Hum Genet (2003) 113 : 391–405
DOI 10.1007/s00439-003-1004-7

ORIGINAL INVESTIGATION

**Qing-Peng Kong · Yong-Gang Yao · Mu Liu
Shu-Ping Shen · Cai Chen · Chun-Ling Zhu
Malliya Gounder Palanichamy · Ya-Ping Zhang**

Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China

Received: 10 March 2003 / Accepted: 11 July 2003 / Published online: 21 August 2003

Abstract To study the mitochondrial DNA (mtDNA) polymorphisms in a total of 232 individuals from five ethnic populations (Daur, *n*=45; Ewenki, *n*=47; Korean, *n*=48; Mongolian, *n*=48; Oroqen, *n*=44) in northern China, we analyzed the control region sequences and typed for a number of characteristic mutations in coding regions (especially the region 14576–16047), by direct sequencing or restriction-fragment-length-polymorphism (RFLP) analysis. With the exception of 14 individuals belonging to the European-specific haplogroups R2, H, J, and T, the mtDNAs considered could be assigned into the East Asian-specific haplogroups described recently. The polymorphisms in cytochrome *b* sequence were found to be very informative for defining or supporting the haplogroups status of East Asian mtDNAs in addition to the reported regions 10171–10659 and 14055–14590 in our previous study. The haplogroup distribution frequencies varied in the five ethnic populations, but in general they all harbored a large amount of north-prevalent haplogroups, such as D, G, C, and Z, and thus were in agreement with their ethnohistory of northern origin. The two populations (Ewenki and Oroqen) with small population census also show concordant features in their matrilineal genetic structures, with lower genetic diversities observed.

Q.-P. Kong · Y.-G. Yao · C.-L. Zhu · M. G. Palanichamy
Y.-P. Zhang (✉)
Laboratory of Cellular and Molecular Evolution,
Kunming Institute of Zoology, Chinese Academy of Sciences,
650223 Kunming, China
Tel.: +86-871-5198993, Fax: +86-871-5195430,
e-mail: zhangyp1@263.net.cn

Q.-P. Kong · Y.-P. Zhang
Laboratory for Conservation and Utilization of Bio-resource,
Yunnan University, 650091 Kunming, China

Q.-P. Kong · Y.-G. Yao
Graduate School of the Chinese Academy of Sciences,
100039 Beijing, China

M. Liu · S.-P. Shen · C. Chen
Department of Pathology, Baotou Medical College,
014010 Baotou, Inner Mongolia, China

Introduction

China stands at a geographical crossroads between Southeast Asia, Central Asia, and Siberia, and consists of a very important region for tracing the migration and expansion of the anatomically modern human to these areas. Although a number of recent studies using different genetic markers have shed some light on the peopling in East Asia, especially in China (Ding et al. 2000; Karafet et al. 2001; Kivisild et al. 2002; Yao et al. 2000b, 2002a), debates have not been compromised. One of the major problems causing the controversy is due to the fact that the samples considered were not fully matched: populations from southern China were much more sampled than populations from northern China (c.f. Ding et al. 2000). Among the available mtDNA data of Chinese ethnic populations (Oota et al. 2002; Yao et al. 2000a, 2002a, 2002b, 2003a; Yao and Zhang 2002), relatively fewer populations from northern China were analyzed compared with those from southern and central China. To better understand the migration pattern of East Asia, more data from northern China is thus necessary.

Based on a satisfactorily resolved East Asian mtDNA phylogeny that was constructed either by combining information provided by control region and coding region (Yao et al. 2002a) or by complete sequences (Kivisild et al. 2002; Kong et al. 2003), we dissected the mtDNA lineages identified in 232 subjects in five ethnic groups sampled from northern China into different haplogroups, then compared the genetic structure of these populations on the basis of distribution frequency of each haplogroup. We followed the strategy for haplogroup classification as described in Yao et al. (2002a, 2003a) and Kivisild et al. (2002). In brief, the mtDNAs were tentatively assigned into respective haplogroups according to their specific mutations observed in the hypervariable segment I (HVS-I) and by (near-)matching with the reported data with coding region information available, then, other specific mutations in the hypervariable segment II (HVS-II) and/or coding region were typed to further characterize the haplogroup status of the mtDNAs. In addition, the region 14576–

Fig. 1 Geographic locations of the five ethnic populations in northern China

16047 [according to the revised Cambridge reference sequence (rCRS; Andrews et al. 1999)], which covering the complete cytochrome *b* gene (Cyt *b*), was sequenced in a total of 63 individuals [25 from this study and 38 from our previous studies (Yao et al. 2002a, 2002b, 2003a) as well as our unpublished data set], with an intention to learn whether this region is as informative as regions 10171–10659 and 14055–14590 in Yao et al. (2002a) for defining and/or supporting the haplogroups of East Asian mtDNAs. Our results showed that there were differences in the maternal genetic structures of the five ethnic populations considered, and the information provided by the region (14576–16047) was helpful for discerning 15 (sub-)haplogroups.

Material and methods

Sampling

A total of 232 individuals from five ethnic populations were sampled in Inner Mongolia, China: 45 Daurs from Ewenkizu Zizhiqi county, 47 Ewenkis from Xin Barag Zuoqi county, 48 Koreans from Arunqi county, 48 Mongolians from Xin Barag Zuoqi county, and 44 Oroqens from Oroqen Zizhiqi county (Fig. 1). All of the individuals were confirmed to be unrelated before sampling and were given informed consent.

DNA amplification and sequencing

The mtDNA HVS-I sequence was amplified and sequenced in all the samples using the procedures described elsewhere (Yao et al. 2000a, 2002a, 2002b). In order to confirm the haplogroup status of some individuals, a number of characteristic mutations in HVS-II sequence and/or coding regions were further detected by direct sequencing or RFLP analysis using the same primer pairs and conditions as in Yao et al. (2002a). Moreover, all of the individuals were screened for the mtDNA 9-bp deletion in the COII/tRNA^{Lys} intergenic region according to our previous studies (Yao et al. 2000b, 2001).

The region 14576–16047, which covers the whole Cyt *b* gene sequence and harbors two characteristic polymorphisms (14783 and 15043) of macro-haplogroup M, was amplified by using primer pair L14575 (5′-ACCCGACCACACCGCTAACA-3′)/H16048 (5′-GTCAATACTTGGGTGGTACC-3′). After being purified on spin columns (Watson BioTechnologies, Shanghai), the PCR products were directly sequenced for both strands by using the two primers for amplification and six internal primers (L14752, 5′-ACTACAAGAACACCAATGACC-3′; L14989, 5′-ATGGCTGAATCATCCGCTAC-3′; L15391, 5′-TAGGAATCACCTCCCATTCC-3′, L15598, 5′-ACACAATTCTCCGATCCGTC-3′; H15086, 5′-AGGAGGATAATGCCGATGTT-3′; H15400, 5′-TGTAGTAAGGGTGGAAGGTG-3′). The numbers in the primer names refer to the position of the 3′ end of the primer sequence relative to rCRS (Andrews et al. 1999). L and H stand for light and heavy strands, respectively.

Data analyses

The sequences were edited and aligned by DNAstar software and compared with the rCRS (Andrews et al. 1999). The length polymorphisms of the A and C stretches in region 16180–16193 (triggered by the 16189 T/C substitution) were disregarded in the analysis.

The mtDNAs were classified into the specific (sub-)haplogroups by using the strategy and haplogroup annotation system as fully described in recent studies (Richards et al. 2000; Kivisild et al. 2002; Yao et al. 2002a, 2003a; Kong et al. 2003). After each mtDNA was assigned into the most-derived named haplogroup, the haplogroup distribution frequencies in each of the five populations were then estimated. We also compared the haplogroup distribution pattern in our samples with that of the recently reported data from Siberia (Buryat and Yakut; Pakendorf et al. 2003). The haplotype diversity and nucleotide diversity (Nei 1987) in the populations were computed by using the DnaSP package (Rozas and Rozas 1999). In addition, using the information provided by the complete sequences of the Cyt *b* genes analyzed in this study and in the reported complete sequences (Kivisild et al. 2002; Kong et al. 2003), we reanalyzed the recently reported Cyt *b* data from Koreans (Lee et al. 2002) and tried to pinpoint the potential errors in their sequence data.

Nomenclature

Gene mutation nomenclature used in this article follows the recommendations of den Dunnen and Antonarakis (2001). Gene symbols used in this article follow the recommendations of the HUGO Gene Nomenclature Committee (Povey et al. 2001). The authors have made every attempt to perform the study in accordance with the recommendations made by Cooper et al. (2002).

Results

Haplogroup identification

The mtDNA sequence variation in the 232 individuals was listed in Table 1. It should be noted that all of the samples, with the exception of two individuals, could be classified into the most-derived named mtDNA haplogroups (Fig. 2). The two mtDNAs that could not be assigned further were labeled M and N, respectively. The one M mtDNA from the Korean sample (Kor92), with mutation motif 16145-16148-16188-16189-16223, might belong to a new (sub-) haplogroup of M that is still not defined. This motif can

Table 1 MtDNA variation in 232 individuals from five ethnic populations in northern China. Positions are numbered according to the revised Cambridge reference sequence (rCRS) of Andrews et al. (1999); the mtDNAs that have no mutations in a sequenced region compared with the reference sequence are labeled as CRS

Sample[a]	Haplo-group	HVS-I (16001–16497)[b] (16000+)	HVS-II (30–407)[b] (73 and 263 in addition)	5176*Alu*I[c]	4831*Hha*I[c]	9-bp[d]	Other polymorphisms
DW13	D	223 270 362		–		2	
DW14	D	223 271 362		–		2	
DW16	D	223 362		–		2	
DW18	D	223 245 311 362 368		–		2	
DW29	D	223 245 362 368		–		2	
DW37	D	223 245 362 368		–		2	
DW38	D	223 245 362 368		–		2	
Oro17	D	145 223 362 368		–		2	
Oro18	D	145 223 362 368		–		2	
Oro2	D	223 291 362		–		2	
Oro20	D	223 274 362		–		2	
Oro22	D	145 223 362 368		–		2	
Oro32	D	223 319 362		–		2	
Oro36	D	223 291 362		–		2	
Oro4	D	223 291 362		–		2	
Oro40	D	223 319 362		–		2	
Oro42	D	223 319 362		–		2	
Oro43	D	223 270 362		–		2	
Oro5	D	223 291 362		–		2	
Oro8	D	223 291 362		–		2	
Oro9	D	223 291 362		–		2	
Mg212	D	223 249 362	152 309+C 315+C	–		2	
Mg218	D	223 320 335 362	279 315+C	–		2	
Mg222	D	172 362	194 315+C	–	–	2	
Mg224	D	174 223 362	309+C 315+C	–	–	2	
Mg227	D	093 129 223 249 362	146 152 196 309+C 315+C	–		2	
Mg230	D	223 362	185 309+C 315+C	–		2	
Mg234	D	218 223 295 362	195 309+C 315+C	–		2	
Mg237	D	224 245 292 362	309+C 315+C	–		2	
Mg244	D	218 223 295 362	195 309+C 315+C	–		2	
Mg248	D	129 223 362	152 217 309+C 315+C	–		2	
Mg249	D	129 223 362	152 217 309+C 315+C	–		2	
Mg251	D	223 320 335 362	279 315+C	–		2	
Mg253	D	223 320 335 362	279 315+C	–		2	
Mg257	D	218 223 295 362	195 309+C 315+C	–		2	
Mg260	D	362	194 309+C 315+C	–		2	10398 10400
Kor48	D	129 223 362		–		2	
Kor50	D	129 223 362		–		2	
Kor53	D	223 287 319 362 399		–		2	
Kor55	D	274 362 365	195 298 309+C 315+C	–		2	10398 10400

Table 1 (continued)

Sample[a]	Haplo-group	HVS-I (16001–16497)[b] (16000+)	HVS-II (30–407)[b] (73 and 263 in addition)	5176*Alu*I[c]	4831*Hha*I[c]	9-bp[d]	Other polymorphisms
Kor62	D	093 129 223 362		–		2	
Kor66	D	223 362		–		2	
Kor76	D	223 362		–		2	
Kor77	D	129 223 362		–		2	
Kor79	D	223 278 362		–		2	
Kor80	D	093 129 223 362		–		2	
Kor87	D	129 223 362		–		2	
EWK10	D	184 213 223 255 274 311 362		–		2	10398 10400 10601
EWK12	D	223 294 362		–		2	
EWK13	D	184 213 223 255 274 311 362		–		2	
EWK15	D	184 213 223 255 274 311 362		–		2	
EWK16	D	184 213 223 255 274 311 362		–		2	
EWK25	D	223 294 362		–		2	
EWK30	D	192 223	195 315+C 340	–		2	10398 10400
EWK39	D	184 213 223 255 274 311 362		–		2	
EWK4	D	093 223 232 290 362	195 309+C 315+C	–		2	
EWK41	D	223 278 362		–		2	
EWK44	D	223 294 362		–		2	
EWK45	D	223 278 362		–		2	
DW15	D5	167 189 223 362		–		2	
DW43	D5	183C 189 223 249 362		–		2	
Oro3	D5	093 183C 189 223 356 362		–		2	
Oro35	D5	093 183C 189 223 356 362		–		2	
Oro6	D5	093 183C 189 223 356 362		–		2	
Mg211	D5	223 311 316 362	150 151 152 309+CC 315+C	–		2	
Mg235	D5	188 189 214 223 362	309+C 315+C	–		2	
Mg256	D5	093 183C 189 217 223 319 362	150 309+C 315+C	–		2	
Kor68	D5	174 189 362		–		2	
Kor69	D5	174 189 362		–		2	
Kor72	D5	174 189 362		–		2	
Kor75	D5	174 189 362		–		2	
Kor90	D5	163 182C 183C 189 223 362		–		2	
EWK47	D5	167 189 223 362		–		2	
EWK48	D5	167 189 223 362		–		2	
DW19	D5a	092 172 182C 183C 189 223 266 362		–		2	
DW23	D5a	092 164 182C 183C 189 223 266 362		–		2	
Oro26	D5a	164 172 182C 183C 189 223 266 274 362		–		2	
Oro41	D5a	092 102 164 182C 183C 189 223 266 362		–		2	
Mg259	D5a	092 164 179 182C 183C 189 223 259 266 362	150 315+C	–		2	
EWK46	D5a	051 172 182C 183C 189 223 266 362		–		2	
Oro29	G	172 223 362	(263) 315+C	+	+	2	10398 10400

Oro30	G	172 223 362		+	+	2	
Oro39	G	172 223 362		+	+	2	10398 10400
Oro44	G	172 223 362		+	+	2	
Oro47	G	172 223 362		+	+	2	
EWK19	G	172 223 362		+	+	2	
EWK23	G	172 223 362		+	+	2	
EWK28	G	172 223 362	(263) 315+C	+	+	2	10398 10400
EWK31	G	172 223 362		+	+	2	
DW45	G1a	223 325 362		+	+	2	
DW48	G1a	223 325 362	150 315+C	+	+	2	10398 10400
Kor47	G1a	075 223 325 362		+	+	2	10398 10400
Mg219	G2	189 223 278 362	309+C 315+C	+	+	2	
Mg220	G2	189 223 278 362	309+CC 315+C	+	+	2	
Kor61	G2	189 223 278 362		+	+	2	10398 10400
Kor64	G2	183C 189 223 269 278 362		+	+	2	10398 10400
Kor73	G2	183C 189 223 269 278 362		+	+	2	
DW1	G2a	223 227 234 278 362		+	+	2	
DW4	G2a	223 227 243G 278 293C 362		+	+	2	
Kor60	G2a	093 223 227 278 362		+	+	2	
Kor65	G2a	223 227 265C 278 362		+	+	2	
Kor78	G2a	223 227 265C 278 362		+	+	2	
Kor81	G3	092 223 274 362		+	+	2	10398 10400
Mg246	M7	CRS	146A 199 204 309+C 315+C			2	10398 10400
Kor46	M7a1	172 209 223 324	195 207 309+C 315+C	+		2	9824; 10398 10400
DW7	M7b	129 180 223 297	150 159 199 204 309+C 315+C			2	
DW11	M7b1	129 192 223 256 297 309	146 150 199 309+C 315+C			2	
DW30	M7b1	129 192 223 297	150 199 315+C			2	9824
DW34	M7b1	129 192 223 297	150 199 315+C			2	
DW36	M7b1	129 192 223 297	150 199 315+C			2	
Kor51	M7b2	129 189 223 297 298				2	9824; 10345 10398 10400
Kor52	M7b2	129 189 223 297 298				2	10345 10398 10400
DW20	M7c	223 295 304	199 309+C 315+C			2	
DW3	M7c	223 295 360	146 152 199 309+CC 315+C			2	
DW5	M7c	223 295 360	146 152 199 309+CC 315+C			2	9824
DW6	M7c	223 295 360	146 152 199 309+CC 315+C			2	
Kor49	M7c	086 145 172 187 223 295	146 199 315+C			2	
DW40	C	223 298 327	146 249d 309+C 315+C			2	
DW41	C	093 223 298 325 327 356	93 249d 290d 291d 309+C 315+C			2	
DW44	C	223 298 327	146 249d 309+C 315+C			2	
Oro11	C	093 129 223 298 327	195 249d 309+C 315+C	+		2	10398 10400
Oro15	C	171 223 298 327 344 357	93 249d 309+C 315+C			2	
Oro21	C	171 223 298 327 344 357	93 249d 309+C 315+C			2	
Oro23	C	148 223 288 298 327	249d 309+C 315+C			2	
Oro24	C	171 223 298 327 344 357	93 249d 309+C 315+C			2	
Oro27	C	093 214 223 261 288 298	249d 315+C	+		2	10398 10400

Table 1 (continued)

Sample[a]	Haplo-group	HVS-I (16001–16497)[b] (16000+)	HVS-II (30–407)[b] (73 and 263 in addition)	5176*Alu*I[c]	4831*Hha*I[c]	9-bp[d]	Other polymorphisms
Oro28	C	171 223 298 327 344 357	93 249d 309+C 315+C			2	
Oro31	C	223 259+A 294 298 327	146 249d 309+CC 315+C			2	
Oro33	C	171 223 298 327 344 357	93 249d 309+C 315+C			2	
Oro37	C	093 214 223 261 288 298	249d 315+C	+		2	10398 10400
Oro38	C	223 259+A 298 327	146 249d 309+C 315+C			2	
Oro48	C	093 214 223 261 288 298	249d 315+C	+		2	10398 10400
Oro7	C	093 129 223 298 327 354	195 249d 315+C			2	
Mg221	C	223 298 327	146 249d 309+C 315+C			2	
Mg245	C	223 291 298 327	249d 309+C 315+C			2	
Mg254	C	051 129 223 248 298 327	152 249d 315+C			2	
EWK11	C	223 298 327	146 249d 309+C 315+C			2	
EWK17	C	129 150 223 298 327	195 249d 309+C 315+C			2	
EWK18	C	129 150 223 298 327	195 249d 309+C 315+C			2	
EWK27	C	129 150 223 298 327	195 249d 309+C 315+C			2	
EWK29	C	129 150 223 298 327	195 249d 309+C 315+C			2	
EWK32	C	093 223 288 298 327 390	249d 315+C			2	
EWK40	C	171 223 298 327 344 357	93 249d 315+C			2	
EWK8	C	223 298 327	146 249d 309+C 315+C			2	
EWK9	C	223 261 288 298	249d 309+CC 315+C	+		2	10398 10400
DW26	Z	185 223 260 298 311	152 249d 309+C 315+C			2	
DW27	Z	185 223 260 298 311	152 249d 309+C 315+C			2	
Oro25	Z	185 223 260 298 311	152 249d 309+C 315+C			2	
Kor45	Z	185 223 260 298	152 214 249d 315+C			2	
Kor56	Z	185 223 260 298	152 249d 315+C			2	
Kor58	Z	185 223 260 298	152 214 249d 315+C			2	
EWK33	Z	185 223 260 298	152 249d 309+CC 315+C			2	
EWK34	Z	185 223 260 298	152 249d 309+CC 315+C			2	
DW12	Z1	129 219 223 224 260 298	151 152 249d 315+C	+		2	10325 10398 10400
DW8	Z1	129 219 223 224 260 298	151 152 249d 315+C	+		2	10325 10398 10400
Mg217	Z1	129 185 223 224 260 298	151 152 249d 309+CC 315+C			2	
Mg250	Z1	129 185 223 224 260 298	151 152 249d 309+CC 315+C			2	
Mg223	M9a	223 234 265C 291 316 362	309+C 315+C	+		2	
Mg239	M9a	223 234 300 316 362	315+C	+		2	
Mg241	M10	093 129 193 223 311 357 497	146 152 309+C 315+C			2	
Kor63	M10	129 148 189 223 311 357				2	10398 10400 10646
Kor71	M10	066 223 261 311	315+C	+		2	10398 10400 10646
Kor74	M10	066 223 261 311	315+C	+		2	
Kor92	M	145 148 188 189 223	152 315+C	+	–	2	10398 10400 10411
DW2	B4	183C 189 217				1	
DW33	B4	183C 189 217				1	
DW42	B4	183C 189 217 478				1	

DW46	B4	183C 189 217			1	
Oro34	B4	183C 189 217			1	
Kor89	B4	183C 189 217 311			1	
DW17	B4a	093 182C 183C 184+C 187 189 214 217 261			1	
DW28	B4a	093 182C 183C 184+C 187 189 214 217 261			1	
DW39	B4a	051 182C 183C 189 213 217 261 299			1	
Mg243	B4a	181d 182C 183C 189 213 217 261 292	61A 62 309+CC 315+C		1	
Mg213	B4b	136 183C 189 217 218	309+CC 315+C		1	
Mg215	B4b	136 183C 189 217 218	309+CC 315+C		1	
Mg231	B4b	136 182C 183C 189 217 218 240	185 315+C		1	
Kor57	B4b	136 183C 189 217 311	315+C	+	1	10171–10659=CRS
Kor67	B4b	136 172 183C 189 217 218 428			1	
EWK14	B4b	136 182C 183C 189 217 218 240			1	
EWK3	B4b	136 182C 183C 189 217 218 240			1	
EWK5	B4b	136 182C 183C 189 217 218 240			1	
EWK6	B4b	136 182C 183C 189 217 218 240			1	
Mg258	B5a	140 183C 189 260 266G	152 210 294 315+C		1	
Kor59	B5b	140 182C 183C 189 243 278			1	
Kor91	B5b	140 182C 183C 189 243 278			1	
EWK35	B5b	017 140 182C 183C 189 227 234 243			1	
Mg225	F1a	129 164 172 304	249d 315+C		2	
Mg240	F1a	129 172 221 304 438	249d 309+C 315+C		2	10310 10609
Oro45	F1b	182C 183C 189 232A 249 304 311	249d 315+C		2	10310 10609
Oro46	F1b	182C 183C 189 232A 249 304 311	249d 315+C		2	
Mg238	F1c	111 129 266 304	152 249d 315+C		2	
EWK49	F1c	111 129 266 304	152 249d 309+C 315+C 368		2	10310 10454 10609
Mg242	F2a	126 203 291 304	249d 309+C 315+C		2	10535 10586
DW9	A	223 290 319 362	151 152 235 309+C 315+C		2	
Oro13	A	039 189 223 290 319 356 362	152 235 309+C 315+C		2	
Oro19	A	039 189 223 290 319 356 362	152 235 309+C 315+C		2	
Mg214	A	104 223 290 319 362	151 152 200 235 309+C 315+C	+	2	
Mg216	A	104 223 290 319 362	151 152 200 235 309+C 315+C	+	2	
Mg226	A	129 213 223 290 319	152 235 309+CC 315+C		2	
Mg233	A	223 290 319 362	152 235 309+C 315+C	+	2	
Kor54	A	223 278 290 319 362	200 235 309+C 315+C		2	
Kor82	A	223 290 319	152 235 309+CC 315+C		2	
Kor83	A	157 223 290 319 362	152 195 235 309+C 315+C	+	2	
EWK42	A	093 223 290 319 362	151 152 200 235 243 315+C	+	2	
EWK43	A	093 223 290 319 362	151 152 200 235 315+C	+	2	
Kor70	A5	187 223 290 319	235 309+C 315+C		2	
Kor84	A5	187 223 290 319	235 309+C 315+C		2	
Kor85	A5	187 223 290 319	235 309+C 315+C		2	
Kor86	A5	187 223 290 319	235 309+C 315+C		2	
DW25	N9a	129 223 257A 261	150 309+CC 315+C		2	
Mg228	N9a	111 129 223 257A 261	150 309+C 315+C		2	

Table 1 (continued)

Sample[a]	Haplo-group	HVS-I (16001–16497)[b] (16000+)	HVS-II (30–407)[b] (73 and 263 in addition)	5176*Alu*I[c]	4831*Hha*I[c]	9-bp[d]	Other polymorphisms
Mg229	N9a	111 129 223 257A 261	150 309+C 315+C			2	5417
Mg236	N9a	129 223 257A 261	150 247 309+C 315+C			2	
Kor88	N9a	129 189 223 257A 261				2	
DW35	Y1	126 231 266	146 315+C			2	5417; 10398; 14178
Mg255	R9b	145 192 243 304 309 390	183 309+C 315+C			2	10410
DW22	R2	071	152 195 315+C	+	–	2	10171–10659=CRS
DW24	R2	071	152 195 315+C		–	2	10171–10659=CRS
EWK22	H4	274	315+C	+		2	10601
EWK36	H4	274	315+C			2	10601
EWK37	H4	274	315+C			2	10601
EWK38	H4	274	315+C			2	10601
DW47	J1	069 126	185 228 295 309+C 315+C			2	10192 10398
EWK7	J1	069 126	185 228 295 309+C 315+C			2	10192 10398
EWK20	J1	069 126 245 325	185 189 228 295 309+C 315+C			2	10398
EWK21	J1	069 126 245 325	185 189 228 295 309+C 315+C			2	10398
EWK24	J1	069 126 245 325	185 189 228 295 309+C 315+C			2	10398
EWK26	J1	069 126 245 325	185 189 228 295 309+C 315+C			2	10398
Mg232	T1	126 163 186 189 294	152 195 309+C 315+C			2	12633A
DW31	T2	126 183C 189 292 294 296				2	14233
Oro16	N	176 223 232 355	195 309+C 315+C	+	–	2	10289 10556

[a]The ethnic populations Daur, Ewenki, Korean, Mongolian, and Oroqen are abbreviated as *DW*, *EWK*, *Kor*, *Mg*, and *Oro*, respectively

[b]Suffixes *A*, *G*, *C*, *T* and *d* or + indicate transversions and deletions, insertions, respectively; Indels are recorded at the last possible site (as usual in forensics)

[c]A – or + denotes the absence or presence of the restriction site, respectively

[d]*1* indicates the deletion of the 9-bp (CCCCCTCTA), *2* indicates nondeletion. Mutation *in parentheses* indicates that this mutation is absent in the mtDNA when compared with rCRS

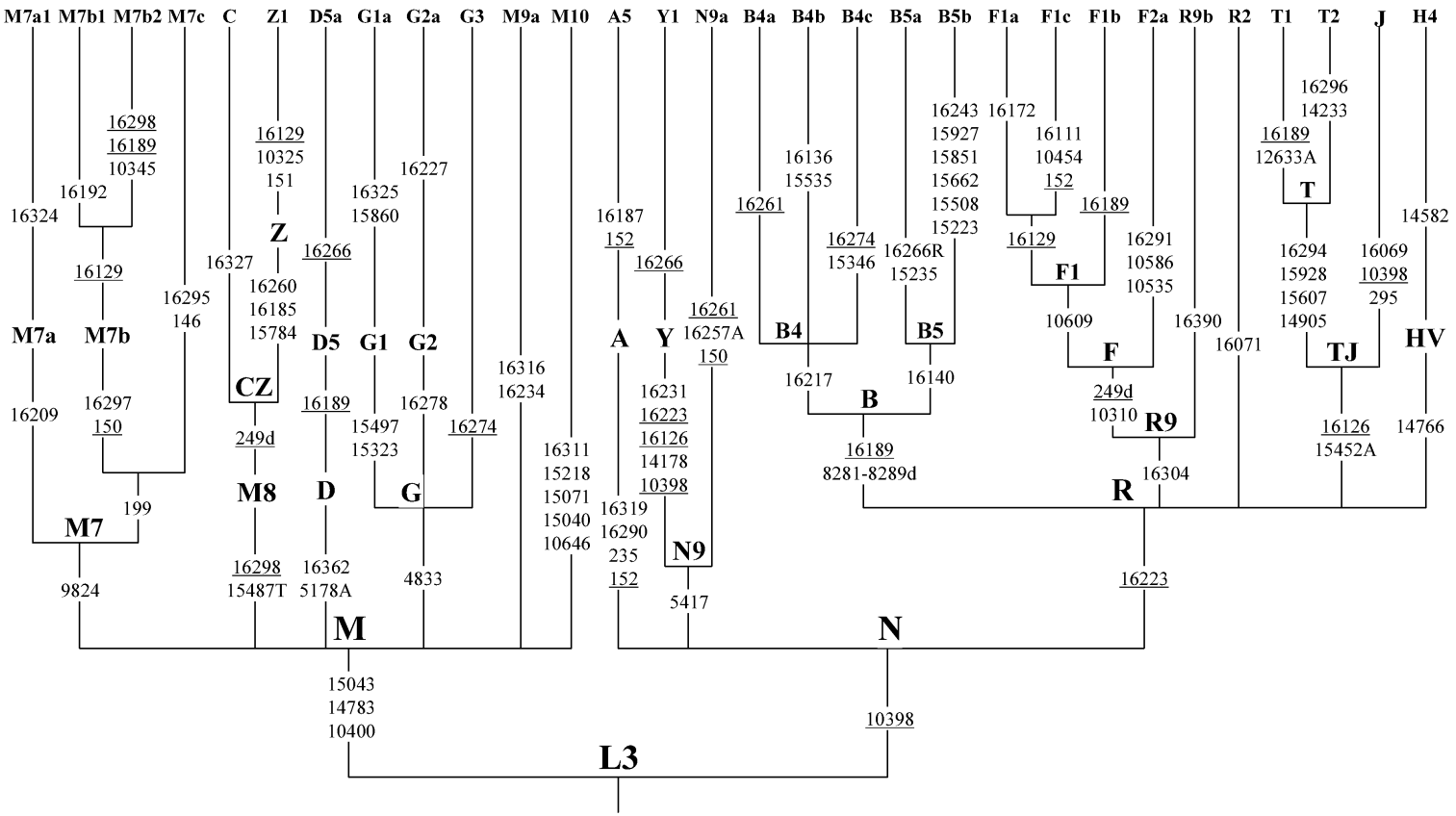

Fig. 2 Classification tree of the mtDNA haplogroups identified in 232 northern Chinese samples. This tree is constructed with reference to the classification trees of Yao et al. (2002a), Kivisild et al. (2002), and Kong et al. (2003). The characteristic mutations (relative to the revised CRS; Andrews et al. 1999) considered here are indicated on the branches with an arbitrary order. The *suffix* indicates a transversion, *d* indicates deletion, and recurrent mutations are *underlined*. The revised CRS branches out from the R node by seven haplogroup-specific mutations at sites 73, 1438, 2706, 4769, 7028, 11719, and 14766, plus four private mutations at sites 263, 750, 8860, and 15326

Table 2 Polymorphisms in region 14576–16047 of 63 samples. The sequenced mtDNA region that is identical to the revised reference sequence (Andrews et al. 1999) is labeled by CRS. *ND* not determined. All mutations are transition unless a *suffix* (i.e., A, C, G, and T) is specified; *d* indicates deletion, + means insertion. Mutation *in parentheses* indicates the absence of this mutation at the site in the sample when compared with rCRS, and *boldface* is used to highlight the specific mutations of a haplogroup relative to the roots of M and N, respectively

Sample[a]	Haplo-group	Region 14576–16047 (14766 and 15326 in addition)	HVS-I 16001–16497 (16000+)	HVS-II 30–407 (73 and 263 in addition)
EWK28	G	14783 15043 15301	172 223 362	(263) 315+C
Oro29	G	14783 15043 15301	172 223 362	ND
DW48	G1a	14783 15043 15301 **15323 15497 15860**	223 325 362	150 315+C
Kor47	G1a	14783 15043 15299 15301 **15323 15497 15860**	075 223 325 362	ND
Kor81	G3	14783 15043 15301 15746	092 223 274 362	ND
Mg246	M7	14783 15043 15236 15301	CRS	146A 199 204 309+C 315+C
Kor46	M7a	14783 15043 15301	172 209 223 324	195 207 309+C 315+C
DW7	M7b	14783 15043 15301	129 180 223 297	150 159 199 204 309+C 315+C
Zhuang34n	M7b	14783 14978 15043 15301 15721	129 189 223 297	ND
Kor51	M7b2	14783 15043 15301	129 189 223 297 298	ND
DW6	M7c	14783 15043 15301	223 295 360	146 152 199 309+CC 315+C
Kor49	M7c	14783 15043 15301	086 145 172 187 223 295	146 199 315+C
SD10331	M8a	14783 15043 15301 **15487T**	184 201 223 298 319 362	309+C
LN7590	M8a	14783 15043 15244 15301 **15487T**	184 223 298 319	315+C
LN7597	M8a	14783 15043 15301 **15487T**	184 223 298 319 400	152 315+C
LN7715	M8a	14783 15043 15301 **15487T**	184 209 223 293 298 311 319	207 309+CC 315+C
QD8120	M8a	14783 15043 15301 **15487T**	134 184 223 298 319	309+C 315+C
QD8150	M8a	14783 15043 15169 15301 **15487T** 15924	184 223 293 298 319	152 200 315+C
QD8159	M8a	14783 15043 15301 **15487T**	184 223 278 298 319	234 309+C 315+C
WH6981	M8a	14783 15043 15301 **15487T**	134 184 223 298 319	315+C
XJ8417	M8a	14783 15043 15301 **15487T**	184 223 293C 298 319	152 309+C 315+C
LN7710	C	14783 14978 15043 15301 **15487T** 15930	217 223 298 311 327	146 249d 309+CC 315+C
XJ8435	C	14783 15043 15204 15301 **15487T** 15968	129 223 298 327	195 249d 309+C 315+C
WH6979	Z	14783 15043 15301 15475 **15487T 15784** 15944d	185 189 223 224 260 261 298 302	152 185 249d 309+C 315+C
DW8	Z	14783 15043 15261 15301 **15487T 15784**	129 219 223 224 260 298	151 152 249d 315+C
Mg223	M9a	14783 15043 15301	223 234 265C 291 316 362	309+C 315+C
SD10334	M10	14783 **15040** 15043 **15071** 15172 **15218** 15301 15924	066 223 311	315+C
Kor63	M10	14783 **15040** 15043 **15071 15218** 15301	129 148 189 223 311 357	ND
Kor74	M10	14783 **15040** 15043 **15071 15218** 15301	066 223 261 311	315+C
QD8122	M10	14783 **15040** 15043 **15071 15218** 15301	129 223 311	315+C
Mg241	M10	14783 **15040** 15043 **15071 15218** 15301	093 129 193 223 311 357 497	146 152 309+C 315+C
LN7593	M10	14783 **15040** 15043 **15071 15218** 15301 15580	093 129 193 223 311 357 497	146 152 309+C 315+C
LN7720	M10	14783 14870 **15040** 15043 **15071 15218** 15301	223 311	195 315+C 331
YN163	M10	14783 **15040** 15043 **15071 15218** 15301	093 129 223 311 357 497	309+CC 315+C
LN7596	M10	14783 **15040** 15043 **15071 15218** 15301	066 086 092 223 311	152 315+C
SD10324	M	14783 14790 15043 15301	223 265T 497	146 215 315+C 318 326
SD10364	M	14783 15043 15301 15731	223 311	198 200 215 309+C 315+C 318 326
Kor92	M	14783 15043 15301 15924	145 148 188 189 223	152 315+C
SD10313	B	CRS	093 179 182C 183C 189	150 309+CC 315+C
GD7832	B	CRS	129 183C 189 352 355	150 152 185 189 309+C 315+C
GD7812	B4a	14905	181d 182C 183C 189 217 261 292	309d 315+C
WH6982	B4b	**15535** 15930	183C 189 234	315+C

LN7589	B4b	15038 **15535** 15930	183C 189 217	309+CC 315+C 316
XJ8428	B4b	14587 15314 15370 **15535**	136 183C 189	114 309+CC 315+C
GD7813	B4b	15301 **15535**	136 183C 189 217 309 354	207 309+C 315+C
GD7814	B4b	15301 **15535**	136 183C 189 217 309 354	146 207 315+C
QD8119	B4b	15301 **15535**	092 136 183C 189 309 354	207 315+C
LN7716	B4b	15236 **15535**	136 183C 189 284	199 202 207 309+CCC 315+C
Kor57	B4b	**15535**	136 183C 189 217 311	315+C
LN7552	B4c	**15346**	140 182d 183C 189 217 274 311	146 150 315+C
YN154	B4c	14687 14763C **15346**	140 183C 189 217 274	150 152 309+C 315+C
Yao11	B5a	**15235**	140 183C 189 266A	ND
Yao8	B5a	**15235**	140 183C 189 266A	ND
SD10308	B5b	**15223 15508 15662 15851 15927**	140 158 182C 183C 189 243	103 114 309+CC 315+C
SD10319	B5b	**15223 15508 15662** 15850 **15851 15927**	111 140 182C 183C 189 234 243 291 463	131 204 309+C 315+C
Mg255	R9b	CRS	145 192 243 304 309 390	183 309+C 315+C
EWK37	H4	**14582 (14766)**	274	315+C
EWK20	J1	**14798 15452A**	069 126 245 325	185 189 228 295 309+C 315+C
EWK7	J1	**14798 15452A**	069 126	185 228 295 309+C 315+C
Mg232	T1	**14905 15452A 15607 15928**	126 163 186 189 294	152 195 309+C 315+C
DW31	T2	**14905** 15287 **15452A 15607 15928**	126 183C 189 292 294 296	ND
DW22	R2	CRS	071	152 195 315+C
Zhuang49	R	CRS	189 311 390 399	ND

[a]The samples were selected according to their control-region information (including HVS-I and/or HVS-II) from this study, our previously reported data (Yao et al. 2002a, 2002b, 2003a) as well as our unpublished data set

also be found in the Tibetan samples from Yunnan (Yao and Zhang 2002). The one N haplotype from the Oroqen sample (Oro16) near-matches a Han Chinese sample from Wuhan, Hubei (WH6976; Yao et al. 2002a).

Table 2 shows the sequence polymorphisms identified in region 14576–16047 in the 63 individuals. This region provides useful information in supporting the poorly characterized haplogroups that only defined by the control region motifs (Kivisild et al. 2002; Yao et al. 2002a, 2003a). For instance, haplogroup B5b, a sub-haplogroup of B5, which was formerly defined by HVS-I transitions at sites 16140, 16189, and 16243, is newly confirmed by five specific mutations (15223, 15508, 15662, 15851, and 15927) in this region. Similarly, haplogroup B5a is identified by the 15235 mutation, haplogroup B4b is supported by a transition at site 15535, haplogroup B4c is recognized by 15346 mutation, haplogroups M8a, C, and Z share a specific transversion at site 15487. Haplogroup M10, which was formerly defined by mutations 10646 and 16311 (Yao et al. 2002a), could be well recognized by mutations at sites 15040, 15071, and 15218. The newly described haplogroup G1 (Bandelt et al. 2003; Kong et al. 2003), which is a sub-haplogroup of haplogroup G, is characteristic of mutations 15323 and 15497. Mutation 15860 and the HVS-I motif (16223-16325-16362) further defined a sub-branch of G1, G1a. The region 14576–16047 also provides information for defining several European specific haplogroups (Finnilä et al. 2001; Herrnstadt et al. 2002), such as JT (characterized by mutation 15452A), T (recognizable by mutations 14905, 15607, and 15928), and H4 (characterized by mutations 14766 and 14582). Note that a subset of J is also identifiable by a transition at site 14798 (Finnilä et al. 2001; Herrnstadt et al. 2002). As a result, 14 individuals from the Daur, Ewenki, and Mongolian samples considered here could be assigned into the west European-specific haplogroups R2, H4, J1, T1, and T2 (Table 1). Our extensive searching for the reported data in Chinese showed that haplogroups J and T also occurred in samples from Liaoning, Shaanxi, Hunan, and Xinjiang Provinces (Oota et al. 2002; Yao et al. 2000a, 2002a). The question about the invasion and spread of these western-Eurasian-specific lineages across China is still unspecific, and further analysis is needed to resolve it.

Errors in the reported Korean data

The mtDNA coding region information has been employed in forensic science recently (Tzen et al. 2001; Lee et al. 2002). The now available data with coding region and control region information (Kivisild et al. 2002; Yao et al. 2002a; Kong et al. 2003; this study) could be used as a benchmark to check the potential reading errors or artificial recombination in the reported data set. In the 98 Korean samples that were sequenced for the complete Cyt *b* sequence and the two hypervariable segments of control region (Lee et al. 2002), several obvious errors caused by possible sample crossover can be easily discerned: sample H84 is a crossover of M7a1 with D4a; F531.2 is a cross-

Table 3 The haplogroup distribution frequencies (%) in the seven northern ethnic populations. Populations Daur, Oroqen, Ewenki, Korean, Mongolian, Buryat, and Yakut are abbreviated as *DW*, *Oro*, *EWK*, *Kor*, *Mg*, *Bur*, and *Yak*, respectively

Haplogroup	DW (*n*=45)	Oro (*n*=44)	EWK (*n*=47)	Kor (*n*=48)	Mg (*n*=48)	Bur (*n*=126)[a]	Yak (*n*=117)[a]
D	15.6	31.8	25.5	22.9	33.3	15.9	14.5
D5	4.4	6.8	4.3	10.4	4.2	1.6	0.9
D5a	4.4	4.5	2.1		2.1	2.4	14.5
G		11.4	8.5				
G1a	4.4			2.1			0.9
G2				6.3	4.2	1.6	
G2a	4.4			6.3		6.3	1.7
G3				2.1			
M7					2.1		
M7a1				2.1			
M7b	2.2					0.8	
M7b1	8.9						
M7b2				4.2			
M7c	8.9			2.1		0.8	1.7
M8a						0.8	
C	6.7	29.5	19.1		6.3	15.9	38.5
Z	8.9	2.3	4.3	6.3	4.2	1.6	
M9a					4.2		
M10				6.3	2.1	1.6	
M				2.1		1.6	1.7
A	2.2	4.5	4.3	6.3	8.3	7.1	1.7
A5				8.3			
N1a						1.6	
N1b						0.8	
N9a	2.2			2.1	6.3	1.6	
Y1	2.2					0.8	
W						0.8	1.7
X						0.8	
N		2.3					
B						0.8	
B4	8.9	2.3		2.1		4.0	
B4a	6.7				2.1		
B4b			8.5	4.2	6.3	0.8	
B5a					2.1		
B5b			2.1	4.2		1.6	
F						1.6	
F1a					4.2	2.4	
F1b		4.5				4.8	1.7
F1c			2.1		2.1		
F2a					2.1		2.6
R2	4.4						
R9b					2.1		
HV1							3.4
H						2.4	
H4			8.5				
J						0.8	4.3
J1	2.2		10.6			0.8	1.7
T						0.8	5.1
T1					2.1		
T2	2.2						
K						0.8	
U2e						0.8	
U4						2.4	
U5a1						3.2	
U5a1a						0.8	
U5b						0.8	
U5b1							0.9
R						6.3	2.6

[a]Data from Pakendorf et al. (2003)

over of F1b and D4a; H81 is a crossover of A5 with D4a; H98 is a crossover of A5 and M. Besides these recombination errors, there are many overlooked polymorphisms: 16290 might be missed in H98; 15326 might be overlooked in H108 and F907.1; 249d might be disregarded in SB41; in the three G1a samples, F916.2 and F844.1 all lacked 15497, while F408.2 missed 15323. The transversion at site 15487, which is shared by haplogroups M8a, C, and Z, was neglected in sample F496.1. Moreover, oversight of mutation 16362 seems to be frequent for the D4a types that were identifiable via transition at site 14979. These seemingly artificial errors caused by sample crossover or other reasons are not infrequent. The 8.8-kb length of mtDNA sequences of Native Americans reported by Silva et al. (2002) also contained such problems (Yao et al. 2003b, 2003c). Even with extreme caution during the bench work, such errors may occur (c.f. Kong et al. 2003; Yao et al. 2003a). Thus, additional quality control measures, such as independent typing for the region by different individuals (Herrnstadt et al. 2002), matching or near-matching with reliable data sets, and detecting errors by phylogenetic analysis as described recently (Bandelt et al. 2001, 2002; Yao et al. 2003b, 2003d; Yao and Zhang 2003) should be extensively employed to avoid possible errors in the data.

Haplogroup distribution

Table 3 shows the haplogroup distribution frequencies in the five northern ethnic groups, as well as in the samples reported by Pakendorf et al. (2003), from which several features can be discerned: (1) haplogroups D, D5, G, and A are distributed widely among the seven populations; (2) haplogroups G2 (including G2a), M9a, Y, and the subhaplogroups of F1 (including F1a, F1b, and F1c) have limited distributions in these samples; (3) some north-prevalent haplogroups – namely, D, G, C, and Z (Yao et al. 2002a, 2003a) – have relative high frequencies (altogether more than 45%) in these populations; (4) the frequency of haplogroup B is quite high (more than 7.1%) in Daur, Ewenki, Korean, Mongolian, and Buryat, but is lower in either Oroqen (2.3%) or Yakut (0.0%); (5) haplogroup M7 shows high frequency in Daur (20.0%) and Korean (8.4%), which, however, were either absent or with low frequency in other populations.

Discussion

The emerging mtDNA phylogeny of East Asian mtDNAs and the available data set with coding region and control region information can serve as the foundation for the East Asian mtDNA haplogroup assignment, and this has been fully described in a series of recent studies (Kivisild et al. 2002; Yao et al. 2002a, 2003a; Kong et al. 2003). The dissection of the 232 mtDNAs from northern China in this study by the same strategy revealed that: (1) most of the mtDNAs could be classified into the most-derived named mtDNA haplogroups (Fig. 2), and (2) some haplogroups, such as D, G, C, and Z, were prevalent in these northern samples and the matrilineal genetic profile was consistent with the genetic pattern observed recently (Yao et al. 2002a).

Two mtDNA coding region segments (10171–10659 and 14055–14590) analyzed in Yao et al. (2002a) were found to be very informative for East Asian mtDNA haplogroup characterization. However, these segments provided little information in supporting the status of haplogroups G1, G1a, M8, M10, B4b, B4c, B5a, and B5b. Our analyses of the region 14576–16047 showed that it contained many characteristic polymorphisms for these haplogroups and filled the lacunae. We suggested that when discerning the haplogroup status of the major haplogroups in East Asian mtDNAs based on short coding-region segments, these three segments should be the ideal choices.

The comparison of the matrilineal genetic structure of the ethnic populations could reflect their ethnohistory more or less (Yao et al. 2002a, 2002b; Yao and Zhang 2002). The Koreans in China are mainly the descendents of migrants from the Korean Peninsula (Du and Yip 1993). Our analysis of the Korean sample revealed that it contained the specific haplogroups A5 and M7a1, which are also prevalent in South Koreans (Kivisild et al. 2002, and references therein), thus revealing a common matrilineal genetic background of the Koreans in China and in the Korean Peninsula.

The Daurs were said to be the descendants of Khitan people in Liao Dynasty (916 AD–1125 AD). Other hypotheses suggested that their ancestors could be traced back to the local people in northern Heilongjiang Province and to some tribes of the Heishui region (same province) during the Shui and Tang Dynasty (581 AD–907 AD) (Du and Yip 1993). Our results demonstrated that Daur contained a high amount of the haplogroups prevalent in northern China (>46%), thus consistent with their northern origin.

According to historical documents, the Ewenkis traced their origin to the populations who lived around Lake Baikal and adjacent eastern regions more than 2,000 years ago, and were divided into three long-separated branches (Solon, Tungus, and Yakut; Du and Yip 1993). Our results supported the suggestion that the Ewenki is a typically

Table 4 Genetic diversities in the seven northern ethnic populations. The genetic diversities in the populations were calculated according to the HVS-I sequence [relative to 16001–16400 in the revised reference sequence (Andrews et al. 1999)]

Population	*n*	No. of haplotypes	Haplotype diversity	Nucleotide diversity
Daur	45	30	0.979±0.009	0.016±0.001
Ewenki	47	21	0.956±0.011	0.017±0.001
Korean	48	30	0.975±0.009	0.015±0.001
Mongolian	48	38	0.989±0.007	0.017±0.001
Oroqen	44	21	0.948±0.015	0.016±0.001
Yakut[a]	117	41	0.961±0.007	0.020±0.001
Buryat[a]	126	91	0.993±0.002	0.020±0.001

[a]Data from Pakendorf et al. (2003)

northern population, for more than 63% of its maternal components are composed of haplogroups D, G, C, and Z. Furthermore, the high frequencies of haplogroups C and D but lower haplotype diversity (0.956±0.011; Table 4) observed in Ewenki sample suggested that Ewenki might have undergone recurrent genetic drifts because of its small population size (about 26,000, 1990 census) and episodes of population fragmentation during its development (Du and Yip 1993).

The Oroqens were regarded as the earliest inhabitants who lived in the Heilong River valley (Du and Yip 1993). The Oroqen population size experienced serious reduction during the past ten decades: the population size was about 18,000 in 1895, but was reduced to 2,256 by 1953. In 1990, the sample size increased to 6,965 (Du and Yip 1993). The genetic structure of our Oroqen sample showed concordant features with its small population size and recorded history: high frequencies (86.4%) of north-prevalent haplogroups, such as D, G, C, and Z, were found, but with low haplotype diversity (0.948±0.015; Table 4).

The formation and development of the Mongolian population was a complex process affected by integrating many Turkic-speaking tribes and some ethnic groups such as Han, Manchu, and Daur. Although the south-prevalent hapologroups F, R9b (formerly R10 in Yao and Zhang, 2002; c.f. Kong et al. 2003), and N9a were found with a low frequency in the Mongolian sample, the main maternal components of the population are composed of the north-prevalent haplogroups, which occupy more than 58% of the total samples.

In short, we identified another coding region segment (region 14576–16047) that is informative for discerning the haplogroup status of East Asian mtDNAs besides the previously reported ones (Yao et al. 2002a). Although the matrilineal genetic components of the five northern Chinese ethnic populations differed, the observed genetic profile was in general consistent with that of the Chinese Han regional samples (Yao et al. 2002a). The presence of northern population prevalent haplogroups D, G, C, and Z in these populations gave direct information in supporting their northern origin. Therefore, the matrilineal structures of these five northern Chinese ethnic populations reflected both the regional features and their ethnohistory.

Acknowledgements We thank Shi-Fang Wu for technical assistance and two anonymous reviewers for their critical comments on the manuscript. We also thank Dr. Brigitte Pakendorf for providing us the sequence data. The work was supported by Natural Science Foundation of China (NSFC) and Natural Science Foundation of Yunnan Province.

Electronic database information

Accession numbers and URLs for the sequence data of mtDNA control region (including HVS-I and HVS-II) and the region 14576–16047 in this article are as follows: GenBank, http://www.ncbi.nlm.nih.gov/web/Genbank (accession numbers: AY243965–AY244330 and AY243877–AY243939).

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

Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mtDNA data. Am J Hum Genet 71:1150–1160

Bandelt H-J, Herrnstadt C, Yao Y-G, Kong Q-P, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang Y-P (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet (in press)

Cooper DN, Nussbaum RL, Michael Krawczak M (2002) Proposed guidelines for papers describing DNA polymorphism-disease associations. Hum Genet 110:207–208

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

den Dunnen JT, Antonarakis SE (2001) Nomenclature for the description of human sequence variations. Hum Genet 109:121–124

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

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152–1171; 71:448–449 (erratum)

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

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

Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet (in press)

Lee SD, Lee YS, Lee JB (2002) Polymorphism in the mitochondrial cytochrome B gene in Koreans: an additional marker for individual identification. Int J Legal Med 116:74–78

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Oota H, Kitano T, Jin F, Yuasa I, Wang L, Ueda S, Saitou, Stoneking M (2002) Extreme mtDNA homogeneity in continental Asian populations. Am J Phys Anthropol 118:146–153

Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M (2003) Mitochondrial DNA evidence for admixed origins of central Siberian populations. Am J Phys Anthropol 120:211–224

Povey S, Lovering R, Bruford E, Wright M, Lush M, Wain H (2001) The HUGO Gene Nomenclature Committee (HGNC). Hum Genet 109:678–680

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Norby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozari R, Torroni A, Bandelt H-J (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Silva Jr WA, Bonatto SL, Holanda AJ, Ribeiro-Dos-Santos AK, Paixão BM, Goldman GH, Abe-Sandes K, Rodriguez-Delfin L, Barbosa M, Paçó-Larson ML, Petzl-Erler ML, Valente V, Santos SEB, Zago MA (2002) Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. Am J Hum Genet 71:187–192

Tzen C-Y, Wu T-Y, Liu H-F (2001) Sequence polymorphism in the coding region of mitochondrial genome encompassing position 8389–8865. Forensic Sci Int 120:204–209

Yao Y-G, Zhang Y-P (2002) Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. J Hum Genet 47:311–318

Yao Y-G, Zhang Y-P (2003) Pitfalls in the analysis of ancient human mtDNA. Chinese Sci Bull 48:826–830

Yao Y-G, Lü X-M, Luo H-R, Li W-H, Zhang Y-P (2000a) 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 (2000b) 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

Yao Y-G, Yuan Z-G, Zhou Z-D, Geng P-L, Li Q-W, Zhang Y-P (2001) Frequency of the mtDNA 9-bp deletion among Chinese ethnic groups. Prog Nat Sci 11:358–364

Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P (2002a) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651

Yao Y-G, Nie L, Harpending H, Fu Y-X, Yuan Z-G, Zhang Y-P (2002b) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118:63–76

Yao Y-G, Kong Q-P, Man X-Y, Bandelt H-J, Zhang Y-P (2003a) Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. Mol Biol Evol 20: 214–219

Yao Y-G, Macaulay V, Kivisild T, Zhang Y-P, Bandelt H-J (2003b) To trust or not to trust an idiosyncratic mitochondrial data set. Am J Hum Genet 72:1341–1346

Yao Y-G, Macaulay V, Kivisild T, Zhang Y-P, Bandelt H-J (2003c) Reply to Silva et al. Am J Hum Genet 72:1348–1349

Yao Y-G, Kong Q-P, Sun C, Zhang Y-P (2003d) Can the occurrence of rare insertion/deletion polymorphisms in human mtDNA be verified from phylogeny? Chinese Sci Bull 48:663–667