The Emerging Limbs and Twigs of the East Asian mtDNA Tree

Toomas Kivisild, Helle-Viivi Tolk,* Jüri Parik,* Yiming Wang,† Surinder S. Papiha,‡ Hans-Jürgen Bandelt,§ and Richard Villems**

*Department of Evolutionary Biology, Tartu University and Estonian Biocentre, Estonia; †Department of Medical Genetics, Sun Yat-Sen University of Medical Sciences, People's Republic of China; ‡Department of Human Genetics, University of Newcastle-upon-Tyne; and §Department of Mathematics, University of Hamburg, Germany

We determine the phylogenetic backbone of the East Asian mtDNA tree by using published complete mtDNA sequences and assessing both coding and control region variation in 69 Han individuals from southern China. This approach assists in the interpretation of published mtDNA data on East Asians based on either control region sequencing or restriction fragment length polymorphism (RFLP) typing. Our results confirm that the East Asian mtDNA pool is locally region-specific and completely covered by the two superhaplogroups M and N. The phylogenetic partitioning based on complete mtDNA sequences corroborates existing RFLP-based classification of Asian mtDNA types and supports the distinction between northern and southern populations. We describe new haplogroups M7, M8, M9, N9, and R9 and demonstrate by way of example that hierarchically subdividing the major branches of the mtDNA tree aids in recognizing the settlement processes of any particular region in appropriate time scale. This is illustrated by the characteristically southern distribution of haplogroup M7 in East Asia, whereas its daughtergroups, M7a and M7b2, specific for Japanese and Korean populations, testify to a presumably (pre-)Jomon contribution to the modern mtDNA pool of Japan.

Introduction

Previous analyses based on mtDNA (Ballinger et al. 1992; Wallace 1995), autosomal (Chu et al. 1998), and Y-chromosomal (Su et al. 1999) genetic loci have revealed regional clustering of northern and southern groups inhabiting East Asia and agree with the opinion that Asian populations derive from a recent out-of-Africa migration of modern humans followed by colonization of East Asia from south to north (Su and Jin 2000). Recently, this opinion has been contested by analyses using three human genetic marker systems (mtDNA, Y chromosome, and autosomal short tandem repeats) and a human born virus (Ding et al. 2000), suggesting that there is no support for a major north-south distinction in these markers, almost no structure in the mtDNA differences among regions, and that the observed frequency distribution in all three marker systems can be explained by simple isolation by distance.

In contrast to West Eurasia (Macaulay et al. 1999*b*), the Americas (Torroni et al. 1993*a;* Brown et al. 1998), and Siberia (Torroni et al. 1993*b;* Schurr et al. 1999), the mitochondria of East Asia (Japan, in particular) have not yet been classified satisfactorily. Eurasian mtDNAs belong to two superhaplogroups (''trunks'') M (Chen et al. 1995) and N (first defined in fig. 3 of Quintana-Murci et al. 1999 by one of its characteristic mutations and then baptized in fig. 2 of Alves-Silva et al. 2000), which differ at nucleotide position (np) 10400 (and further coding region sites). These trunks encompass the known Asian-specific haplogroups C, D, E, G, Z (''limbs'' and ''boughs'' of the M ''trunk'') and A, B, F, Y (limbs and boughs of the N trunk). Although high-resolution restriction fragment length polymorphisms (RFLPs) define all these haplogroups unambiguously except for Z (Starikovskaya et al. 1998; Schurr et al. 1999), their branching order within M and N, respectively, remained undetermined. Moreover, the nine haplogroups do not quite cover all East Asian mtDNAs, and toward the periphery of the mtDNA tree, potentially region-specific boughs and ''twigs'' that may be recognized by certain hypervariable segments (HVS)-I motifs in conjunction with single coding region sites still await full description.

We have therefore used hitherto nonsynthesized information concerning Asian complete mtDNA sequences (Yoneda et al. 1990; Ozawa et al. 1991; Jun, Brown, and Wallace 1994; Ikebe, Tanaka, and Ozawa 1995; Ozawa 1995; Nishino et al. 1996; Ingman et al. 2000; Shin et al. 2000; Tawata et al. 2000; Maca-Meyer et al. 2001), RFLPs (Ballinger et al. 1992; Torroni et al. 1993*b*, 1994; Starikovskaya et al. 1998; Schurr et al. 1999; Derbeneva et al. 2002), and HVS-I (and -II) sequences from Japan, Korea, central Asia, Southeast Asia, mainland China, and Taiwan (Horai and Hayasaka 1990; Horai et al. 1996; Kolman, Sambuughin, and Bermingham 1996; Lee et al. 1997; Comas et al. 1998; Pfeiffer et al. 1998; Seo et al. 1998; Nishimaki et al. 1999; Redd and Stoneking 1999; Qian et al. 2001), and have compared these data with a sample of 69 mtDNAs from southern China (Han Chinese), for which the two hypervariable segments (HVS-I, HVS-II) and specific coding region sites were analyzed. This allows us a better recognition of several limbs, which deeply branch off from the Eurasian mtDNA tree, as well as the identification of region-specific boughs and twigs that might testify to prehistoric settlement processes in the eastern fringe of Asia.

Methods

Subjects, Sample Preparation, and Sequence Analyses

After obtaining informed consent, 4–6 blood spots were collected on Guthrie cards from 69 healthy and

Abbreviations: RFLP, restriction fragment length polymorphism; CRS, Cambridge reference sequence; HVS-I (and -II), first (and second) hypervariable segment.

Key words: human mitochondrial DNA, population genetics, phylogeography.

Address for correspondence and reprints: Toomas Kivisild, Estonian Biocentre, Riia 23, Tartu 51010, Estonia. E-mail: tkivisil@ebc.ee.

Mol. Biol. Evol. 19(10):1737–1751. 2002

Table 1
mtDNA Variation in 69 Han Chinese from Southern China and in Five West Asian Controls

		Control Region Sequence[a]		
Sample	Haplogroup	HVS-1 (minus 16000)	HVS-II (73 263 in addition)	16519
CRS[b] . . .	H			T
1	B4	129 182C 183C 189 217 274 289 301 311	41 309d 310d (372)	C
2	B4	129 140 145 166 183C 189 217 274 335	146 150 309+CC 315+C	C
3	B4	140 168A 183C 189 217 311	150 315+C	C
4	B4	140 182C 183C 189 217 242A 274 335	146 150 315+C	C
5	B4	140 182C 183C 189 217 274 305T 335	150 195 309+CC 315+C	C
6	B4	182C 183C 189 217 362	309+CC 315+C	C
7	B4a	129 145 182C 183C 189 217 261	199 308d 309d	T
8	B4a	129 182C 183C 189 217 261	308d 309d 310d	C
9	B4a	129 182C 183C 189 217 261 354 357	309d 310d	T
10	B4a	129 182C 183C 189 261	309+C 315+C	C
11	B4a	150 182C 183C 189 217 240 261	195 309+C 315+C	T
12	B4a	154 182C 183C 189 217 240 261	309+C	T
13	B4a	182C 183C 189 217 221 240 261	150 309+C 315+C	T
14	B4a	182C 183C 189 217 248 261 295 (494)	150 309+C	C
15	B4a	93 182C 183C 189 217 261	146 204 309+CC	C
16	B4a	93 182C 183C 189 217 261 344	315+C	C
17	B4b	136 179 182C 183C 189 217	150 204 207 250G 309+CC 315+C	C
18	B4b	136 183C 189 217	315+C	C
19	B4b	136 183C 189 217 309 354	195 207 309+C 315+C	C
20	B4b	136 189 217 260 287 325	315+C	C
21	B5b	111 126 140 183C 189 234 243	103 309+CC 315+C	C
22	D	184 189 223 311 362	146 152 315+C	T
23	D4	92 218 223 362	315+C	C
24	D4	209 223 266 362	194 309+C 315+C	C
25	D4	223 362	207 310	T
26	D4	95 209 223 362	153 315+C	T
27	D4	201 223 362	152 315+C	T
28	D5	189 223 362	150 309+C 315+C	C
29	D5	182C 183C 189 223 319 360 362	150 309+CC 315+C	T
30	D5	182C 183C 189 223 362	150 194 309+CC 315+C	T
31	D5	182C 183C 189 223 362	150 194 309+CC 315+C	T
32	E	223 362 (390)	152 315+C	C
33	F	129 218 304 311	150 249d 315+C	T
34	F1a	129 172 304	195 249d 309+C 315+C	C
35	F1a	129 162 172 304 335	249d 251 315+C	C
36	F1a	129 172 304	52 53 54C 71d 249d 309+CC 315+C 318	C
37	F1a	108 129 162 172 304	249d 315+C	C
38	F1a	129 162 172 292 304	152 249d	C
39	F1a	129 172 304	249d 309+C	C
40	F1a	129 162 172 304 (399)	249d 309+C 315+C	C
41	F1a	129 172 304	249d 309+C 315+C	C
42	F1a	108 129 162 172 214 304	165 249d 315+C	C
43	F1a	129 162 172 304 335	249d 251 309+C	C
44	F1a	108 129 162 172 274 304	93 95C 249d 315+C	C
45	F1a	172 304 311	249d 315+C	C
46	F1b	183C 189 249 300 304	150 195 249d 309+C	T
47	F1c	111 129 243 266 304	152 249d 309+C 315+C	C
48	F2	CRS	235 249d 315+C	T
49	F2	209 304	249d 309+C 315+C	T
50	F2a	92A 291 304	249d 315+C	T
51	G2	223 311 362	152 309+C 315+C	T
52	M	223 234 287 290 362	125 127 128 309+CC 315+C 318	T
53	M10	93 129 223 311 (497)	315+C	T
54	M10	129 223 311	315+C	T
55	M7	CRS	146A 199 204 309+C 315+C	C
56	M7b	223 297	150 199 204 309+C	T
57	M7b	223 297	150 152 199 204 315+C	T
58	M7b	102 223 297 300	150 199 204 249d 315+C	T
59	M7b	38 129 223 297	n.d.	
60	M7b1	129 192 223 297	150 199 309+C 315+C 341	T
61	M7b1	129 192 223 243 297	150 199 315+C	T
62	M7c	129 223 295	146 199 309+C	C
63	M8a	184 189 223 298 311 319 (390)	309+C 315+C	T

Table 1
Continued

Sample	Haplogroup	Control Region Sequence[a] HVS-1 (minus 16000)	HVS-II (73 263 in addition)	16519
64	M8a	184 223 298 319	152 309+C 315+C	T
65	N	69 172 223 278 291A 298 325 362	150 152 199 309+C 315+C	T
66	N9a	223 257A 261 311	150 309+CC 315+C	T
67	R	183C 189 311 365	165 185 189 309+CC	C
68	R9a	93 111 192 249 298 355 362 (390)	207 249d 309+C 315+C	T
69	W	111 172 189 223 311 362	185 189 195 229 234 309+C 315+C	C
	A4	222 223 249 290 319 362	n.d.	
	Y	126 231 266	n.d.	
	W	223 292	n.d.	
	N1a	147G 172 223 248	n.d.	
	N	111 153 223 319	n.d.	

Coding Region Sequence and RFLP Variation[c]

Sample	Haplogroup	663e	1004o	1715c	3010	4830n	4833	4850	4883	4913	5176a	5178	5231	5263	5301	5319	5351f	5417	5446y	6455w	7025a	7853o	7598	7598f	7828f	7600	7933j
CRS[b]	H	−	+	+	G	−	A	C	C	A	+	C	G	C	A	A	−	G	−	+	−	+	G	+	−	G	−
1	B4																		−								
2	B4																										
3	B4																										
4	B4																		−								
5	B4																		−		+						
6	B4																										
7	B4a																		−								
8	B4a																		+								
9	B4a																		−								
10	B4a																		−								
11	B4a																		−								
12	B4a																		−								
13	B4a																		−								
14	B4a																		+								
15	B4a																		+								
16	B4a																		+								
17	B4b			+	G						+	C	G	C	A			−	G	−							
18	B4b																		−								
19	B4b																		−								
20	B4b																		−								
21	B5b																		−								
22	D				G	−	A	C	T	A	−	A	G	C	A	A	−	G	−			+		+			
23	D4				A	−	A	C	T	A	−	A	G	C	A	A	−	G						+			
24	D4				A	−					−	A	G	C	A	A	−	G		+							
25	D4				A	−	A	C	T	A	−	A	G	C	A	A	−	G						+			
26	D4				A	−	A	C	T	A	−	A	G	C	A	A	−	G					G	+		G	
27	D4				A	−	A	C	T	G	−	A	G	C	A	A	−	G						+			
28	D5				G	−	A	C	T	A	−	A	G	C	G	A	−	G						+			
29	D5				G	−	A	C	T	A	−	A	G	C	G	A	−	G						+			
30	D5				G	−					−												+			
31	D5				G	−					−	A	G	C	G	A	−	G						+			
32	E				G	−	A	C	C	A	+	C	G	C	A	A	−	G				+	A	−		G	
33	F		+																						−		
34	F1a		+																						−		
35	F1a																										
36	F1a		+																						−		
37	F1a																										
38	F1a																										
39	F1a																										
40	F1a																										
41	F1a																										
42	F1a																										
43	F1a																										
44	F1a																										
45	F1a																										

Table 1
Continued

Sample	Haplogroup	Coding Region Sequence and RFLP Variation[c]																														
		663e	1004o	1715c	3010	4830n	4833	4850	4883	4913	5176a	5178	5231	5263	5301	5319	5351f	5417	5464y	6455w	7025a	7853o	7598	7598f	7828f	7600	7933j
46	F1b	−	+																						−		
47	F1c																										
48	F2	−	−																		+				+		
49	F2		−																		+				+		
50	F2a		−																		+				+		
51	G2				G	+	G	C	C	A	+	C	G	C	A	A	−	G				+	G	+		G	
52	M				G	−	A	C	C	A	+	C	G	C	A	A	−	G				+	G			G	
53	M					−	A	C	C	A	+	C	G	C	A	A	−	G		+		+					
54	M					−	A	C	C	A	+	C	G	C	A	A	−	G		+		+					
55	M7					−	A	T	C	A	+	C	G	C	A	A	−	G	−	−	+	+					
56	M7b					−											+									
57	M7b					−											+			−		−				
58	M7b					−											+									
59	M7b					−											+			−		−				
60	M7b1					−											+			−		−				
61	M7b1					−											+			−		−				
62	M7c					−					+	C	G	C	A	A	−	G		−		+					
63	M8a					−	A	C	C	A	+	C	G	C	A	A	−	G		+			G			G	
64	M8a					−	A	C	C	A	+	C	G	C	A	A	−	G		+							
65	N	−		+							+	C	G	C	A	A	−	G						+			
66	N9a	−		+							+	C	A	T	A	A	−	A									
67	R	−	+																−		+				−		
68	R9a																		−		+						
69	W	−		+							+	C	G	C	A	A	−	G						+			
	A4	+		+							+	C	G	C	A	A	−	G									
	Y	−		+							+	C	G	C	A	A	−	A									+
	W	−		+							+	C	G	C	A	A	−	G									
	N1a	−		−							+	C	G	C	A	A	−	G									
	N	−		+							+	C	G	C	A	A	−	G									

c) Coding Region Sequence and RFLP Variation[c]

Sample	Haplo-group	9bp	8148e	8249b	8391e	8414x	8701	8994	9052n	9540r	9820g	10373	10394c	10397a	10397	10398	10400	10464	10871z	11465u	11583	12372	12406o	12705w	13259o	14953	15043	15079	15094	15235i
CRS[b]	H	2	−	−	+	−	A	G	+	−	−	G	−	−	A	A	C	−	+	+	A	G	+	+	+	C	G	A	C	−
1	B4	1							+	−			−																	
2	B4	1							+	−			−																	−
3	B4	1							+				−																	
4	B4	1							+	−			−																	
5	B4	1							+	−			−	−						+			+							
6	B4	1							−				−																	
7	B4a	1							+	−			−																	−
8	B4a	1							+	−			−																	−
9	B4a	1							+	−			−																	
10	B4a	1							+	−			−																	−
11	B4a	1							+	−			+																	−
12	B4a	1							+				+																	
13	B4a	1							+				+																	
14	B4a	1							+				−																	
15	B4a	1							+	−			−																	−
16	B4a	1							+	−			−																	
17	B4b	1							+	−			−																	−
18	B4b	1							+	−			−																	−
19	B4b	1							+				−						+					+						−
20	B4b	1							+	−			−																	
21	B4b	1							+				+																	−
22	D	2				−			+	+	−		+	+	A	G	T								+					
23	D4	2				+			+	+			+	+	A	G	T							−	+					
24	D4	2							+	+	−		+	+	A	G	T													

Table 1
Continued

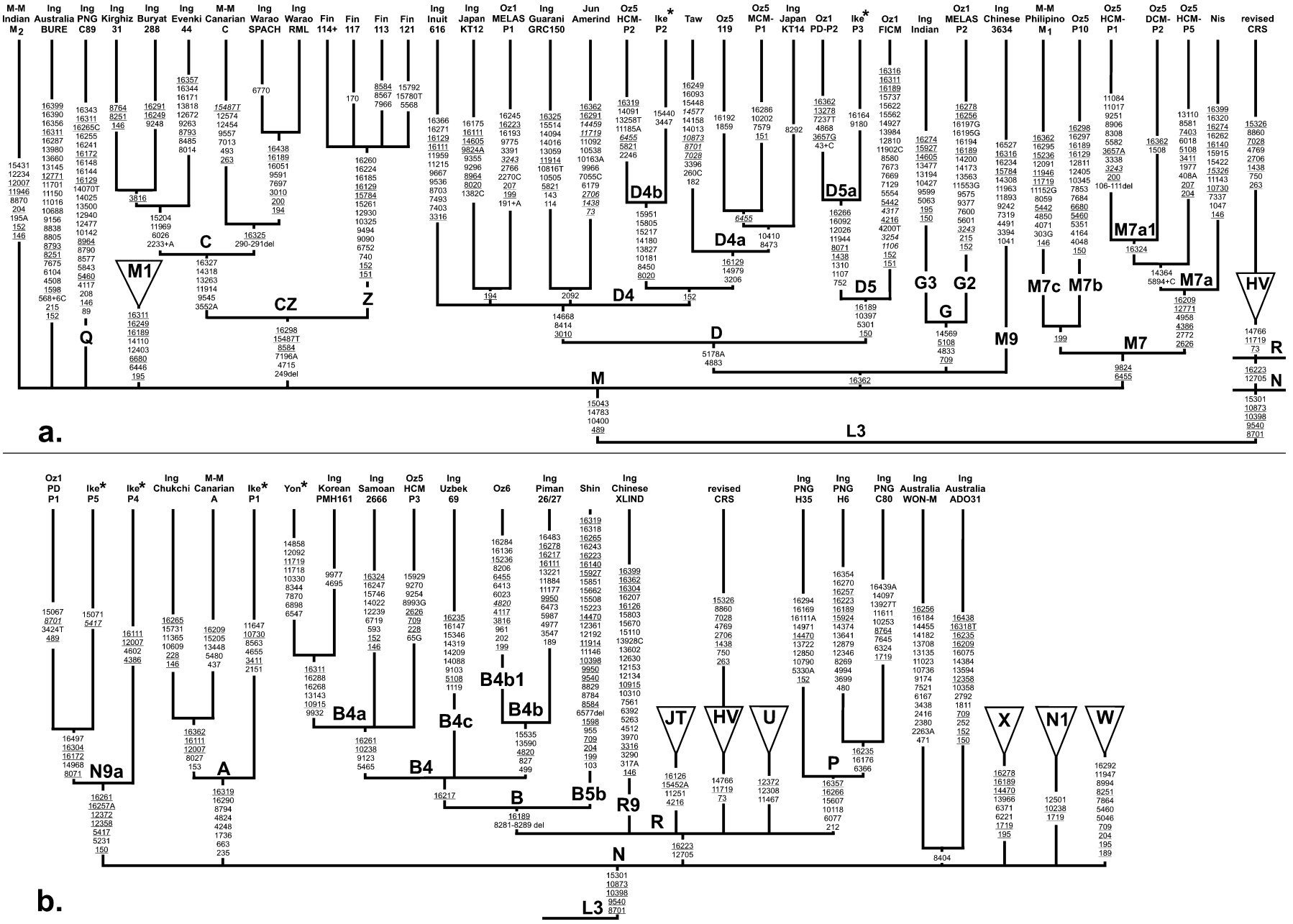
c) Coding Region Sequence and RFLP Variation[c]

Sample	Haplo-group	9bp	8148e	8249b	8391e	8441x	8701	8994	9052n	9540r	9820g	10373	10394c	10397a	10397	10398	10400	10646	10871z	11465u	12358	12372	12406o	12705w	13259o	14953	15043	15079	15094	15235i
25	D4	2				+			+	+			+	+	A	G	T								+					
26	D4	2				+			+	+			+	+	A	G	T								+					
27	D4	2				+			+	+			+	+	A	G	T								+					
28	D5	2				−			+	+		G	−	−	G	G	T								+					
29	D5	2				−							−	−	G	G	T							−	+					
30	D5	2				−			+	+			−	−	G	G	T								+					
31	D5	2				−			+	+			−	−	G	G	T								+					
32	E	2				−			+	+			+	+	A	G	T							−	+					
33	F	2							+				−	−									+							
34	F1a	2							−	−			−	−									−	+						
35	F1a	2	−						−	−			−	−									−							
36	F1a	2							−				−	−									−							
37	F1a	2	+						−				−	−									−							
38	F1a	2							−				−	−									−							
39	F1a	2											−	−					+				−	+						
40	F1a	2											−	−									−	+						
41	F1a	2							−				−	−									−							
42	F1a	2	+						−	−			−	−									−							
43	F1a	2							−				−	−									−							
44	F1a	2	+						−	−			−	−									−							
45	F1a	2	−						−				−										−							
46	F1b	2	−						+	−			−	−									−	+						
47	F1c	2	−						−	−			−	−									−							
48	F2	2	−						+				−	−					+				+							
49	F2	2	−						+	−			−	−					+				+							
50	F2a	2							+	−			−	−					+				+							
51	G2	2				−			+	+			+	+	A	G	T								+					
52	M	2				−			+	+	−		+	+	A	G	T							−						
53	M	2				−			+	+	−		+	+	A	G	T	+												
54	M	2				−			+	+	−		+	+	A	G	T	+												
55	M7	2				−			+	+	+		+	+	A	G	T			+			+							
56	M7b	2				−					+		+	+	A	G	T													
57	M7b	2				−			+	+	+		+	+	A	G	T													
58	M7b	2				−			+	+	+		+	+	A	G	T													
59	M7b	2				−			+	+	+		+	+	A	G	T													
60	M7b1	2				−			+	+	+		+	+	A	G	T							−						
61	M7b1	2				−			+	+	+		+	+	A	G	T													
62	M7c	2				−	G	G	+	+	+		+	+	A	G	T													
63	M8a	2				−			+	+	−		+	+	A	G	T													
64	M8a	2				−			+	+	−		+	+	A	G	T													
65	N	2		−			A	G	+	−			−	−	A	A	C		+					−	+	C	G	A	C	
66	N9a	2		−			A	G	+	−		G	−	−	A	A	C		+		G	A		−		C	G	G	T	
67	R	2							+	−			+	−						+			+	+						
68	R9a	2							+				−	−					+	+				+						
69	W	2		−			A	A	+	−		A	−	−	A	A	C		+					−	+	T	G	A	C	
...	A4			−			A	G		−			−	−					+		A	G		−		C	G	A	C	
...	Y			−	+		A	G		−			−	−					+		A	G		−		C	G	A	C	
...	W			+			A	A		−			−	−					+		A	G		−		C	G	A	C	
...	N1a			−			A	G		−			−	−					+		A	G		−						
...	N			−			A	G		−			−	−					+		A	G		−						

[a] Two hypervariable segments of the control region were sequenced: HVS-I (16024–16383) and HVS-II (16511–00360). Nucleotide change is specified for transversions; d = deletion; + = insertion. Length variation in the C stretch (16184–16193) is not shown. Mutations that were detected outside the specified range are shown in parentheses.

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

[c] Restriction-endonuclease sites are indicated as follows: a = *Alu*I, b = *Ava*II, c = *Dde*I, e = *Hae*III, f = *Hha*I, g = *Hin*fI, i = *Msp*I, j = *Mbo*I, k = *Rsa*I; n = *Hae*II, o = *Hinc*II, r = *Bfa*I, u = *Mse*I, w = *Mbo*II, x = *Ssp*I, y = *Hph*I, z = *Mnl*I. Haplogroup assignments are according to Figure 2.


M-M Indian M2
Ing Australia BURE
Ing PNG C89
Ing Kirghiz 31
Ing Buryat 288
Ing Evenki 44
M-M Canarian C
Ing Warao SPACH
Ing Warao RML
Fin 114+
Fin 117
Fin 113
Fin 121
Ing Inuit 616
Ing Japan KT12
Oz1 MELAS P1
Ing Guarani GRC150
Jun Amerind
Oz5 HCM-P2
Ike* P2
Taw
Oz5 119
Oz5 MCM-P1
Ing Japan KT14
Oz1 PD-P2
Ike* P3
Oz1 FICM
Ing Indian
Oz1 MELAS P2
Ing Chinese 3634
M-M Philipino M1
Oz5 P10
Oz5 HCM-P1
Oz5 DCM-P2
Oz5 HCM-P5
Nis
revised CRS
M1
C
CZ
Z
Q
D4b
D4a
D4
D5a
D5
D
G3
G2
G
M9
M7c
M7b
M7
M7a1
M7a
HV
R
N
M
L3
a.
Oz1 PD P1
Ike* P5
Ike* P4
Ing Chukchi
M-M Canarian A
Ike* P1
Yon*
Ing Korean PMH161
Ing Samoan 2666
Oz5 HCM P3
Ing Uzbek 69
Oz6
Ing Piman 26/27
Shin
Ing Chinese XLIND
revised CRS
Ing PNG H35
Ing PNG H6
Ing PNG C80
Ing Australia WON-M
Ing Australia ADO31
N9a
A
B4a
B4
B4c
B4b1
B4b
B
B5b
R9
R
N
JT
HV
U
P
X
N1
W
L3
b.

maternally unrelated Cantonese from southern China, Guangdong Province, and five West Asian controls from Turkey. Blood spots were perforated into 4–5 circles (3 mm diameter) for DNA extraction. Circles were soaked in 1 ml of H_2O for 30 min, treated with methanol for 10 min, immersed in a 200-μl solution (5 mM NaOH, 2 mM EDTA·Na_2) and covered with mineral oil. The samples were heated (100°C) for 6–7 min, then chilled on ice, and centrifuged. The DNA was purified with DNA zol® BD reagent (Gibco-BRL). Selected regions of mtDNA were PCR amplified in short fragments (200–700 bp; primer sequences are available from the authors on request), and purified products were sequenced with a DYEnamic™ ET terminator cycle sequencing kit (Amersham Pharmacia Biotech) and analyzed on an ABI 377 DNA Sequencer. Where possible (see table 1), RFLP tests were performed to assay phylogenetic signals revealed either by the analysis of complete sequences or available high-resolution RFLP studies. Sequences were aligned and analyzed with the Genetics Computer Group (GCG) Wisconsin Package. Control region sequences have been submitted to the EMBL data library under accession nos. AJ401470–AJ401608.

Phylogenetic Analyses

Published complete mtDNA sequences were aligned manually, and a phylogenetic tree was inferred combining information available from published RFLP data (see fig. 1). The unrooted tree comprising Asian complete mtDNA sequences was rooted in African L3 (cf. Alves-Silva et al. 2000, fig. 2). A network of haplogroup M7 HVS-I haplotypes, comprising the superposition of the most parsimonious trees for the three postulated sets of M7a, M7b, and M7c sequences (plus respective ancestral types) was constructed by departing from the median-joining network (Bandelt, Forster, and Röhl 1999) and then adding additional edges to capture all equally parsimonious solutions. The age of mtDNA clades is calculated (for a plausible tree within this network) from the observed transitions in positions 16090–16365 according to Forster et al. (1996), with standard deviation estimated as in Saillard et al. (2000). Some of the (early) HVS-I sequences of Horai et al. (1996) do not cover the variation preceding 16129, but the potential minor effect on the age estimation was ignored.

The haplogroup is to be understood as a monophyletic clade in the rooted mtDNA tree, i.e., a group of haplotypes that comprises all descendants of their most recent common ancestor, as inferred from the shared mutations (Torroni et al. 2000). When speaking about nested haplogroups, e.g., M encompasses D, D includes D4, or N encompasses R (Macaulay et al. 1999*b*), R includes R9, etc., we use informal terms reflecting the branching order of a tree: trunk > limb > bough > twig. We reserve the name trunk for the ubiquitous ancient Eurasian superhaplogroups M, N and its subgroup R, whereas, at the other extreme, twigs signify typically region-specific young haplogroups (of age ~10,000 years). In clade naming, we follow the RFLP-based nomenclature started by Torroni et al. (1993*a*) and the hierarchical scheme set up by Richards et al. (1998). Because subgroup names, such as A1, A2, B1, B2, D1, D2, and D3 have been used in the context of Native American mtDNAs (Forster et al. 1996, 1997; Saillard et al. 2000), we let the subgrouping of A–D start with number 4 to avoid confusion. For M, the numbering of new subhaplogroups will begin with 7 because M1 is already defined (Quintana-Murci et al. 1999) and M2–M6 signify potential M subgroups of India (Kivisild et al. 1999*b*). To keep the distinction between West and East Eurasian subhaplogroups of N and R transparent, the East Asian branches will receive numbers close to 10.

Results

More than 60 complete mtDNA sequences have been published, mainly in the field of medical genetics, before the 53 complete mtDNA sequences of Ingman et al. (2000) and others published recently (Finnilä, Lehtonen, and Majamaa 2000; Finnilä et al. 2001; Maca-Meyer et al. 2001). The former, however, suffer from numerous problems, so that additional confirmation of characteristic sites and comparison with the RFLP and control region database for consistency is indispensable

←

FIG. 1.—An inferred tree linking complete non-African human mtDNA sequences, which is rooted in African L3 (see fig. 2 in Alves-Silva et al. 2000). Superhaplogroups M (*a*) and N (*b*) cover all sequences. Abbreviations and sources of the sequences are as follows: Japan: Ike—Ikebe, Tanaka, and Ozawa (1995), Ing Japan—Ingman et al. (2000), Nis—Nishino et al. (1996), Oz1—Ozawa et al. (1991), Oz5—Ozawa et al. (1995), Shin—Shin et al. (2000), Taw—Tawata et al. (2000), Yon—Yoneda et al. (1990); Mainland Asia: Ing Buryat, Chinese, Chukchi, Indian, Kirghiz, Korean, Uzbek—Ingman et al. (2000); India: M-M Indian—Maca-Meyer (2001); Philippines: M-M Philip—Maca-Meyer (2001); Oceania: Ing Australia, Samoa, PNG (Papua New Guinea)—Ingman et al. (2000); Native American: Ing Guarani, Piman, Warao—Ingman et al. (2000), Jun Amerind—Jun, Brown, and Wallace (1994), M-M Canarian—Maca-Meyer et al. (2001); Europe: Fin Finn—Finnilä et al. (2001); revised CRS—Andrews et al. (1999); individuals are coded as in the original publications. All mtDNA sequences from those publications are shown except for the partial B sequence Oz1 MERFF (the variant sites of which are nearly all covered by the Yon sequence) and the European complete mtDNA sequences. The latter all fall into the West Eurasian haplogroups JT, HV, U, W, X, and N1. For these haplogroups and the African bough, M1 (Quintana-Murci et al. 1999), of M ancestral sequences are reconstructed (indicated by triangles), using the information from Finnilä (2000, 2001), Levin, Cheng, and Reeder (1999), and Maca-Meyer et al. (2001). Positions are numbered according to the CRS, and all nucleotide changes are transitions unless a suffix indicates a transversion (to base A, G, C, or T) or a deletion (del) or an insertion (+). Length polymorphisms in the C stretches of the control region and for CA repeats at around 514, heteroplasmic mutations, and variation at 16519 are not displayed. Recurrent mutations are underlined. Problematic mutations are mainly potential sequencing errors or omissions (Kivisild and Villems 2000) and together with known disease causing mutations are shown in italics. Sequences marked by an asterisk evidently lack some control region information, so that their branching points with regard to HVS-I and HVS-II have been inferred from the corresponding haplogroup motifs where possible (Alves-Silva et al. 2000).

(Macaulay, Richards, and Sykes 1999*a;* Quintana-Murci et al. 1999; Kivisild and Villems 2000). Figure 1 summarizes our editing efforts, with particular focus on the East Asian (and Native American) lineages and their position in the worldwide mtDNA tree. The reconstructed ancestral sequences for M and N as well as R can be regarded as quite reliable because the corresponding nodes in the tree are multifurcation points. There is evidence that the node at which the trunks M and N diverge is the root of other African offshoots (Ingman et al. 2000). From figure 1 in conjunction with table 1 we obtain the major internal branches of the East Asian mtDNA tree (bold links in fig. 2). The remaining branches in figure 2 are then inferred with the additional help of published RFLP and control region data as well as coding region sequences (Herrnstadt et al. 2002). One should bear in mind that the encoding and characterization of the more peripheral segments of the mtDNA tree, defined by no more than two complete sequences, is not yet absolutely stable, but is rather intended to direct future sequencing efforts.

In the Asian-specific trunk M the combination of the currently available data (see fig. 2) suggests that the limbs D, G, and M9 may all stem from a common node which is distinguished from the ancestral node of the M trunk solely by C at position 16362. The branching point, however, is somewhat ambiguous because position 16362 is highly variable (Hasegawa et al. 1993) and the boughs of G differ at it. The limb M9 is defined by a transition at 4491 (as inferred from the data of Herrnstadt et al. 2002) and thus carries the boughs E and M9a (previously referred to as M9 by Yao et al. 2002). The next limb of M, called M7 here, is well supported by complete sequences: it is defined by transitions at 6455 and 9824, the latter of which is recognized by the gain of a *Hin*fI site at 9820. The limb M7 branches further into three boughs, M7a, M7b, and M7c (see fig. 3). The bough M7b was already identified by the loss of a *Hin*cII site at 7853 and distinguished as a group by Ballinger et al. (1992). M7c is not yet exactly characterized by coding region sites and is difficult to distinguish from other M subgroups on the basis of HVS-I alone. Another major limb, M8 (Yao et al. 2002), is defined by transitions at 4715, 8584, 16298 and transversions at 7196 and 15487. Its principal boughs are the sister haplogroups C, Z, and a new haplogroup, M8a, that is characterized by transitions at 14470, 16184, and 16319 (Yao et al. 2002). M also has a major limb, named Q by Forster et al. (2001), in Papua New Guinea as well as at least one limb in Australia.

The N trunk of East Asia has three major branches: R, which may itself be called a trunk because it constitutes a founder superhaplogroup for all Eurasia (cf. Macaulay et al. 1999*b*), then the limb A, and a new branch defined by a transition at 5417 that we baptize N9, carrying the Y bough and a not previously described bough N9a. The latter is characterized by three further mutations in the coding region as well as a rare transversion in HVS-I (see fig. 2). Further branches of N are found in Australia. Almost all East Asian R lineages belong to the two major limbs B and F. These limbs together with another limb, called P by Forster et al. (2001), virtually encompass all non-M mtDNAs found in Papua New Guinea. Originally, F was defined by the loss of the (highly mutable) *Hin*cII site at position 12406 (Torroni et al. 1994). It is now becoming evident, however, that this marker defines only one particular branch of a deeper branching limb of similar geographic spread. Haplogroup R9 includes both F and R9a and is defined by the deletion of one A at np 248/249 and a transition at np 10310 (Yao et al. 2002). This deletion in HVS-II is a fairly rare but not unique event because it occurs in parallel also in CZ (see fig. 2) and has been found in few haplogroup U sequences in Europe (Finnilä, Lehtonen, and Majamaa 2000). We propose to expand the range of F by letting it include all R9 mtDNAs with C at 16304 but anticipating potential instances of back mutation. The previous F is now named F1. It has a sister branch, F2, defined by two coding region mutations that can be detected by 14-enzyme RFLPs (inferred from Torroni et al. 1994 and table 1) and positions 10535 and 10586 (Yao et al. 2002). F has a potential sister branch, called R9a, that is distinguished by three HVS-I transitions (inferred from Betty et al. 1996 and table 1) and a coding region site 10320 (Yao et al. 2002). Within F1, F2, and B, further branches are inferred from the published data (see fig. 2). It seems that the Native American mtDNAs of haplogroup B belong to B4b (defined by transitions at 499, 827, 13590, and 15535) and could form a sister clade to B4b1 (characterized by 16136, previously referred to as B4b by Yao et al. 2002).

To clarify the positions of the East Asian haplogroups stemming from the pan-Eurasian trunk N, we also analyzed five West Asian mtDNAs belonging to haplogroups A, Y, W, and N1a (defined by transitions at 1719, 10238, 12501, 16172, and a transversion at 16147), and an unspecified offshoot of N. It turns out (see table 1) that the East and West Asian limbs of N do not share any mutations, which implies an early split of the West Eurasian, South Asian (Kivisild et al. 1999*a*, 1999*b*), and East Asian founder lineages. For haplogroup R we may anticipate a similar situation.

It should be emphasized that the abundant HVS-I information alone, without support from coding region sites, may be quite misleading in a number of cases. For instance, HVS-I sequences bearing the transition at 16295 may belong to either M7c or a novel M subgroup (characterized by transitions at 200, 215, 318, and 326 in HVS-II; see Yao et al. 2002). The distinction of A4 and A5 from A* currently hinges upon a single hypervariable HVS-I site for each. More problematically, a HVS-I sequence differing from the Cambridge reference sequence (CRS) only at 16223 and 16362 (by transitions) can belong to any of the haplogroups D, E, G; the seeming R motif of transitions at 16245 and 16362 in fact characterizes a bough of D4 (cf. sequence MP1 in Ozawa et al. 1991); the CRS (belonging to haplogroup H) may even match M and F sequences within HVS-I (see 33 and 37 in table 1). Comparing just short stretches of HVS-I sequences could therefore easily create the impression of specific European and East Asian mtDNA affinities, e.g., as asserted by figure 3 in Wang

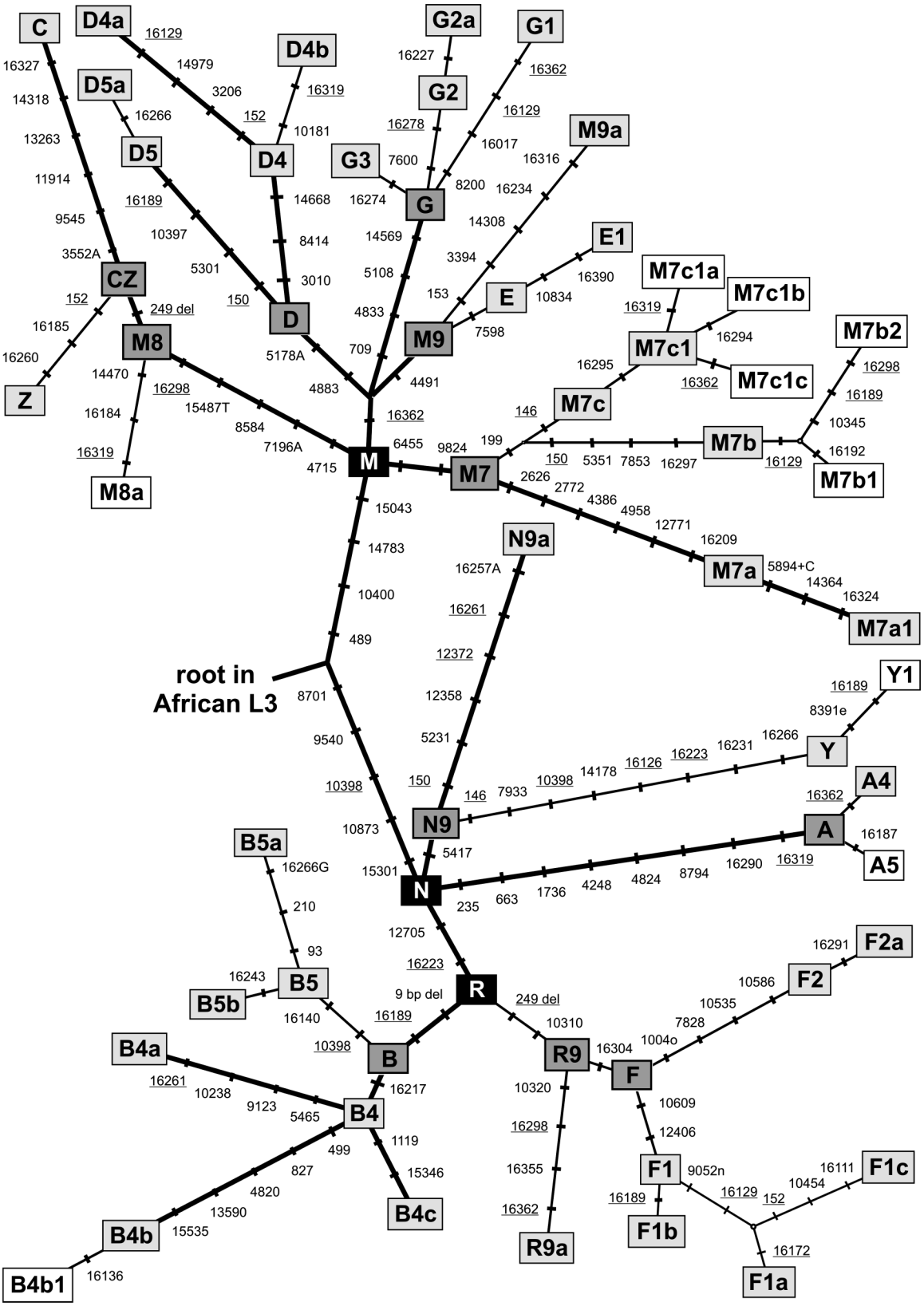


FIG. 2.—The major branches of the Asian mtDNA tree displaying the East Asian–specific haplogroups. The tree is rooted in haplogroup L3 (see fig. 2 in Alves-Silva et al. 2000). The positions at which two nodes differ are listed along links (in arbitrary order); nucleotide change is specified for transversions by suffixes, and ''del'' indicates deletion; recurrent mutations are underlined. Restriction fragment length polymorphism sites, where the exact nucleotide change is not known, are indicated as follows: e = *Hae*III; n = *Hae*II; o = *Hin*cII. Length polymorphisms in C stretches and variation at 16519 are disregarded. Bold links are defined by at least two complete sequences. Each haplogroup is indicated by its ancestral haplotype. Informal distinction between the trunks (black), limbs (dark gray), boughs (gray), and twigs (white background boxes) is according to the main text.

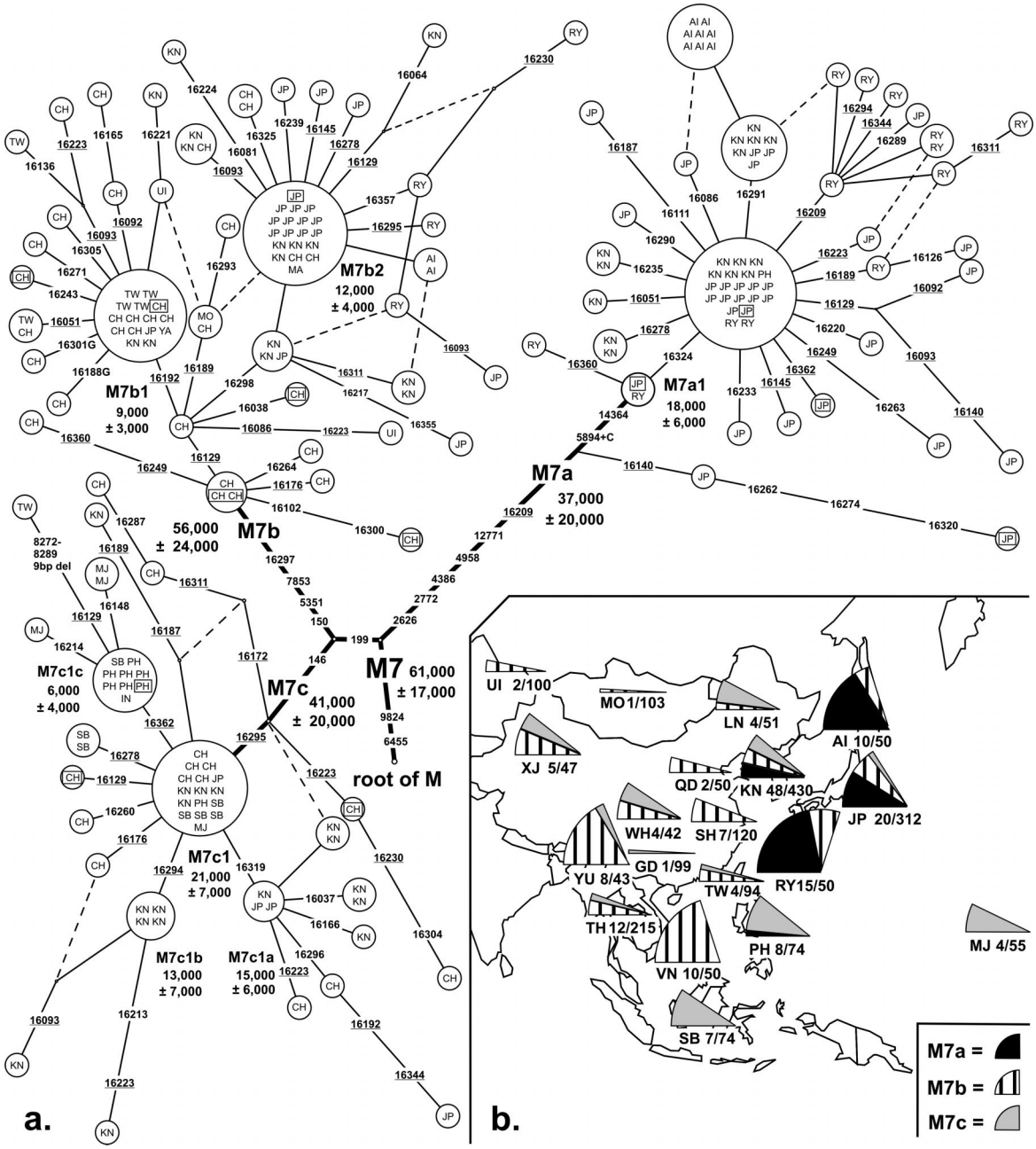


FIG. 3.—Phylogenetic reconstruction and geographic distribution of haplogroup M7. *a,* A network of HVS-I haplotypes, which comprises the superposition of the most parsimonious trees for the three postulated sets of M7a, M7b, and M7c sequences. The mutations along the bold links were only analyzed for a few Japanese sequences (Ozawa et al. 1991; Ozawa 1995; Nishino et al. 1996) and—toward the root of M—for some Chinese sequences (this study): the corresponding individuals with (partial) coding region information are boxed. Numbers along links indicate transitions; recurrent HVS-I mutations are underlined. The age of mtDNA clades is calculated (along the tree indicated by unbroken lines) according to Forster et al. (1996), with standard errors estimated as in Saillard et al. (2000). Sample codes (and sources): AI—Ainu (Horai et al. 1996); CH—Chinese (Betty et al. 1996; Nishimaki et al. 1999; Qian et al. 2001; Yao et al. 2002; this study); IN—Indonesian (Redd and Stoneking 1999); JP—Japanese (Ozawa et al. 1991; Ozawa 1995; Horai et al. 1996; Nishino et al. 1996; Seo et al. 1998; Nishimaki et al. 1999); KN—Koreans (Horai et al. 1996; Lee et al. 1997; Pfeiffer et al. 1998); MA—Mansi (Derbeneva et al. 2002); MJ—Majuro (Sykes et al. 1995); MO—Mongolians (Kolman, Sambuughin, and Bermingham 1996); PH—Philippines (Sykes et al. 1995; Maca-Meyer 2001); RY—Ryukyuans (Horai et al. 1996); SB—Sabah (Sykes et al. 1995); TW—Taiwanese Han (Horai et al. 1996) and aboriginals (Melton et al. 1998); UI—Uighur (Comas et al. 1998; Yao et al. 2000); YA—Yakuts (Derenko and Shields 1997). *b,* Frequencies of the subgroups of M7 in Asian populations are inferred from the preceding HVS-I as well as partial HVS-I and RFLP data (VN—Vietnamese: Ballinger et al. 1992; Lum et al. 1998). Mainland Han Chinese are denoted as follows: GD—Guangdong, LN—Liaoning, QD—Qingdao, WH—Wuhan, XJ—Xinjiang, YU—Yunnan (Yao et al. 2002), SH—Shanghai (Nishimaki et al. 1999). The number of M7 sequences in relation to the sample size is indicated under each pie slice proportional to the M7 frequency.

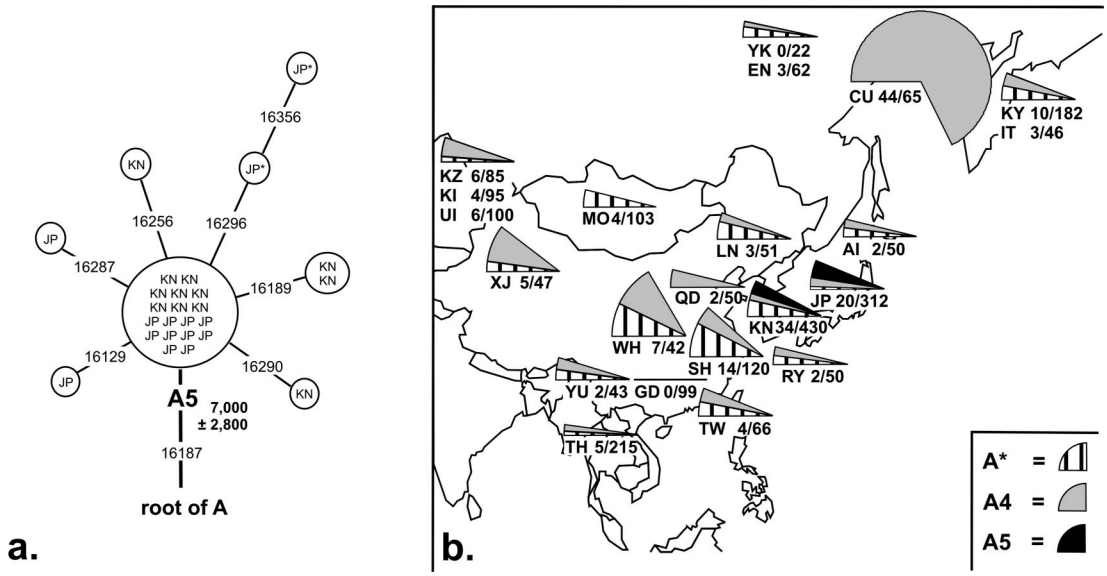


FIG. 4.—Phylogeography of haplogroup A. *a,* The most parsimonious tree of all Asian HVS-I haplotypes bearing transitions at 16187, 16223, 16290, and 16319 relative to CRS. Sequences marked by an asterisk are partial HVS-I sequences. *b,* Frequencies of A*, A4, and A5 in Asian populations inferred from HVS-I sequences. Sample codes (and sources): AI—Ainu (Horai et al. 1996); CU—Chukchis (Starikovskaya et al. 1998); EN—Evens (Derenko and Shields 1997); GD—Guangdong, Han Chinese (Yao et al. 2002; this study); IT—Itelmen (Schurr et al. 1999); JP—Japanese (Horai et al. 1996; Seo et al. 1998; Nishimaki et al. 1999); KI—Kirghiz (Comas et al. 1998); KN—Koreans (Horai et al. 1996; Lee et al. 1997; Pfeiffer et al. 1998); KY—Koryaks (Derenko and Shields 1997; Schurr et al. 1999); KZ—Kazakhs (Comas et al. 1998); LN—Liaoning, Han Chinese (Yao et al. 2002); MO—Mongols (Kolman, Sambuughin, and Bermingham 1996); QD—Qingdao, Han Chinese (Yao et al. 2002); RY—Ryukyuans (Horai et al. 1996); SH—Shanghai, Han Chinese (Nishimaki et al. 1999); TH—Thais (Fucharoen, Fucharoen, and Horai 2001); TW—Taiwanese Han (Horai et al. 1996); UI—Uighurs (Comas et al. 1998); WH—Wuhan, Han Chinese (Yao et al. 2002); YK—Yakutians (Derenko and Shields 1997; Schurr et al. 1999); XJ—Xinjiang, Han Chinese (Yao et al. 2002); YU—Yunnan, Han Chinese (Yao et al. 2002). The number of A sequences in relation to the sample size is indicated under each pie slice proportional to the A frequency.

et al. (2000). Their Linxi data likely encompass several haplotypes from haplogroup B4, the ancestral type of which cannot be distinguished from CRS within the short HVS-I stretch sequenced. Any subdivision of mtDNA into clusters based on HVS-I sequences alone, as attempted by Horai et al. (1996), thus runs into serious problems. For instance, only 10 of the 18 clusters proposed by Horai et al. (1996) turn out to be monophyletic (as long as back mutations would not cause trouble): their clusters C1, C6, C11, C14, C17, and C18 by and large correspond to B4, A, D5, M8, M7b2, and M7b1, respectively, whereas C1, C4, C7, and C15 constitute subclades of Y1, F1a, G2a, and M7a1, respectively. The potentially paraphyletic clusters C5, C8, C9, and C10 would essentially cover D4 but inevitably also include some other M haplotypes. The remaining four clusters (C2, C9, C12, and C16) constitute poly- or paraphyletic groups.

Although considerable differences in the geographic distribution of the Asian-specific haplogroups can already be revealed by comparing RFLP profiles of Siberian and Southeast Asian populations (Wallace 1995), a finer resolution is obtained when extending the marker list on the mtDNA molecule. Haplogroups A, C, D, G, Y, and Z almost completely cover the mtDNA pool of Northeast Asians, whereas in Southeast Asians C, Y, or Z mtDNAs have rarely been found, but instead haplogroups B and F are predominant. N9a is, compared with its sister bough Y, widely spread, although at very low frequencies, among most East Asian populations considered here. Considering the geographic distribution of the boughs and twigs we see further regional patterns. In contrast to A4, which is widely spread, the A5 twig, with its low diversity suggesting shallow time depth, is specific to Koreans and Japanese (see fig. 4). Similarly, B4 is the prevailing bough in haplogroup B (see fig. 2), covering all haplogroup B types in Native Americans and Polynesians. B5 is found most frequent, accounting for about one third to one half of the B types, in eastern China, Korea, and Japan (Horai and Hayasaka 1990; Horai et al. 1996; Seo et al. 1998; Nishimaki et al. 1999). D4 seems to be the predominant bough of D (see fig. 2), but as a whole it cannot be identified without coding region markers. D5 is characterized by a transition at 16189 in conjunction with the reversal of the RFLP marker for M (caused by a transition at 10397); it is most frequent in southern China but rare or absent in Central Asians and Siberians. E1 is so far found only in Southeast Asia (Ballinger et al. 1992; our table 1); what was called E in the Tibetan data (Torroni et al. 1994) seems to be mixed with G2a (as signalized by 16227), and what was labeled E in the Asian mtDNA tree of Herrnstadt et al. (2002, fig. 2) turns out to constitute a mixed bag of M1, M7c, M9a, and real E haplotypes. F1a is the main branch of F (see fig. 2) in Southeast Asia, whereas F1b is more frequent in Central Asians and Mongols, Koreans, and Japanese. F2, being much less frequent, may have a wide geographic distri-

bution, as judged from the few occurrences of F2a. It seems that G1 is restricted to Northeast Siberia. G2a is highest among Central Asians (8.8%) and also above 3% in Tibetans and Ainu and rare or absent among southern Chinese, Vietnamese, island Southeast Asians (including Ryukyuans), and Siberians. G3 is not yet well screened, but evidently it is seen in Korea, Mongolia, and Central Asia. Y1 seems to be restricted to Northeast Asian populations and Ainu.

Haplogroup M7, although characteristic for East Asian populations, has not been found in the northeast of the continent (Torroni et al. 1993*b;* Derenko and Shields 1997; Starikovskaya et al. 1998; Schurr et al. 1999). It is also very rare in Central Asians (Torroni et al. 1994; Kolman, Sambuughin, and Bermingham 1996; Comas et al. 1998). This haplogroup has been detected so far in China and Vietnam, the Korean peninsula and Japanese islands, as well as among Mongols, the West Siberian Mansi, and island Southeast Asia. Koreans possess lineages from both the southern and the northern haplogroup complex and share M7a with Japanese, Ainu, and Ryukyu islanders. The geographic specificity of the boughs and twigs of M7 (see fig. 2) is most intriguing: M7c1c is specific to island Southeast Asia and M7b1 is of Chinese provenance, whereas M7a, M7b2, and M7c1b are found almost exclusively in Korea and Japan. In fact, M7 is one of the prevailing haplogroups not only among Japanese (of Honshu and Kyushu) but also for Ainu and Ryukyuans, thus testifying to a common genetic background. There is very little haplotype sharing in M7 across the distinguished populations except for the ancestral types of the (named) nested clades (see fig. 3); in particular, no single type is shared between Ainu and Ryukyuans (Horai et al. 1996).

In summary, we see characteristic regional features at several levels of the mtDNA phylogeny that testify to geographic structure (cf. Yao et al. 2002). The tentative interpretation of this structuring is best accomplished when taking additional information about climatic conditions and archaeological findings into consideration, so that the genetic data can be used to address hypotheses about early or recent prehistoric events in East Asia.

Discussion

Population histories drawn from genetic data do not necessarily always correspond to those drawn from linguistic and archaeological studies. On the other hand, analysis of mtDNA lineages in the broader cultural and geological contexts might be of help in distinguishing early settlement patterns in East Asia. The estimated coalescence times for the subclades of M7a, M7b, and M7c range between 6,000 and 18,000 years (see fig. 3), which suggests that some of these starlike clades (with star indices >0.9; Torroni et al. 1998) could reflect a (re-)settlement process in the area around the (southern) Japanese Sea after the Last Glacial Maximum (LGM), contemporary with the spread of microblades, e.g., of the Suyanggae-type (Imamura 1996), and before the onset of the Jomon culture. This does not exclude the possibility that M7a and M7b entered Japan with the pioneer settlers more than 30,000 years ago and were bottlenecked toward the LGM. In any case, one has to reckon with a long presence of M7 haplotypes in the coastal areas next to or south of the Yellow Sea (most of which fell dry around the LGM).

In contrast to M7, the sharing of haplogroups A5, B5, C, F1a, N9a, and Z between Koreans and Japanese and their virtual absence in Ryukyuans and Ainu indicate later migrations and contacts between Korea and Japan. This is probably mainly due to the influx of Yayoi people (2,300 years ago) who brought agriculture and the Japanic language (Janhunen 1996) to Japan. It is then remarkable in view of those events that the variation seen in M7 is not shuffled between Koreans and Japanese, quite in contrast to B4. This could point to the considerable geographic substructure of the Korean mtDNA pool at that time (paralleling the probable linguistic situation (Janhunen 1996), so that the migrants carried only a limited number of M7 types (perhaps M7c1 types and the ancestral type of M7a together with a one-step descendant).

The presence of Y1 lineages among Ainu points to another migration route, namely, from the native Siberian populations to the northernmost populations of the Japanese islands. This mitochondrial connection fits well with the archaeological record (Imamura 1996) and classical anthropological data (Hanihara 1998). The peopling of Japan can therefore be seen as a complex process with the early pioneer settlement and several recent migrations which affected the resident populations differentially. By and large, the mtDNA findings support the opinions of Hanihara (1993) and Kazumichi (1993).

The now emerging picture of the East Asian mtDNA tree that incorporates complete mtDNA information helps to shed light on prehistoric human migrations to and within the eastern belt of Asia. Quite consistent with the Y-chromosome data (Su et al. 1999), mtDNA analyses show that although East Asian haplotypes are regionally specific, they all derive (together with the West Eurasian haplotypes) ultimately from one or two ancestral lineages of African origin. Subdividing the hierarchy of the mtDNA tree into limbs, boughs, and twigs is instrumental for telling apart the settlement processes of any particular region in appropriate time scale. We have demonstrated by way of example that the limbs and boughs, geographically widespread, reflect the earliest events in the settlement of East Asia, whereas the twigs of greater geographic specificity can be of value to uncover more recent events. The broad claim that in East Asia ''there is . . . almost no structure in the mtDNA differences among regions'' and that ''mtDNA phylogeny . . . could not reveal much of interest about population history here'' (Ding et al. 2000) is clearly unjustified and rather reflects insufficient resolution of the used mtDNA markers and lack of phylogenetic reasoning.

Acknowledgments

We thank Peter Forster, Max Ingman, Vincent Macaulay, Masato Tawata, Shintaroh Ueda, and Wee Soo

Shin for helpful information and Ille Hilpus and Jaan Lind for technical assistance. This work was supported by an exchange program between the Universities of Tartu and Hamburg, by the Citrina Foundation, and an Estonian Science Foundation grant to R.V.

LITERATURE CITED

ALVES-SILVA, J., M. DA SILVA SANTOS, P. E. GUIMARÃES, A. C. FERREIRA, H.-J. BANDELT, S. D. PENA, and V. F. PRADO. 2000. The ancestry of Brazilian mtDNA lineages. Am. J. Hum. Genet. **67**:444–461.

ANDREWS, R. M., I. KUBACKA, P. F. CHINNERY, R. N. LIGHTOWLERS, D. M. TURNBULL, and N. HOWELL. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. **23**:147.

BALLINGER, S. W., T. G. SCHURR, A. TORRONI, Y. Y. GAN, J. A. HODGE, K. HASSAN, K. H. CHEN, and D. C. WALLACE. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics **130**:139–152.

BANDELT, H.-J., P. FORSTER, and A. RÖHL. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. **16**:37–48.

BETTY, D. J., A. N. CHIN-ATKINS, L. CROFT, M. SRAML, and S. EASTEAL. 1996. Multiple independent origins of the COII/tRNA(Lys) intergenic 9-bp mtDNA deletion in aboriginal Australians. Am. J. Hum. Genet. **58**:428–433.

BROWN, M. D., S. H. HOSSEINI, A. TORRONI, H.-J. BANDELT, J. C. ALLEN, T. G. SCHURR, R. SCOZZARI, F. CRUCIANI, and D. C. WALLACE. 1998. mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? Am. J. Hum. Genet. **63**:1852–1861.

CHEN, Y.-S., A. TORRONI, L. EXCOFFIER, A. S. SANTACHIARA-BENERECETTI, and D. C. WALLACE. 1995. Analysis of mtDNA variation in African populations identifies the most ancient of all human continent-specific haplogroups. Am. J. Hum. Genet. **57**:133–149.

CHU, J. Y., W. HUANG, S. Q. KUANG et al. (14 co-authors). 1998. Genetic relationship of populations in China. Proc. Natl. Acad. Sci. USA **95**:11763–11768.

COMAS, D., F. CALAFELL, E. MATEU et al. (12 co-authors). 1998. Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. Am. J. Hum. Genet. **63**:1824–1838.

DERBENEVA, O. A., E. B. STARIKOVSKAYA, D. C. WALLACE, and R. SUKERNIK. 2002. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am. J. Hum. Genet. **70**:1009–1014.

DERENKO, M. V., and G. F. SHIELDS. 1997. Diversity of mitochondrial DNA nucleotide sequences in three groups of aboriginal inhabitants of Northern Asia. Mol. Biol. (Mosk) **31**: 784–789.

DING, Y. C., S. WOODING, H. C. HARPENDING et al. (11 co-authors). 2000. Population structure and history in East Asia. Proc. Natl. Acad. Sci. USA **97**:14003–14006.

FINNILÄ, S., I. E. HASSINEN, L. ALA-KOKKO, and K. MAJAMAA. 2001. Phylogenetic network of the mtDNA haplogroup U in Northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. Am. J. Hum. Genet. **66**:1017–1026.

FINNILÄ, S., M. S. LEHTONEN, and K. MAJAMAA. 2000. Phylogenetic network for European mtDNA. Am. J. Hum. Genet. **68**:1475–1484.

FORSTER, P., R. HARDING, A. TORRONI, and H.-J. BANDELT. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. Am. J. Hum. Genet. **59**:935–945.

———. 1997. Reply to Bianchi and Bailliet. Am. J. Hum. Genet. **61**:246–247.

FORSTER, P., A. TORRONI, C. RENFREW, and A. RÖHL. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol. Biol. Evol. **18**:1864–1881.

FUCHAROEN, G., S. FUCHAROEN, and S. HORAI. 2001. Mitochondrial DNA polymorphisms in Thailand. J. Hum. Genet. **46**:115–125.

HANIHARA, K. 1993. The population history of the Japanese. Nippon Ronen Igakkai Zasshi **30**:923–931.

———. 1998. Reanalysis of local variations in the Ainu crania. Anthropol. Sci. **106**:1–15.

HASEGAWA, M., A. DI RIENZO, T. D. KOCHER, and A. C. WILSON. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. J. Mol. Evol. **37**:347–354.

HERRNSTADT, C., J. L. ELSON, E. FAHY et al. (11 co-authors). 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.

HORAI, S., and K. HAYASAKA. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am. J. Hum. Genet. **46**:828–842.

HORAI, S., K. MURAYAMA, K. HAYASAKA, S. MATSUBAYASHI, Y. HATTORI, G. FUCHAROEN, S. HARIHARA, K. S. PARK, K. OMOTO, and I. H. PAN. 1996. mtDNA polymorphism in East Asian populations, with special reference to the peopling of Japan. Am. J. Hum. Genet. **59**:579–590.

IKEBE, S., M. TANAKA, and T. OZAWA. 1995. Point mutations of mitochondrial genome in Parkinson's disease. Brain Res. Mol. Brain Res. **28**:281–295.

IMAMURA, K. 1996. Prehistoric Japan. UCL Press Ltd, London.

INGMAN, M., H. KAESSMANN, S. PÄÄBO, and U. GYLLENSTEN. 2000. Mitochondrial genome variation and the origin of modern humans. Nature **408**:708–713.

JANHUNEN, J. 1996. Manchuria. An ethnic history. The Finno-Ugrian Society, Helsinki.

JUN, A. S., M. D. BROWN, and D. C. WALLACE. 1994. A mitochondrial DNA mutation at nucleotide pair 14459 of the NADH dehydrogenase subunit 6 gene associated with maternally inherited Leber hereditary optic neuropathy and dystonia. Proc. Natl. Acad. Sci. USA **91**:6206–6210.

KAZUMICHI, K. 1993. The Japanese as an Asia-Pacific population. Pp. 19–30 *in* D. DENOON, M. HUDSON, G. MCCORMACK, and T. MORRIS-SUZUKI, eds. Multicultural Japan: palaeolithic to postmodern. Cambridge University Press, Cambridge, U.K.

KIVISILD, T., M. J. BAMSHAD, K. KALDMA et al. (15 co-authors). 1999*a*. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr. Biol. **9**: 1331–1334.

KIVISILD, T., K. KALDMA, M. METSPALU, J. PARIK, S. S. PAPIHA, and R. VILLEMS. 1999*b*. The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World. Pp. 135–152 *in* R. DEKA and S. S. PAPIHA, eds. Genomic diversity. Kluwer-Academic-Plenum Publishers, New York.

KIVISILD, T., and R. VILLEMS. 2000. Questioning evidence for recombination in human mitochondrial DNA. Science **288**: 1931.

KOLMAN, C., N. SAMBUUGHIN, and E. BERMINGHAM. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics **142**:1321–1334.

LEE, S., C. SHIN, K. KIM, Y. LEE, and J. LEE. 1997. Sequence variation of mitochondrial DNA control region in Koreans. Forensic Sci. Int. **87**:99–116.

LEVIN, B. C., H. CHENG, and D. J. REEDER. 1999. A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection. Genomics **55**:135–146.

LUM, J. K., R. L. CANN, J. J. MARTINSON, and L. B. JORDE. 1998. Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. Am. J. Hum. Genet. **63**:613–624.

MACA-MEYER, N., A. M. GONZÁLEZ, J. M. LARRUGA, C. FLORES, and V. C. CABRERA. 2001. Major genomic mitochondrial lineages delineate early human expansions. BMC Genetics **2**:13.

MACAULAY, V., M. RICHARDS, and B. SYKES. 1999*a*. Mitochondrial DNA recombination—no need to panic. Proc. R. Soc. Lond. B. Biol. Sci. **266**:2037–2039.

MACAULAY, V. A., M. B. RICHARDS, E. HICKEY, E. VEGA, F. CRUCIANI, V. GUIDA, R. SCOZZARI, B. BONNÉ-TAMIR, B. SYKES, and A. TORRONI. 1999*b*. The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am. J. Hum. Genet. **64**:232–249.

MELTON, T., S. CLIFFORD, J. MARTINSON, M. BATZER, and M. STONEKING. 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am. J. Hum. Genet. **63**:1807–1823.

NISHIMAKI, Y., K. SATO, L. FANG, M. MA, H. HASEKURA, and B. BOETTCHER. 1999. Sequence polymorphism in the mtDNA HV1 region in Japanese and Chinese. Legal Med. **1**:238–249.

NISHINO, I., A. SEKI, Y. MAEGAKI, K. TAKESHITA, S. HORAI, I. NONAKA, and Y. GOTO. 1996. A novel mutation in the mitochondrial tRNA(Thr) gene associated with a mitochondrial encephalomyopathy. Biochem. Biophys. Res. Commun. **225**:180–185.

OZAWA, T. 1995. Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. Biochem. Biophys. Acta **1271**:177–189.

OZAWA, T., M. TANAKA, H. INO et al. (14 co-authors). 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.

PFEIFFER, H., R. STEIGHNER, R. FISHER, H. MÖRNSTAD, C. L. YOON, and M. M. HOLLAND. 1998. Mitochondrial DNA extraction and typing from isolated dentin-experimental evaluation in a Korean population. Int. J. Legal. Med. **111**:309–313.

QIAN, Y. P., Z.-T. CHU, Q. DAI, C.-D. WEI, J. Y. CHU, A. TAJIMA, and S. HORAI. 2001. Mitochondrial DNA polymorphism in Yunnan nationalities in China. J. Hum. Genet. **46**: 211–220.

QUINTANA-MURCI, L., O. SEMINO, H.-J. BANDELT, G. PASSARINO, K. MCELREAVEY, and A. S. SANTACHIARA-BENERECETTI. 1999. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat. Genet. **23**:437–441.

REDD, A. J., and M. STONEKING. 1999. Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am. J. Hum. Genet. **65**:808–828.

RICHARDS, M. B., V. A. MACAULAY, H.-J. BANDELT, and B. C. SYKES. 1998. Phylogeography of mitochondrial DNA in western Europe. Ann. Hum. Genet. **62**:241–260.

SAILLARD, J., P. FORSTER, N. LYNNERUP, H.-J. BANDELT, and S. NØRBY. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am. J. Hum. Genet. **67**:718–726.

SCHURR, T. G., R. I. SUKERNIK, Y. B. STARIKOVSKAYA, and D. C. WALLACE. 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.

SEO, Y., B. STRADMANN-BELLINGHAUSEN, C. RITTNER, K. TAKAHAMA, and P. M. SCHNEIDER. 1998. Sequence polymorphism of mitochondrial DNA control region in Japanese. Forensic Sci. Int. **97**:155–164.

SHIN, W. S., M. TANAKA, J. SUZUKI, C. HEMMI, and T. TOYO-OKA. 2000. A novel homoplasmic mutation in mtDNA with a single evolutionary origin as a risk factor for cardiomyopathy. Am. J. Hum. Genet. **67**:1617–1620.

STARIKOVSKAYA, Y. B., R. I. SUKERNIK, T. G. SCHURR, A. M. KOGELNIK, and D. C. WALLACE. 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.

SU, B., and L. JIN. 2000. Natives or immigrants: modern human origin in East Asia. Nat. Rev. **1**:126–133.

SU, B., J. XIAO, P. UNDERHILL et al. (21 co-authors). 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last ice age. Am. J. Hum. Genet. **65**:1718–1724.

SYKES, B., A. LEIBOFF, J. LOW-BEER, S. TETZNER, and M. RICHARDS. 1995. The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am. J. Hum. Genet. **57**:1463–1475.

TAWATA, M., J. I. HAYASHI, K. ISOBE, E. OHKUBO, M. OHTAKA, J. CHEN, K. AIDA, and T. ONAYA. 2000. A new mitochondrial DNA mutation at 14577 T/C is probably a major pathogenic mutation for maternally inherited type 2 diabetes. Diabetes **49**:1269–1272.

TORRONI, A., H.-J. BANDELT, L. D'URBANO et al. (11 co-authors). 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am. J. Hum. Genet. **62**:1137–1152.

TORRONI, A., J. A. MILLER, L. G. MOORE, S. ZAMUDIO, J. ZHUANG, T. DROMA, and D. C. WALLACE. 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., M. RICHARDS, V. MACAULAY, P. FORSTER, R. VILLEMS, S. NØRBY, M. L. SAVONTAUS, K. HUOPONEN, R. SCOZZARI, and H.-J. BANDELT. 2000. mtDNA haplogroups and frequency patterns in Europe. Am. J. Hum. Genet. **66**: 1173–1177.

TORRONI, A., T. G. SCHURR, M. F. CABELL, M. D. BROWN, J. V. NEEL, M. LARSEN, D. G. SMITH, C. M. VULLO, and D. C. WALLACE. 1993*a*. Asian affinities and continental radiation of the four founding Native American mtDNAs. Am. J. Hum. Genet. **53**:563–590.

TORRONI, A., R. I. SUKERNIK, T. G. SCHURR, Y. B. STARIKORSKAYA, M. F. CABELL, M. H. CRAWFORD, A. G. COMUZZIE, and D. C. WALLACE. 1993*b*. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am. J. Hum. Genet. **53**:591–608.

WALLACE, D. 1995. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. Am. J. Hum. Genet. **57**:201–223.

WANG, L., H. OOTA, N. SAITOU, F. JIN, T. MATSUSHITA, and S. UEDA. 2000. Genetic structure of a 2,500-year-old human population in China and its spatiotemporal changes. Mol. Biol. Evol. **17**:1396–1400.

YAO, Y.-G., Q.-P. KONG, H.-J. BANDELT, T. KIVISILD, and Y.-P. ZHANG. 2002. Phylogeographic differentiation of mito-

chondrial DNA in Han Chinese. Am. J. Hum. Genet. **70**: 635–651.

YAO, Y.-G., X.-M. LÜ, H.-R. LUO, W.-H. LI, and Y.-P. ZHANG. 2000. Gene admixture in the Silk Road region of China—evidence from mtDNA and melanocortin 1 receptor polymorphism. Genes Genet. Syst. **75**:173–178.

YONEDA, M., Y. TANNO, S. HORAI, T. OZAWA, T. MIYATAKE, and S. TSUJI. 1990. A common mitochondrial DNA mutation in the t-RNA(Lys) of patients with myoclonus epilepsy associated with ragged-red fibers. Biochem. Int. **21**:789–796.

JEFFERY LONG, reviewing editor

Accepted May 27, 2002