J Hum Genet (2004) 49:414–423
DOI 10.1007/s10038-004-0170-3

ORIGINAL ARTICLE

**Qing-Peng Kong · Yong-Gang Yao · Chang Sun
Chun-Ling Zhu · Li Zhong · Cheng-Ye Wang
Wang-Wei Cai · Xiang-Min Xu · An-Long Xu
Ya-Ping Zhang**

Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic

Received: 17 February 2004 / Accepted: 6 May 2004 / Published online: 22 July 2004

Abstract In this report, we studied on a homoplasmic T12338C change in mitochondrial DNA (mtDNA), which substituted methionine in the translational initiation codon of the NADH dehydrogenase subunit 5 gene (*ND5*) with threonine. This nucleotide change was originally identified in two mtDNAs belonging to haplogroup F2 by our previous complete sequencing of 48 mtDNAs. Since then, a total of 76 F2 mtDNAs have been identified by the variations occurring in the hypervariable segments and coding regions among more than 3,000 individuals across China. As the T12338C change was detected in 32 samples representing various sub-clades of the F2 haplogroup while not in 14 non-F2 controls, we believe that the T12338C change is specific to the F2 haplogroup. As F2 and its sub-clades were widely distributed in normal individuals of various Chinese populations, we conclude that T12338C is not pathogenic. In addition, based on the average distribution frequency, haplotype diversity and nucleotide diversity of haplogroup F2 in the populations across China, the T12338C nucleotide substitution seems to have been occurred in north China about 42,000 years ago. Our results provided a good paradigm for distinguishing a polymorphic change from a pathogenic mutation based on mtDNA phylogeny.

Keywords Mitochondrial DNA · Haplogroup F2 · T12338C · ND5 · Phylogeny · East Asian

Accession numbers and URLs for the sequence data of mtDNA control region (including HVS-I and HVS-II) of F2 types in this article are as follows: GenBank, http://www.ncbi.nlm.nih.gov/web/Genbank (accession numbers: AY522667–AY522718).

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

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

Q.-P. Kong · C. Sun · C.-Y. Wang
Graduate School of the Chinese Academy of Sciences,
Beijing, 100039, China

Present address: Y.-G. Yao
The Wilmer Eye Institute,
Johns Hopkins University School of Medicine,
Baltimore, MD 21287, USA

W.-W. Cai
Department of Biochemistry,
Hainan Medical College, Haikou,
571101, China

X.-M. Xu
Department of Medical Genetics,
First Military Medical University, Guangzhou,
510515, China

A.-L. Xu
College of Life Sciences, Zhongshan University,
Guangzhou, 510275, China

Introduction

In the past two decades, many mutations, including point mutations, insertions, deletions, and rearrangements, have been detected in protein-coding genes as well as tRNA and rRNA genes in mitochondrial DNA (mtDNA), and were suggested to be pathogenic (Wallace et al. 1999; DiMauro and Schon 2001; Chinnery and Schon 2003). However, among these pathogenic mutations (http://www.mitomap.org), it is rare to find a mutation occurring in the conserved initiation codon of coding genes. Since the change of initiator amino acid from methionine to another

residue would hamper protein translation, it was natural to regard such a mutation as pathogenic. Hitherto, only three mtDNA mutations have been reported as occurring in the translational initiation codon: T3308C in the *ND1* gene (Campos et al. 1997), T7587C in the *COX II* gene (Clark et al. 1999), and A8527G in the *ATP6* gene (Dubot et al. 2004). Mutation T7587C was identified in a family with mitochondrial encephalomyopathy, and suggested to affect translation of COX II (Clark et al. 1999). Mutation T3308C was originally reported to be associated with bilateral striatal necrosis and MELAS (mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes) syndrome (Campos et al. 1997), however, it was later proved to neither impair synthesis of ND1 polypeptide nor affect the activity of complex I, thus suggesting that T3308C might be a benign mutation (Vilarinho et al. 1999; Fernandez-Moreno et al. 2000). Rocha et al. (1999) reached the same conclusion by a phylogenetic analysis. All the mtDNAs harboring T3308C were classified into a particular haplogroup, L1b, which is widely distributed in current African and Iberian populations. The third mutation, A8527G, was recently identified in apparent normal individuals. Although the mutation disrupts the ATG initiation codon of the *ATP6* gene, GTG, thus generated, usually encodes valine, may serve as a translational initiation codon in mitochondria (Dubot et al. 2004). The different effects of these mutations in respective cases raise the possibility of some unknown remediation pathway(s). Since the third codon in ND1 is ATG, one of the plausible explanations is that the third codon may serve as an alternative initiation codon to produce similar, but two amino acid short, polypeptide (Rocha et al. 1999; Fernandez-Moreno et al. 2000). Additional mutations and variations occurring in the initiation codon of mtDNA genes in human and other species would clarify the mechanism.

In this study, we report a homoplasmic nucleotide change T12338C in mtDNA, which occurs in the initiation codon of the ND5 gene and substitutes methionine with threonine (M1T). The nucleotide change was originally detected in two mtDNAs (GD7809 and QD8147) when we systematically surveyed mtDNAs belonging to all the major haplogroups specific to East Asian by complete mtDNA sequencing (Kong et al. 2003b). To substantiate whether the T12338C change is specific to haplogroup F2 and to learn more about the origin of haplogroup F2, we performed an extensive search for F2 in more than 3,000 Chinese mtDNAs in reported studies (Yao et al. 2000, 2002a,c, 2003; Tsai et al. 2001; Kivisild et al. 2002; Oota et al. 2002; Yao and Zhang 2002; Kong et al. 2003a; Tajima et al. 2003) and our unpublished data by motif-searching and/or (near-)matching methods (Yao et al. 2002a, 2003). Our phylogeographic study revealed that the T12338C change was specific to haplogroup F2 and occurred in normal individuals across China, thus suggesting a polymorphic change rather than a pathogenic mutation.

Material and methods

Sampling

A total of 1,494 subjects from 28 populations across China were screened in this study. All of the individuals were confirmed to be unrelated before sampling and were given informed consent. To better understand the phylogeny of haplogroup F2, the previously reported data sets (Yao et al. 2000, 2002a,c, 2003; Tsai et al. 2001; Kivisild et al. 2002; Oota et al. 2002; Yao and Zhang 2002; Kong et al. 2003a; Tajima et al. 2003) were also included. As a result, a total of 3,090 mtDNAs from 57 populations across China were examined, and their detailed information was illustrated in Table 1.

DNA amplification and sequencing

The hypervariable segments I (HVS-I) and II (HVS-II) of mtDNA control region as well as the regions including potential variations were amplified and sequenced as described elsewhere (Yao et al. 2002a; Kong et al. 2003a,b).

Data analyses

The mtDNA sequences were edited and aligned by the DNAStar software package. Mutations were scored according to the revised Cambridge reference sequence (rCRS; Andrews et al. 1999). Length polymorphisms of A and/or C stretches in region 16180–16193 in HVS-I and region 303–315 in HVS-II were disregarded in the subsequent analysis. Each mtDNA was tentatively assigned to a haplogroup on the basis of the variations in the HVS-I and II control regions. The haplogroup status was further confirmed by detecting additional variations in other regions as described in our previous studies (Yao et al. 2002a, 2003; Kong et al. 2003a). A segment covering region 10171–10659 of the rCRS, which was suggested to be informative in defining East Asian specific haplogroups (Yao et al. 2002a), was adopted to specify the phylogenetic status of the F* or R9* mtDNAs (the asterisk attached to haplogroups indicates that the sample was not able to be further classified into the sub-clade(s) of the haplogroup) in our previous studies (Yao et al. 2000, 2002a,c, 2003; Yao and Zhang 2002; Kong et al. 2003a) and unpublished data. For those published data sets (not from our laboratory) with only HVS-I and/or HVS-II information available, we recognized the potential F2 types by

Table 1 Frequency of mtDNA haplogroup F2 in Chinese ethnic populations

Population	Abbreviation	Location	No. of samples examined	No. of samples with F2 haplotype	Frequency (%)	References
XJ Han	XJ	Yili, Xinjiang	47	2	4.3	Yao et al. (2002a)
Kazak	Kaz	Xinjiang	30	1	3.3	Yao et al. (2000)
Uygur	Uyg	Xinjiang	45	0	0	Yao et al. (2000)
Huizu	Hui	Yili, Xinjiang	45	1	2.2	This study
Mongolian	Mg	Yili, Xinjiang	49	1	2.0	This study
QH Han	QH	Qinghai	95	1	1.1	This study
QH DM	QHD	Qinghai	78	1	1.3	This study
Tibetan	Tibet	Qinghai	37	2	5.4	This study
Tuzu	Tu	Qinghai	64	2	3.1	This study
Mongolian	Mg	Qinghai	15	1	6.7	Yao et al. (2002c)
Mongolian	Mg	Inner Mongolia	48	1	2.1	Kong et al. (2003a)
Xibe	Xibe	Inner Mongolia	49	0	0	This study
Oroqen	Oro	Inner Mongolia	44	0	0	Kong et al. (2003a)
Korean	Kor	Inner Mongolia	48	0	0	Kong et al. (2003a)
Ewenki	Ewk	Inner Mongolia	47	0	0	Kong et al. (2003a)
Daur	Daur	Inner Mongolia	45	0	0	Kong et al. (2003a)
Buryat	Bur	Inner Mongolia	58	0	0	This study
Manchu	Man	Inner Mongolia	72	1	1.4	This study
QD Han	QD	Qingdao, Shandong	50	1	2.0	Yao et al. (2002a)
SD Han	SD	Tai'an, Shandong	76	6	7.9	Yao et al. (2003)
LN Han	LN	Fengcheng, Liaoning	51	1	2.0	Yao et al. (2002a)
Manchu	Man	Fengcheng, Liaoning	30	0	0	This study
Lisu	Lisu	Gongshan, Yunnan	37	4	10.8	Yao et al. (2002c)
Nuzu	Nu	Gongshan, Yunnan	30	5	16.7	Yao et al. (2002c)
Baizu	Bai	Dali, Yunnan	31	1	3.2	Yao et al. (2002c)
Lahu	Lahu	Yunnan	35	9	25.7	Yao and Zhang (2002)
Naxi	Naxi	Lijiang, Yunnan	56	3	5.4	This study
Tibetan	Tibet	Yunnan, Qinghai	40	0	0	Yao et al. (2002c)
Sani (Yizu)	Sani	Nuxi, Yunnan	31	1	3.2	Yao et al. (2002c)
Hani	Hani	Nuxi, Yunnan	25	0	0	This study
Jinozu	Jino	Yunnan	31	0	0	This study
Kucong	Kc	Yunnan	34	0	0	This study
WH Han	WH	Wuhan, Hubei	42	2	4.8	Yao et al. (2002a)
QJ Han	QJ	Qijiang, Chongqing	51	2	3.9	This study
GD Han	GD	Zhanjiang, Guangdong	30	4	13.3	Yao et al. (2002a)
GD Han	GD	Chaoshan, Guangdong	51	0	0	This study
GD Han	GD	Meizhou, Guangdong	61	1	1.6	This study
GD Han	GD	Guangdong	69	3	4.3	Kivisild et al. (2002)
YN DM	YND	Kunming, Yunnan	97	1	1.0	This study
YN Han	YNC	Kunming, Yunnan	82	2	2.4	This study
YN Han	YN	Kunming, Yunnan	43	1	2.3	Yao et al. (2002a)
Vazu	Va	Yunnan	36	0	0	Yao and Zhang (2002)
Daizu	Dai	Xishuangbanna, Yunnan	41	1	2.4	Yao et al. (2002c)
SX Han	SX	Xi'an, Shaanxi	85	0	0	Oota et al. (2002)
HN Han	HN	Changsha, Hunan	82	0	0	Oota et al. (2002)
Tujia	Tj	Fenghuang, Hunan	46	0	0	This study
Miaozu	Miao	Fenghuang, Hunan	48	0	0	This study
Miaozu	Miao	Kaili, Guizhou	35	4	11.4	This study
Shezu	She	Fuquan, Guizhou	54	1	1.9	This study
Shuizu	Shui	Sandu, Guizhou	64	0	0	This study
Bouyei	By	Zhenning, Guizhou	26	0	0	This study
Zhuangzu	Zhuang	Guangxi	83	0	0	Yao et al. (2002c)
Mulam	Mulam	Luocheng, Guangxi	36	2	5.6	This study
Jingzu	Jing	Guangxi	61	5	8.2	This study
Lizu	Li	Hainan	59	2	3.4	This study
TW Han	TW	Taiwan	155	0	0	Tsai et al. (2001)
Taiwanese	Tai	Taiwan	180	0	0	Tajima et al. (2003)
Total			3,090	76		

matching and/or near-matching with the identified F2 types that have been tested for coding region information. To better understand the relationships among haplogroups, a network profile of haplogroup F2 was constructed according to Bandelt et al. (2000). We also estimated the haplotype diversity and nucleotide diversity (Nei 1987) of haplogroup F2 by using the DnaSP package (Rozas and Rozas 1999).

Table 2 Sequence variations in 76 mtDNAs of haplogroup F2. When the analyzed sequences were identical to the revised Cambridge reference sequence, items are indicated with CRS. When (a) nucleotide change(s) was detected compared to the CRS sequence, only the number of position is indicated for transition, the number with a suffix (i.e. A, C, G, and T) for transversion, with "d" for deletion and with "+" for insertion. When sequence information was not available, items leave blank

Sample[a]	Location	Haplogroup	HVS-I (16000+)	HVS-II (30–407) (73, 263, and 315+C in addition)[b]	10171–10659 (10000+)[c]	Nucleotide at 12338	References
GD48	Guangdong	F2	CRS	235 249d			Kivisild et al. (2002)
GD7810	Zhanjiang, Guangdong	F2	261	194 235 249d 309+C	310 535 586	C	Yao et al. (2002a)
GD49	Guangdong	F2	209 304	249d 309+C			Kivisild et al. (2002)
Kaz28	Xinjiang	F2	304				Yao et al. (2000)
Miao10	Kaili, Guizhou	F2	304	195 249d	310 535 586		This study
Miao5	Kaili, Guizhou	F2	304				This study
SD10316	Tai'an, Shandong	F2	304	249d 309+C	310 535 586	C	Yao et al. (2003)
Tibet27	Qinghai	F2	304		310 535 586	C	This study
SD10301	Tai'an, Shandong	F2	304 519	249d	310 535 586	C	Yao et al. (2003)
QJ516	Qijiang, Chongqing	F2	304 327 362	249d	310 535 586	C	This study
Mg11	Qinghai	F2	217 221 304		310 535 586		Yao et al. (2002c)
WH6948	Wuhan, Hubei	F2	299 304	249d 309+C	310 535 586	C	Yao et al. (2002a)
She15	Fuquan, Guizhou	F2	051 304	195 249d 309+CC	310 535 586		This study
Man153	Inner Mongolia	F2a	187 221 291 304	42+T 195 249d 275 309+CC	310 535 586		This study
YN281	Kunming, Yunnan	F2a	051 291 304	195 249d	310 535 586		Yao et al. (2002a)
QD8147[d]	Qingdao, Shandong	F2a1	266 291 304	146 249d	310 535 586	C	Yao et al. (2002a)
SD10335	Tai'an, Shandong	F2a1	266G 291 304 519	249d 309+C	310 535 586	C	Yao et al. (2003)
SD10340	Tai'an, Shandong	F2a1	266G 291 304 519	249d 309+CC	310 535 586	C	Yao et al. (2003)
Tu58	Qinghai	F2a1	266G 291 304 325 519	249d 309+CC	310 535 586	C	This study
Naxi80	Lijiang, Yunnan	F2a1	185 266A 291 304 519	249d	310 535 586	C	This study
Naxi83	Lijiang, Yunnan	F2a1	185 266G 304 519	249d			This study
Naxi73	Lijiang, Yunnan	F2a1	185 266A 291 304 519	195 249d			This study
Bai21	Dali, Yunnan	F2a1	185 266A 291 304				Yao et al. (2002c)
Nu17	Gongshan, Yunnan	F2a1	185 258 266A 291 304				Yao et al. (2002c)
Nu22	Gongshan, Yunnan	F2a1	185 258 266A 291 304				Yao et al. (2002c)
Nu26	Gongshan, Yunnan	F2a1	185 258 266A 291 304				Yao et al. (2002c)
Nu30	Gongshan, Yunnan	F2a1	185 258 266A 291 304				Yao et al. (2002c)
Miao1	Kaili, Guizhou	F2a2	092A 291 304				This study
Miao32	Kaili, Guizhou	F2a2	092A 291 304				This study
Mulam13	Luocheng, Guangxi	F2a2	092A 291 304				This study
Mulam15	Luocheng, Guangxi	F2a2	092A 291 304				This study
GD50	Guangdong	F2a2	092A 291 304	249d			Kivisild et al. (2002)
SD10360	Tai'an, Shandong	F2a2	092A 291 304	249d	310 535 586		Yao et al. (2003)
XJ8414	Yili, Xinjiang	F2a2	092A 291 304	249d 309+CC	310 535 586	C	Yao et al. (2002a)
QJ534	Qijiang, Chongqing	F2a2	092A 291 299 304 362 519	249d 309+C	310 535 586	C	This study
SD10302	Tai'an, Shandong	F2a2	092A 289 291 304	249d 309+CC	310 535 586	C	Yao et al. (2003)
GD7836n	Zhanjiang, Guangdong	F2a2	092A 291 304 359	249d 309+C	310 535 586		Yao et al. (2002a)
Jing100	Guangxi	F2a2	092A 291 304 311	195 249d		C	This study
Jing126	Guangxi	F2a2	092A 291 304 311	195 249d			This study
Jing140	Guangxi	F2a2	092A 291 304 311	195 249d			This study
Jing73	Guangxi	F2a2	092A 291 304 311	195 249d		C	This study
Jing83	Guangxi	F2a2	092A 291 304 311	195 249d			This study
Lisu12	Gongshan, Yunnan	F2a2	092A 170C 189 291 294 304				Yao et al. (2002c)

Table 2 (Continued)

Sample[a]	Location	Haplogroup	HVS-I (16000+)	HVS-II (30–407) (73, 263, and 315+C in addition)[b]	10171–10659 (10000+)[c]	Nucleotide at 12338	References
Dai51	Xishuangbanna, Yunnan	F2a2	092A 170T 189 291 304				Yao et al. (2002c)
Lisu34	Gongshan, Yunnan	F2a2	092A 170T 189 291 304				Yao et al. (2002c)
Lisu45	Gongshan, Yunnan	F2a2	092A 170T 189 291 304				Yao et al. (2002c)
Lisu30	Gongshan, Yunnan	F2a2	092A 170T 183T 189 291 304 400				Yao et al. (2002c)
Nu14	Gongshan, Yunnan	F2a2	092A 170T 183T 189 291 304				Yao et al. (2002c)
XJ8407	Yili, Xinjiang	F2a3	203 239 291 304	249d 309+C	310 535 586	C	Yao et al. (2002a)
YND303	Kunming, Yunnan	F2a3	203 291 304 519	249d 309+C	310 535 586	C	This study
Tu40	Qinghai	F2a3	203 291 304 519	249d 309+C			This study
Hui38	Yili, Xinjiang	F2a3	093 203 291 304		310 535 586	C	This study
QH9477	Qinghai	F2a3	093 203 291 304		310 535 586	C	This study
Mg242	Inner Mongolia	F2a3	126 203 291 304	249d 309+C	310 535 586	C	Kong et al. (2003a)
QHD20	Qinghai	F2a3	126 203 291 304				This study
GD7842	Zhanjiang, Guangdong	F2b	129 189 304	207 249d	310 535 586	C	Yao et al. (2002a)
Mg46	Yili, Xinjiang	F2b	203 304		310 535 586	C	This study
Tibet34	Qinghai	F2b	203 304		310 535 586	C	This study
LN7601	Fengcheng, Liaoning	F2b	129 203 304	195 249d	310 535 586	C	Yao et al. (2002a)
YNC13	Kunming, Yunnan	F2b	093 203 304 519	249d	310 535 586	C	This study
GD7809[d]	Zhanjiang, Guangdong	F2b	086 203 304	249d	310 535 586	C	Yao et al. (2002a)
Lahu95	Yunnan	F2b	086 167 203 304 519				Yao and Zhang (2002)
Lahu63	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Lahu65	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Lahu74	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Lahu81	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Lahu82	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Lahu84	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Lahu87	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Lahu88	Yunnan	F2b	086 167 203 304 318				Yao and Zhang (2002)
Sani21	Nuxi, Yunnan	F2b	086 167 203 304 318 474+G				Yao et al. (2002c)
Li39	Hainan	F2c	304 527	249d 309+C	265 310 535 586	C	This study
Li50	Hainan	F2c	304 527	249d 309+C	265 310 535 586	C	This study
GD124	Meizhou, Guangdong	F2c	304 527	249d	265 310 325 535 586	C	This study
WH6974	Wuhan, Hubei	F2c	192 304	249d	265 310 535 586	C	Yao et al. (2002a)
YNC105	Kunming, Yunnan	F2c	086 304G	151 152 249d 309+C	265 310 535 586 598	C	This study

[a]The abbreviation for each sample is indicated in Table 1
[b]All the samples analyzed had 73, 263, and 315+C in addition to (a) change(s) indicated. For a example, GD48 had 73, 235, 249d, 263, and 315+C
[c]For the indicated samples, region 10171–10659 was sequenced in this study except for those reported by Yao et al. (2002a)
[d]Samples QD8147 and GD7809 were completely sequenced by Kong et al. (2003b)

Results

Identification of mtDNA haplogroup F2

As had been revealed by our recent study (Kong et al. 2003b), all the characteristic nucleotide variations for haplogroup F2 were located in the coding regions. Thus, it is generally hard to select F2 mtDNAs solely on the basis of sequence information from the hypervariable segments in the control region, but F2a is easily recognized by the variation at the 16291 position from the other F haplogroups and R9* (Yao et al. 2002a). By sequencing of the segment 10171–10659, it is easy to distinguish the haplogroup F2 from the rests by nucleotide variations at 10310, 10535, and 10586. As a result, a total of 76 haplogroup F2 mtDNAs were identified (Table 2). Comparing with 19 other mtDNAs belonging to other haplogroups (Table 3), several features are discernible. (1) Five mtDNAs with the variations at both 10310 and 10609 were assigned to haplogroup F1, and their haplogroup status was further confirmed by the variation at 12406. Two of these had motif 16302–16304–16497–249d, and three others were with motif 16172–16304–249d. (2) Ten mtDNAs bearing only the variation at 10310 in segment 10171–10659 was assigned to (an) unidentified lineage(s) in F. Seven of them had motif 16207–16304–16399–146–249d, and three others had motif 16218–16304–16311–249d. Further analysis showed that the seven mtDNAs with motif 16207–16304–16399–146–249d were characterized by two specific variations at 12396 and 12408, hence belonging to a new haplogroup designated as "F4." (3) The last four samples (two with motif 16157–16256–16304–16335–236–249d and the other two with motif 16304–16362) were found not to have any variations characteristic to F1, F2, F3 (Kong et al. 2003b), or F4, thus belonging to a new lineage (viz., pre-F) in haplogroup R9. Additional information is needed to further specify the phylogenetic positions of these samples.

Table 3 Sequence variations in non-F2 type mtDNAs. When the analyzed sequence was identical to the revised Cambridge reference sequence, items are indicated with CRS. When (a) nucleotide change(s) was detected compared to the CRS sequence, only the number of position is indicated for transition, the number with a suffix (i.e. A, C, G, and T) for transversion, with "d" for deletion and with "+" for insertion. When sequence information was not available, items leave blank

Sample[a]	Location	Haplo-group	HVS-I (16000+)	HVS-II (30–407) (73, 263, and 315+C in addition)[b]	10171–10659 (10000+)[c]	Nucleotide at 12338	Other polymorphisms	References
Man10	Fengcheng, Liaoning	F1	302 304 497		310 609	T	12406	This study
Man17	Fengcheng, Liaoning	F1	302 304 497		310 609	T	12406	This study
Li26	Hainan	F1	172 187 304 343 519	249d 309+C	310 609	T	12406	This study
QJ81	Qijiang, Chongqing	F1	172 304 465 519	249d 309+C	310 609	T	12406	This study
Zhuang73	Luocheng, Guangxi	F1	172 304 465					Yao et al. (2002c)
GD0091	Chaoshan, Guangdong	F4	207 304 399	146 152 281 249d 309+C	310	T		This study
YND617	Kunming, Yunnan	F4	179 207 304 399	146 249d 309+CC	310	T		This study
QH9678	Qinghai	F4	207 304 362 399		310	T		This study
XJ8440	Yili, Xinjiang	F4	207 304 362 399	146 152 249d	310	T	12396 12408	Yao et al. (2002a)
Zhuang30	Guangxi	F4	126 140 207 304 362 399					Yao et al. (2002c)
YNC237	Kunming, Yunnan	F4	207 304 362 399 497	146 152 249d	310	T	12396 12408	This study
YND379	Kunming, Yunnan	F4	093 207 304 362 399 497	146 152 207 249d 309+C	310	T	12396 12408	This study
Mulam6	Luocheng, Guangxi	F*	218 304 311		310	T		This study
Li56	Hainan	F*	218 304 311 519	249d 309+C	310	T		This study
Zhuang43	Guangxi	F*	129 218 265 304 311 355					Yao et al. (2002c)
Mulam42	Luocheng, Guangxi	R9*	304 362	CRS	CRS	T		This study
Dai100	Xishuangbanna, Yunnan	R9*	304 362	CRS	CRS	T		Yao et al. (2002c)
She61	Fuquan, Guizhou	R9*	157 256 266 304 311 335	236 249d	CRS			This study
YN170	Kunming, Yunnan	R9*	157 256 304 335	236 249d	CRS			Yao et al. (2002a)

[a]The abbreviation for each sample is indicated in Table 1

[b]All the samples analyzed had 73, 263, and 315+C in addition to (a) change(s) indicated. Mulam42 and Dai100 had 73, 263, and 315+C, too

[c]For the indicated samples, region 10171–10659 was sequenced in this study except for except for XJ8440 and YN170

Phylogeny of the haplogroup F2

The network profile of haplogroup F2 revealed that this haplogroup was divided into three major branches, designated as "F2a" (defined by the variation at 16291; Yao et al. 2002a), "F2b" (recognizable by the variation at 10810; Table 4), and "F2c" (characterized by the variation at 10265) (Fig. 1). Haplogroup F2a comprises three major clades, F2a1 (recognizable by the variation at 16266), F2a2 (characterized by a transversion (T/A) at 16092), and F2a3 (defined by the variation at 16203). It is evident that the sub-clades of F2 show regional distribution. For instance, most of the F2b and F2a3 types, as well as all the F2a1 types, are confined to north or north-origin populations. In contrast, individuals belonging to haplogroup F2c are prevalent in south China.

T12338C is characteristic of haplogroup F2

Our previous analyses of the major haplogroups in East Asian by using complete sequences revealed that the T12338C variation is exclusively detected in two samples of haplogroup F2 (GD7809 and QD8147; Kong et al. 2003b). Extensive search with more than 1,000 published complete mtDNA sequences (Ingman et al. 2000; Finnilä et al. 2001; Maca-Meyer et al. 2001; Torroni et al. 2001; Derbeneva et al. 2002; Herrnstadt et al. 2002; Kivisild et al. 2002 and references therein; Ingman and Gyllensten 2003; Mishmar et al. 2003) revealed two additional cases with the T12338C variation, with one in a P type (Ingman and Gyllensten 2003) and another in a H1 type (Herrnstadt et al. 2002). To further clarify whether the T12338C variation is characteristic of haplogroup F2 or not, 32 samples were selected from our collection described in Table 2 so as to include at least one sample from respective sub-clades, and subjected to PCR amplification and subsequent direct sequencing for detecting the variation. As a control, 14 samples belonging to non-F2 haplogroups selected from the samples listed in Table 3 were also analyzed. Our results revealed that the T12338C nucleotide substitution was detected in all the F2 samples analyzed but not in any non-F2 samples (Tables 2, 3). Further analyses showed that the T12338C variation was completely linked with the T1005C, T1824C, A7828G, T10535C, G10586A, and G13708A variations (Table 4), thus showing T12338C is one of the characteristic variations specific to haplogroup F2.

Discussion

In this study, we report a homoplasmic nucleotide change, T12338C, which results in substitution of highly conserved methionine at the translation start site of the ND5 gene with threonine (M1T). The T12338C change is tightly associated with other nucleotide variations

Table 4 Relevant nucleotide variations in major sub-clades of haplogroup F2. Only varied nucleotides are indicated with a letter while the identical nucleotide to rCRS is indicated with dots. When information for nucleotide sequences is not available, the items have been left blank

Haplogroup	Representative sample	Nucleotide variations at the indicated position																										
		1005	1709	1824	3290	4811	5417	5846	6392	7828	8636	9137	10310	10535	10586	10801	10810	12338	12609	12366	13500	13708	13928	14016	14560
H	rCRS	T	G	T	T	A	G	C	T	A	T	T	G	T	G	G	T	T	T	A	T	G	G	G	G			
F2a1	QD8147	C	A	C	C	G	A	T	C	G	C	C	A	C	A	A	.	C	C	.	C	A	C	C	A	.		
F2a1	Tu58	C	.	C	.	.	.	.	C	G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.			
F2a1	Naxi80	C	.	C	.	.	.	.	C	G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.			
F2a2	XJ8414	C	.	C	.	G	.	.	C	G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.			
F2a3	XJ8407	C	.	C	.	.	.	.	C	G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.			
F2b	GD7809	C	.	C	.	.	.	.	C	G	.	.	A	C	A	.	C	C	.	G	.	A	C	.	.	A		
F2b	Mg46	C	.	C	.	.	.	.		G	.	.	A	C	A	.	C	C	.	.	.	A	C	.	.	A		
F2b	LN7601	C	.	C	.	.	.	.		G	.	.	A	C	A	.	C	C	.	.	.	A	C	.	.	.		
F2b	GD7842	C	.	C	.	.	.	.		G	.	.	A	C	A	.	C	C	.	.	.	A	C	.	.	.		
F2a3	Mg242	C	.	C	.	.	.	.		G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.	.		
F2a3	QH9477	C	.	C	.	.	.	.		G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.	.		
F2a3	Hui38	C	.	C	.	.	.	.		G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.	.		
F2c	WH6974	C	.	C	.	.	.	.		G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.	.		
F2	WH6948	C	.	C	.	.	.	.		G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.	.		
F2	GD7810	C	.	C	.	.	.	.		G	.	.	A	C	A	.	.	C	.	.	.	A	C	.	.	.		

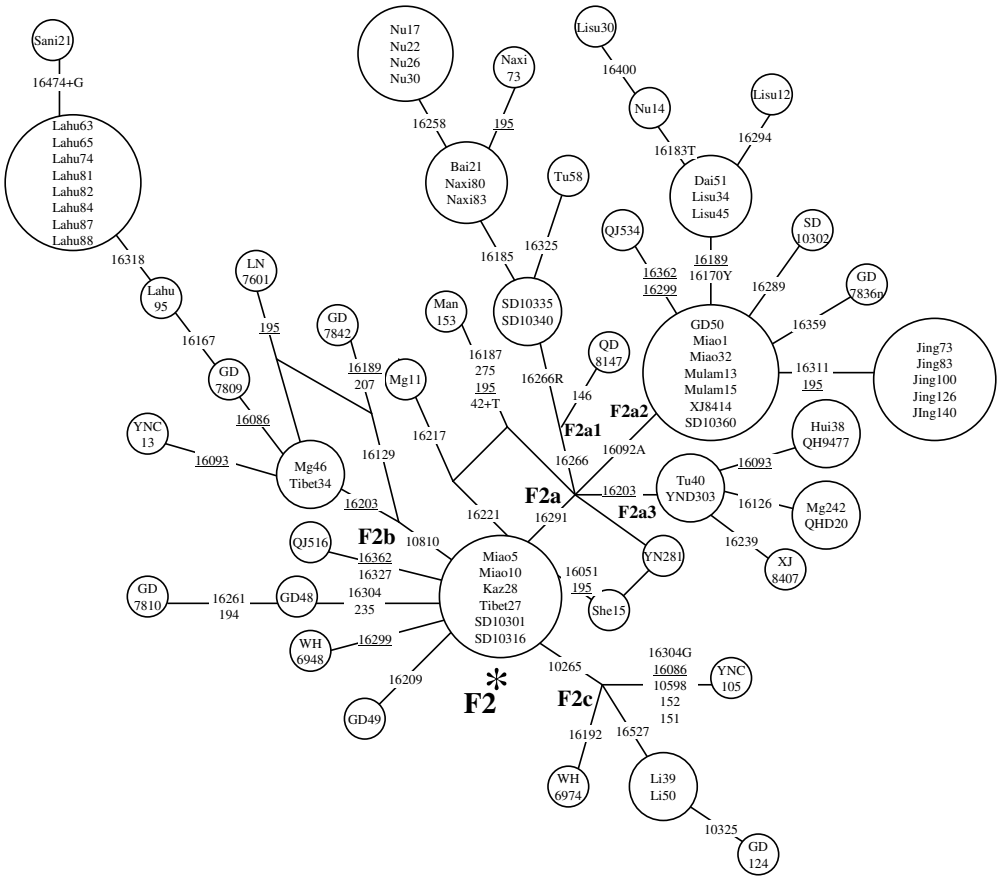

Fig. 1 Network profile of haplogroup F2 samples observed in Chinese. The network is constructed on the basis of 76 F2 mtDNAs identified in 57 Chinese ethnic populations. As the nucleotide at site 16519 is known to be extremely hypervariable, it is disregarded in the construction. Representative sample names are indicated in the *circles*, and the circle size is proportional to the number of samples to be belonged among the pool. Relevant nucleotide variations compared with the revised Cambridge Reference Sequence (rCRS) indicate on the branches in the number of the position. Recurrent mutations are *underlined*, and the *asterisk* indicates the root of haplogroup F2

including T1005C, T1824C, A7828G, T10535C, G10586A, and G13708A. Thus, these variations together characterize the F2 haplogroup. Considering the fact that mitochondria of haplogroup F2 was distributed widely in normal populations across China though with relatively low frequencies (Table 1), and no evidence suggesting any association of the haplogroup with mitochondrial disorders, it is clear that T12338C is a polymorphic variation rather than a pathogenic mutation. This case is similar to that of T3308C, which is specific to haplogroup L1b and has been proven to be benign (Rocha et al. 1999; Fernandez-Moreno et al. 2000). Intriguingly, since the third codon of the ND5 gene also encodes methionine, similar to that of the ND1 gene, it is plausible that the third codon of the ND1 and ND5 genes would act as a surrogate when the initiation codon was impaired (Rocha et al. 1999; Fernandez-Moreno et al. 2000). In the COX II gene, the second methionine codon in frame is located 48 nucleotides downstream of the translational initiation site. Even though this codon could take a similar proxy role as discussed for the ND1 and ND5 genes, it is still uncertain whether the resultant polypeptide with 16 residues shorter is functional. Thus, it is not surprising that a similar phenomenon caused by T7587C in the COX II gene leads to a mitochondrial disorder (Clark et al. 1999). Nucleotide changes occurring in the translational initiation codon are also observed in mtDNAs of primates (Table 5). Two and four cases have been reported for the ND1 and ND5 genes, respectively, in chimpanzees (Horai et al. 1995), macaques, and patas (Hayasaka et al. 1996). It seems that the first codon in proposed cDNA sequence for the ND1 and ND5 genes are not crucial, and the adjacent downstream ATG in frame could take a role in translational initiation.

Haplogroup is normally of higher prevalence and has more genetic diversity at the place it occurred than the radiated area. To speculate where the F2 haplogroup (viz., T12338C) originated, we simply aggregated the data form respective populations into northern and southern groups according to their locations and/or ethono-origin (we did not include samples WH Han and QJ Han because they were located in the Changjiang River region where populations migrate considerably). Although this strategy has been criticized (Yao et al. 2002a), this method would enlarge the sample size and make the comparison less affected by the sample size bias. A slightly higher frequency of haplogroup F2 was identified in the northern group (45/1,473) than in the southern group (27/1,524). The haplogroup diversity and nucleotide diversity in F2 are slightly higher in the northern group (0.940 ± 0.021 and 0.0106 ± 0.0006, respectively) than those in the southern group (0.915 ± 0.033 and 0.0075 ± 0.0008, respectively). These results suggest that the F2 haplogroup might originate in north China. This suggestion is also supported by two

Table 5 Mutations and/or variations occurring in the initiation codon of mitochondria genes. NA, not available

Changes in nucleotide sequences	Changes in amino acid sequences	Gene	Accession number	Organism	References
ATA → ACA	MPM → TPM	ND1	NA	*Homo sapiens*	Campos et al. (1997), Fernandez-Moreno et al. (2000), Polyak et al. (1998), Vilarinho et al. (1999), Rocha et al. (1999) and Opdal et al. (2002)
ATA → AGA	M → termination	ND1	NA	*Homo sapiens*	Opdal et al. (2002)
ATG → ACG	MAH → TAH	COX II	NA	*Homo sapiens*	Clark et al. (1999)
ATG → GTG	MNE → VNE	ATP6	AY370877	*Homo sapiens*	Dubot et al. (2004)
ATA → ACA	MTM → TTM	ND5	AY255168	*Homo sapiens*	Kong et al. (2003b)
ATA → ACA	MTM → TTM	ND5	AY255180	*Homo sapiens*	Kong et al. (2003b)
ATA → ACA	MTM → TTM	ND5	AY289092	*Homo sapiens*	Ingman and Gyllensten (2003)
ATA → ACA	MTM → TTM	ND5	NA	*Homo sapiens*	Herrnstadt et al. (2002)
ATA → ACA	MPM → TPM	ND1	NC_001643	*Pan troglodytes*	Horai et al. (1995)
ATA → ACA	MPM → TPM	ND1	NC_001644	*Pan paniscus*	Horai et al. (1995)
ATA → ACA	MIM → TIM	ND5	D85291	*Erythrocebus patas*	Hayasaka et al. (1996)
ATA → ACA	MIM → TIM	ND5	D85287	*Macaca nigra*	Hayasaka et al. (1996)
ATA → GTA	MIM → VIM	ND5	D85285	*Macaca nemestrina*	Hayasaka et al. (1996)
ATA → GTA	MIM → VIM	ND5	D85286	*Macaca silenus*	Hayasaka et al. (1996)

features in the phylogeographic analysis of F2 (Fig. 1). (1) The potential root types of F2 were more prevalent in the populations from north China or of the northern origin. (2) Almost all the major sub-clades of F2 were distributed in the northern populations except for F2c (which is exclusive to south China). By counting the transitions in region 16090–16365 (Forster et al. 1996; Saillard et al. 2000), we calculated the age of haplogroup F2 to be $41{,}700 \pm 13{,}700$ years. These results suggested that haplogroup F2 might originate and expand in north China before the last Glacier Age. The prevalence of sub-haplogroups of F2, for examples F2c and F2a2, in south China might reflect back-migration events from north China to south China afterwards (Yao et al. 2002a).

Our current study also raises a concern in identifying pathogenic mtDNA mutations. Several surveys have revealed that mutation C5178A is associated with longevity and other disorders (Kokaze et al. 2003 and references therein). According to our study, the C5178A variation is exclusive associated with haplogroup D (Yao et al. 2002b), and it is well known that haplogroup D is prevalent in northern Chinese (Yao et al. 2002a; Kong et al. 2003a) and Japanese (Maruyama et al. 2003). Thus, the most results of association of C5178A with, i.e., a disease phenotype would rather reflect the existence of population stratification (Ardlie et al. 2002) and/or inadequate sampling. Another paper reporting an association of G15497A with obesity (Okura et al. 2003) comprises a similar case, since G15497A together with T8200C and G15323A are characteristic to haplogroup G1, a prevalent form in northeastern Asia (Bandelt et al. 2003; Kong et al. 2003b). Such a hasty conclusion in association studies between nucleotide changes in mtDNA with disorders could be avoidable if the authors should refer to phylogenetic information of mtDNA.

In conclusion, our mutational and phylogeographical analyses with human mtDNA indicate that the T12338C change is one of the characteristic variations associated with haplogroup F2, thus polymorphic, and not a pathogenic mutation.

Acknowledgements The work was supported by the National Key Natural Science Foundation Program of China, Chinese Academy of Sciences (KSCX2-SW-2010), Natural Science Foundation of China (NSFC), and Natural Science Foundation of Yunnan Province.

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

Ardlie KG, Lunetta KL, Seielstad M (2002) Testing for population subdivision and association in four case-control studies. Am J Hum Genet 71:304–311

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

Bandelt H-J, 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 67:512–524

Campos Y, Martin MA, Rubio JC, Gutierrez del Olmo MC, Cabello A, Arenas J (1997) Bilateral striatal necrosis and MELAS associated with a new T3308C mutation in the mitochondrial ND1 gene. Biochem Biophys Res Commun 238:323–325

Chinnery PF, Schon EA (2003) Mitochondria. J Neurol Neurosurg Psychiatry 74:1188–1199

Clark KM, Taylor RW, Johnson MA, Chinnery PF, Chrzanowska-Lightowlers ZM, Andrews RM, Nelson IP, Wood NW, Lamont PJ, Hanna MG, Lightowlers RN, Turnbull DM (1999) An mtDNA mutation in the initiation codon of the cytochrome C oxidase subunit II gene results in lower levels of the protein and a mitochondrial encephalomyopathy. Am J Hum Genet 64:1330–1339

Derbeneva OA, Sukernik RI, Volodko NV, Hosseini SH, Lott MT, Wallace DC (2002) Analysis of mitochondrial DNA diversity in the aleuts of the commander islands and its implications for the genetic history of beringia. Am J Hum Genet 71:415–421

DiMauro S, Schon EA (2001) Mitochondrial DNA mutations in human disease. Am J Med Genet 106:18–26

Dubot A, Godinot C, Dumur V, Sablonniere B, Stojkovic T, Cuisset JM, Vojtiskova A, Pecina P, Jesina P, Houstek J (2004) GUG is an efficient initiation codon to translate the human mitochondrial ATP6 gene. Biochem Biophys Res Commun 313:687–693

Fernandez-Moreno MA, Bornstein B, Campos Y, Arenas J, Garesse R (2000) The pathogenic role of point mutations affecting the translational initiation codon of mitochondrial genes. Mol Genet Metab 70:238–240

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

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Hayasaka K, Fujii K, Horai S (1996) Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. Mol Biol Evol 13:1044–1053

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

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536

Ingman M, Gyllensten U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res 13:1600–1606

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

Kivisild T, Tolk HV, 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

Kokaze A, Ishikawa M, Matsunaga N, Yoshida M, Sekine Y, Sekiguchi K, Satoh M, Harada M, Teruya K, Takeda N, Uchida Y, Tsunoda T, Takashima Y (2003) Longevity-associated mitochondrial DNA 5178 A/C polymorphism modulates effects of daily drinking and cigarette consumption on serum triglyceride levels in middle-aged Japanese men. Exp Gerontol 38:1071–1076

Kong Q-P, Yao Y-G, Liu M, Shen S-P, Chen C, Zhu C-L, Palanichamy MG, Zhang Y-P (2003a) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. Hum Genet 113:391–405

Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P (2003b) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73:671–676

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

Maruyama S, Minaguchi K, Saitou N (2003) Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. Int J Legal Med 117:218–225

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100:171–176

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

Okura T, Koda M, Ando F, Niino N, Tanaka M, Shimokata H (2003) Association of the mitochondrial DNA 15497G/A polymorphism with obesity in a middle-aged and elderly Japanese population. Hum Genet 113:432–436

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

Opdal SH, Vege A, Egeland T, Musse MA, Rognum TO (2002) Possible role of mtDNA mutations in sudden infant death. Pediatr Neurol 27:23–29

Polyak K, Li Y, Zhu H, Lengauer C, Willson JK, Markowitz SD, Trush MA, Kinzler KW, Vogelstein B (1998) Somatic mutations of the mitochondrial genome in human colorectal tumours. Nat Genet 20:291–293

Rocha H, Flores C, Campos Y, Arenas J, Vilarinho L, Santorelli FM, Torroni A (1999) About the "Pathological" role of the mtDNA T3308C mutationellipsis. Am J Hum Genet 65:1457–1459

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

Saillard J, Forster P, Lynnerup N, Bandelt H-J, Norby S (2000) mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am J Hum Genet 67:718–726

Tajima A, Sun CS, Pan IH, Ishida T, Saitou N, Horai S (2003) Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. Hum Genet 113:24–33

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69:1348–1356

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

Vilarinho L, Chorao R, Cardoso ML, Rocha H, Nogueira C, Santorelli FM (1999) The ND1 T3308C mutation may be a mtDNA polymorphism. Report of two Portuguese patients. J Inherit Metab Dis 22:90–91

Wallace DC, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. Gene 238:211–230

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

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, Kong Q-P, Zhang Y-P (2002b) Mitochondrial DNA 5178A polymorphism and longevity. Hum Genet 111:462–463

Yao Y-G, Nie L, Harpending H, Fu Y-X, Yuan Z-G, Zhang Y-P (2002c) 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 (2003) Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. Mol Biol Evol 20:214–219