucharoen et al. [2001]) resulted in highly polyphyletic clusters: haplogroup B was distributed over two clusters, 1 and 3b; cluster 3a includes haplogroups D5, M7c, N9a, and M*; cluster 4 groups C and Z together with R9a; and cluster 8 harbors D4, D5, and A lineages. Most of the apparent clades of this NJ tree intermingle lineages from macrohaplogroups M and N and therefore would not pass the test with complete sequence data. The same kind of problem is also manifest in the NJ analysis of the HVS-I data performed by Qian et al. (2001). Even a mass screening of East Asian mtDNA data based on HVS-I alone, assisted by a network method, cannot provide a much more favorable picture. Among the six "radiation groups" I–VI, erected by Oota et al. (1999), three groups (I–III) each comprise

both M and N lineages, one group (IV) comprises Y and R lineages, and only two groups (V and VI) could potentially serve as proxies for monophyletic groups (B4 and F, respectively).

The comparison of the regional Han mtDNA samples revealed an obvious geographic differentiation in the Han Chinese, as shown by the haplogroup-frequency profiles and the PC maps. The south-to-north cline observed in the frequencies of haplogroups F1, B, and D4 is quite similar to the distributions of immunoglobulin Gm allotypes Gm$^{1,3;5}$ and Gm$^{1;21}$ in Chinese populations (Zhao and Lee 1989). Hence, the grouping of different Han populations into just "Southern Han" and "Northern Han" (Su et al. 1999, 2000) or the use of one or two Han regional populations to stand for all Han Chinese (Horai et al. 1996; Hou et al. 2001; Karafet et al. 2001) constitutes a procrustean bed and does not appropriately reflect the genetic structure of the Han. Intriguingly, despite the numerous historically recorded migrations and substantial gene flow across China from the Bronze Age to the present time (Ge et al. 1997), differences between geographic regions have been maintained. The regional difference is more pronounced in south and southwest China: in the PC map, the southern and southwestern populations show a more diverse pattern than the populations from central, east, and northeast China. The Zhanjiang and Guangzhou samples, though from the same province (Guangdong), differ considerably in their mtDNA haplogroup distribution. It thus seems that the Neolithic expansions from the Yellow River basin and later from the Yangtze River basin to other parts of China, as well as Bronze Age movements, did not erase local populations. The subsequent conquest of the Han in historical time, starting from central China, constituted mainly a political expansion process that led to the cultural assimilation of numerous ethnic groups under the dominant Han culture (Ge et al. 1997).

The spread of Han people to Yunnan, Xinjiang, and Taiwan happened relatively recently—within the past several hundred years. For the Yunnan Han, according to historical records, many movements were caused by an expansion policy, especially during the Ming dynasty (1368–1644 A.D.) (Ge et al. 1997). Since at that time the local population density was very high, the relative contribution of the Han to the local gene pools was overall rather minor, although eventually Han culture was generally accepted. Therefore, the genetic makeup of the Yunnan Han should show more influence from the autochthonous people than that of Han people from their early historical homelands in the basins of the Yellow River and the Yangtze River (see Du et al. 1998). The Taiwanese and Xinjiang Han have similar demographic histories: after World War II, both populations received a heavy influx of Han people from across almost all of China. However, before the withdrawal of the Guomingtang, Han people from the proximal Fujiang and Guangdong provinces and other parts of China continually migrated to Taiwan, with two main waves arriving in the 18th and 19th centuries (Ge et al. 1997). The high frequencies of haplogroups F1a and M7b in the Taiwanese Han, if not an autochthonous signal, might well reflect this connection with south mainland China, whereas other haplogroups—such as G2 and Y, mainly present in the north—hint at recent migrations from north and northeast China. The presence of two R9a types in Xinjiang (incidentally matching the two R9a haplotypes from Hong Kong; Betty et al. 1996), as well as the M7b haplotypes, point to connections with south and southwest China, where R9a and M7b are prevalent. On the other hand, the relatively high percentage of haplogroups A, C, and Z in this population may stem from recent migrations of Han people from central and east China to Xinjiang Province during the 1950s and 1960s. Evidence for recent migration is also reflected by the fact that no west Eurasian mtDNA types were found in the Xinjiang Han, whereas, among the Uygurs and Kazakhs from the same geographic areas (Yao et al. 2000a), >30% of individuals belong to west Eurasian haplogroups (Macaulay et al. 1999).

In summary, our phylogenetic analysis of 263 Han mtDNAs shows that ~94% of the lineages can be allocated to specific subhaplogroups of the Eurasian founder haplogroups M, N, and R (which is itself a subhaplogroup of N shared between Europe and East Asia). Most of the nested haplogroups that are not infrequent have ages >30,000 years. It is conspicuous that the potentially most ancient of these haplogroups, R9 and B, may have their earliest diversification in southern China and/or Southeast Asia. A few possibly basal branches of M, present in Guangdong but absent or rare in northern China, still await a full description with more data from Southeast Asia. Only a restricted number of major subhaplogroups of M and N—namely, G, M8, M9, A, and N9—may be of central or northern Chinese provenance. All this makes an initial pioneer colonization of China ~60,000 years ago from Southeast Asia conceivable (as proposed by Su et al. 1999; Jin and Su 2000) but still leaves much room for speculation about the population dynamics during the long period between then and the Last Glacial Maximum. The contrast between the northern and southern genetic pools might have its roots in this period. Subsequent migration events may have somewhat blurred this early distinction, with the genetic pools of central China possessing mtDNA features of both the northern and the southern pools.

Acknowledgments

Electronic-Database Information

Accession numbers and the URL for data in this article are as follows:

GenBank Overview, http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html (for mtDNA control region data; accession numbers AY052834–AY053358)

References

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

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

Bandelt H-J, Macaulay V, Richards M (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol Phylogenet Evol 16:8–28

Bandelt H-J, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med 115:64–69

Betty DJ, Chin-Atkins AN, Croft L, Sraml M, Easteal S (1996) Multiple independent origins of the COII/tRNA^Lys intergenic 9-bp mtDNA deletion in aboriginal Australians. Am J Hum Genet 58:428–433

Chen R, Ye G, Geng Z, Wang Z, Kong F, Tian D, Bao P, Liu R, Liu J, Song F, Fan L, Zhang G, Guo S, Xu L, Xu X, Cheng D, Zhao X (1993) Revelations of the origin of Chinese nation from clustering analysis and frequency distribution of HLA polymorphism in major minority nationalities in mainland China. Acta Genetica Sinica 20:389–398 (in Chinese)

Chu JY, Huang W, Kuang SQ, Wang JM, Xu JJ, Chu ZT, Yang ZQ, Lin KQ, Li P, Wu M, Geng ZC, Tan CC, Du RF, Jin L (1998) Genetic relationship of populations in China. Proc Natl Acad Sci USA 95:11763–11768

Ding Y-C, Wooding S, Harpending H, Chi H-C, Li H-P, Fu Y-X, Pang J-F, Yao Y-G, Xiang YJG, Moyzis R, Zhang Y-P (2000) Population structure and history in East Asia. Proc Natl Acad Sci USA 97:14003–14006

Du R, Xiao CJ, Cavalli-Sforza LL (1998) Genetic distances between Chinese populations calculated on gene frequencies of 38 loci. Sci China C 28:83–89

Du R, Yip VF (1993) Ethnic groups in China. Science Press, Beijing

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

Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. J Hum Genet 46:115–125

Ge JX, Wu SD, Chao SJ (1997) Zhongguo yimin shi (The migration history of China). Fujian People Press, Fuzhou, China (in Chinese)

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH (1996) mtDNA polymorphism in east Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590

Hou YP, Zhang J, Li YB, Wu J, Zhang SZ, Prinz M (2001) Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han population. Forensic Sci Int 118:147–152

Ikebe S, Tanaka M, Ozawa T (1995) Point mutations of mitochondrial genome in Parkinson's disease. Brain Res Mol Brain Res 28:281–295

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

Jin L, Su B (2000) Natives or immigrants: modern human origin in East Asia. Nat Rev Genet 1:126–133

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of east Asia: sources, patterns, and microevolutionary process. Am J Hum Genet 69:615–628

Macaulay V, Richards M, 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

Nachman MW, Brown WM, Stoneking M, Aquadro CF (1996) Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142:953–963

Nishimaki Y, Sato K, Fang L, Ma M, Hasekura H, Boettcher B (1999) Sequence polymorphism in the mtDNA HV1 region in Japanese and Chinese. Legal Med 1:238–249

Oota H, Saitou N, Matsushita T, Ueda S (1999) Molecular genetic analysis of remains of a 2,000-year-old human population in China—and its relevance for the origin of the modern Japanese population. Am J Hum Genet 64:250–258

Ozawa T (1995) Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. Biochim Biophys Acta 1271:177–189

Ozawa T, Tanaka M, Ino H, Ohno K, Sano T, Wada Y, Yoneda M, Tanno Y, Miyatake T, Tanaka T, Itoyama S, Ikebe S, Hattori N, Mizuno Y (1991) Distinct clustering of point mutations in mitochondrial DNA among patients with mitochondrial encephalomyopathies and with Parkinson's disease. Biochem Biophys Res Commun 176:938–946

Qian YP, Chu Z-T, Dai Q, Wei C-D, Chu JY, Tajima A, Horai S (2001) Mitochondrial DNA polymorphism in Yunnan nationalities in China. J Hum Genet 46:211–220

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441

Richards M, Côrte-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, Bandelt H-J, Sykes B (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260

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

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

Shinoda K, Kanai S (1999) Intracemetery genetic analysis at the Nakazuma Jomon site in Japan by mitochondrial DNA sequencing. Anthropol Sci 107:129–140

Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R, Jin L (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet 107:582–590

Su B, Xiao J, Underhill P, Deka R, Zhang W, Akey J, Huang W, Shen D, Lu D, Luo J, Chu J, Tan J, Shen P, Davis R, Cavalli-Sforza L, Chakraborty R, Xiong M, Du R, Oefner P, Chen Z, Jin L (1999) Y-chromosome evidence for a north-ward migration of modern humans into eastern Asia during the last ice age. Am J Hum Genet 65:1718–1724

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

Tsai LC, Lin CY, Lee JC, Chang JG, Linacre A, Goodwin W (2001) Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. Forensic Sci Int 119: 239–247

Wang L, Oota H, Saitou N, Jin F, Matsushita T, Ueda S (2000) Genetic structure of a 2,500-year-old human population in China and its spatiotemporal changes. Mol Biol Evol 17: 1396–1400

Wu R, Wu X, Zhang S (1989) Early humankind in China. Science Press, Beijing (in Chinese)

Yao Y-G, Lü X-M, Luo H-R, Li W-H, Zhang Y-P (2000*a*) Gene admixture in the silk road of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. Genes Genet Syst 75:173–178

Yao Y-G, Watkins WS, Zhang Y-P (2000*b*) Evolutionary history of the mtDNA 9-bp deletion in Chinese populations and its relevance to the peopling of East and Southeast Asia. Hum Genet 107:504–512

Zhang H, Ding M, Jiao Y, Wang X, Yan Z, Jin G, Meng X, Bai C, Lu Z, Chen R (1998) A dermatoglyphic study of the Chinese population III. Dermatoglyphics cluster of fifty-two nationalities in China. Acta Genetica Sinica 25:381–391 (in Chinese)

Zhao TM, Lee TD (1989) Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. Hum Genet 83:101–110