doi: 10.1046/j.1529-8817.2003.00049.x

Identification of Native American Founder mtDNAs Through the Analysis of Complete mtDNA Sequences: Some Caveats

H.-J. Bandelt[1,*], C. Herrnstadt[2], Y.-G. Yao[3], Q.-P. Kong[3], T. Kivisild[4,5], C. Rengo[6], R. Scozzari[7], M. Richards[8], R. Villems[4], V. Macaulay[9], N. Howell[2,10], A. Torroni[11] and Y.-P. Zhang[3]

[1] *Fachbereich Mathematik, Universität Hamburg, Hamburg, Germany*
[2] *MitoKor, San Diego, California, USA*
[3] *Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China*
[4] *Department of Evolutionary Biology, Tartu University and Estonian Biocentre, Tartu, Estonia*
[5] *Department of Genetics, Stanford University, Stanford, California, USA*
[6] *Istituto di Medicina Legale, Università Cattolica del Sacro Cuore, Rome, Italy*
[7] *Dipartimento di Genetica e Biologia Molecolare, Università "La Sapienza", Rome, Italy*
[8] *Department of Chemical and Biological Sciences, University of Huddersfield, Huddersfield, UK*
[9] *Department of Statistics, University of Oxford, Oxford, UK*
[10] *Department of Radiation Oncology, University of Texas Medical Branch, Galveston, Texas, USA*
[11] *Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy*

Summary

In this study, a detailed analysis of both previously published and new data was performed to determine whether complete, or almost complete, mtDNA sequences can resolve the long-debated issue of which Asian mtDNAs were founder sequences for the Native American mtDNA pool. Unfortunately, we now know that coding region data and their analysis are not without problems. To obtain and report reasonably correct sequences does not seem to be a trivial task, and to discriminate between Asian and Native American mtDNA ancestries may be more complex than previously believed. It is essential to take into account the effects of mutational hot spots in both the control and coding regions, so that the number of apparent Native American mtDNA founder sequences is not erroneously inflated. As we report here, a careful analysis of all available data indicates that there is very little evidence that more than five founder mtDNA sequences entered Beringia before the Last Glacial Maximum and left their traces in the current Native American mtDNA pool.

Introduction

Within the past three years, the sequencing of the entire human mitochondrial DNA (mtDNA), previously performed almost exclusively in patients to detect disease-associated mtDNA mutations, has become a feasible tool also for population and evolutionary studies. This approach, however, cannot and should not replace the wealth of previously gathered mtDNA information; rather, complete mtDNA genome analysis will confirm and refine the picture obtained so far. In particular, the origin and evolution of Native American mtDNA was first assessed by high-resolution RFLP analysis as well as through sequencing of the hypervariable segment I (HVS-I) of the control region. The results of those studies led to the view that all extant Native American mtDNAs (if not of recent admixture) descended from five founder haplogroups, A–D (Torroni *et al*. 1993) and X (Brown *et al*. 1998).

*Correspondence: Prof. Dr. Hans-Jürgen Bandelt, FB Mathematik, Universität Hamburg, Bundesstr. 55, 20146 Hamburg, Germany. Fax: +49-40-42838-5190. E-mail: bandelt@math.uni-hamburg.de

 © University College London 2003

Four recent reports potentially impact our understanding of the Native American mtDNA founders. Herrnstadt *et al.* (2002) sequenced the entire coding region of a large mixed UK/US sample, Silva *et al.* (2002) sequenced the fragment of the mitochondrial genome that spans nucleotides 7148–15976 of the coding region in a small South American sample, Derbeneva *et al.* (2002) analyzed complete mtDNA sequences from a specific Native American group, the Aleuts of the Commander Islands, and Mishmar *et al.* (2003) obtained complete mtDNA sequences from a small representative worldwide sample selected according to haplogroup status. In the first study, the mtDNAs of Native (North) American ancestry were distinguished from a large number of mt-DNAs of diverse continental origins by reference to previously reported – and incomplete – classification criteria that utilized a small number of polymorphisms in the coding region and HVS-I (Herrnstadt *et al.* 2002). Unfortunately, this approach is not entirely reliable in terms of geographical assignment. In contrast, the samples of Silva *et al.* (2002) and Derbeneva *et al.* (2002) were taken directly, for the most part, from Native American groups, so that the assignment problem was not as acute. The data of Mishmar *et al.* (2003) include a total of 13 members of haplogroups A, B, C, D and X, each selected from individuals of Asian and Native American ancestries. The sequencing results of Silva *et al.* (2002), however, provide no more than an approximate picture of Native American mtDNA variation because of the large number of errors in the reported sequences (as we will discuss below; Yao *et al.* 2003b). With the coding region information provided by Herrnstadt *et al.* (2002) and the complete mtDNA sequences of Mishmar *et al.* (2003), we can now derive more complete, and less ambiguous, characteristics of the Native American mtDNA pool.

Material and Methods

DNA Amplification and Sequencing

The polymorphisms in HVS-I, hypervariable segment II (HVS-II), and segments 395–901, 4500–5007, 6010–6710, 10171–10659, 13304–13665, and 14576–16047, as well as at the diagnostic sites 5176 *Alu*I and 4831 *Hha*I, were determined according to the procedures described elsewhere (Maca-Meyer *et al.* 2001; Yao *et al.* 2002).

The complete sequence of the Native American mtDNA belonging to haplogroup X was determined as described by Torroni *et al.* (2001).

A total of 30 complete Asian/Native American mtDNAs, including both coding and control regions, from the MitoKor database are discussed in the present report. These sequences were determined as described previously (Herrnstadt *et al.* 2002), and they can be obtained from the MitoKor website (www.mitokor.com/science/30asianmtdnas.php).

mtDNA Haplogroup Notation

A database of complete mtDNA sequences of Asian and Native American ancestries has been surveyed and classified by Kivisild *et al.* (2002; see Figures 1 and 2 therein). We employ their classification scheme but expand and modify the haplogroup nomenclature, in the case of haplogroups B4 and G, in view of the new data obtained (see below).

Results and Discussion

Distinguishing Asian from Native American mtDNAs

Because Herrnstadt *et al.* (2002), when distinguishing Asian from Native American mtDNAs, did not include control region polymorphisms, which can discriminate between alternative mtDNA haplogroups by the presence of specific HVS-I motifs (Yao *et al.* 2002), a small number of mtDNAs were misassigned in their candidate set of Native American mtDNAs. Thus, in their Figure 2, mtDNA sequences 214 and 169 were assigned to haplogroup E, and hence both should be of Asian descent. Using the classification system of Kivisild *et al.* (2002), both of these mtDNAs should now be assigned to the larger Asian haplogroup M9 (214 to E1 and 169 to M9a) because they carry the transition at nucleotide (nt) 4491. Sequence 287, originally misassigned to haplogroup E, belongs instead to the northeast African branch M1 of haplogroup M, which is also found along the Mediterranean coast, especially the subbranch with mutations at nucleotides 16359 and 12346

© University College London 2003

Table 1 Coding region polymorphisms in Asian haplogroup B4 mtDNAs

Haplogroup	Sample[1]	Region[2]				
		375–922	4500–5007	6010–6710	13304–13665	14576–16047
B4a	GD7812	522–523del	CRS	CRS	CRS	14905
B4b1	GD7813	498 499 827	4820	6023 6216 6413	13590	15301 15535
B4b1	GD7814	499 827	4820	6023 6216 6413	13590	15301 15535
B4b1	LN7716	499 827	4820	6023 6413	13590	15236 15535
B4b1	QD8119	499 827	4820	6023 6216 6378 6413	n.d.[3]	15301 15535
B4b	XJ8428	499 827	4820	CRS	13590	14587 15314 15370 15535
B4	WH6982	556T 827	n.d.	CRS	n.d.	15535 15930
B4	LN7589	827	CRS	CRS	CRS	15038 15535 15930
B4c	LN7552	709	CRS	CRS	CRS	15346
B4c	YN154	709	n.d.	n.d.	n.d.	14687 14763C 15346

[1]Chinese mtDNA samples with HVS-I and -II information are from Yao *et al.* (2002);
[2]All sequences carry the transitions at nucleotides 750, 4769, 14766, and 15326 relative to the revised CRS;
[3]n.d. = not determined.

(the latter recognizable as a *Rsa*I site (Quintana-Murci *et al.* 1999)).

All mtDNA sequences designated as haplogroup B in Figure 2 of Herrnstadt *et al.* (2002) should now be assigned to haplogroups B4b (those labelled B1 and B2 in Figure 2 of Herrnstadt *et al.* 2002) and B4c. In order to clarify the relationship of these mtDNAs to those of confirmed Asian origin, we have analyzed a subset of mtDNAs from the set of Yao *et al.* (2002) for four non-contiguous spans of the coding region. The results (Table 1) show that branch B4b1, characterized by the transition at nt 16136 in HVS-I, has a sub-branch that carries mutations at nucleotides 6023 and 6413 in the coding region. Because the three haplogroup B1 sequences 213, 220, and 372 of Herrnstadt *et al.* (2002) all carry the 16136 transition (as well as a transition at nt 207) that is virtually absent in the Native American HVS-I database, and because these mtDNAs carry the same subset of coding region mutations as does Asian sequence GD7814 in Table 1, we conclude that these three sequences are likely to be of recent Asian origin.

The single B4c member (sequence 178 of Herrnstadt *et al.* 2002), as identified by mutations at nucleotides 1119 and 15346, may also be of Asian origin, but the HVS-I pattern of the Asian B4c variation is not yet well understood because the root HVS-I motif of B4c is indistinguishable from that of B4. From other studies, one sub-branch of B4c can be identified on the basis of HVS-I variation; *viz*, a transition at nt 16147 that occurs both in Central Asian mtDNAs (Comas *et al.* 1998; Ingman *et al.* 2000) and in two individuals of Dai descent from Yunnan Province, China (Yao & Zhang, 2002). B4c also occurs in Japan as inferred from a polymorphism at nt 15346 (Fuku *et al.* 2002; Tanaka *et al.* 2002). Future screening of haplogroup B sequences that are directly drawn from Native American populations should allow us to conclude with more certainty whether B4c sequences contributed to the set of Native American founder mtDNAs.

The HVS-I motif 16223–16325–16362 typical of Native American haplogroup D mtDNAs is also found in Asia. The three sequences with this motif that appear in the Korean data of Snäll *et al.* (2002), however, tested positively for the RFLP site +4830 *Hae*II (recognizing a transition at nt 4833) of haplogroup G. We additionally found the 4833 transition (confirmed by +4831 *Hha*I) in four mtDNAs from China (Table 2). Because all of these mtDNA sequences that belong to haplogroup G have (near-)matches with sequences sampled in Japan (Horai *et al.* 1996; Imaizumi *et al.* 2002; Koyama *et al.* 2002), Korea (Pfeiffer *et al.* 1998; Lee *et al.* 1997, 2002), and China (Tsai *et al.* 2001; Yao *et al.* 2000, 2003a), it appears that these sequences also belong to haplogroup G. The vast majority of these HVS-I and HVS-II sequences share the 150 polymorphism. In particular, eleven HVS-I and HVS-II sequences from the Korean sample of Lee *et al.* (1997) carry the partial motif 16223–16325; seven of them have the additional part 16362–150 of the motif,

 © University College London 2003

Table 2 Polymorphisms in Asian haplogroups G1 and C1 mtDNAs

		Region[2]				
Haplogroup	Sample[1]	HVS-I 16001–16497 (16000+)	HVS-II 30–407 (all with 73, 263)	10171–10659	14576–16047 (all with 14766, 15326)	RFLP 5176a/4831f
G1a	DW45	223 325 362	n.d.[2]	n.d.	n.d.	+/+
G1a	DW48	223 325 362	150 315+C	10398 10400	14783 15043 15301 15323 15497 15860	+/+
G1a	Kor47	223 325 362	150 315+C	10398 10400	14783 15043 15299 15301 15323 15497 15860	+/+
G1a	SD10298	223 325 362	150 315+C	n.d.	14783 15043 15301 15323 15497 15860	+/+
G1b	as7G	017 129 223	207	10398 10400	14783 15043 15301 15323 15497 16017	+/+
C1	DW41	093 223 298 325 327 356	93 249del 290–291del 309+C 315+C	n.d.	n.d.	n.d.

[1]Samples with prefix DW and Kor are from the Daur and Korean populations of Inner Mongolia (China). The sequence information about samples SD10298 and as7G is derived from table 1 of Yao *et al*. (2003a) and the data underlying figure 1 of Mishmar *et al*. (2003), respectively;
[2]n.d. = not determined.

whereas two sequences lack the transition at nt 16362 and another two at 150. This Korean data set, however, is known to contain errors (Bandelt *et al*. 2001). There is also confusion because a more recent study by the same investigators (Lee *et al*. 2002) shows a reverse pattern for the 16362 polymorphism, so that out of nine mtDNAs with the partial motif 16223–16325 seven lack the 16362 transition, whereas the other two (P173 and H120) were evidently taken from the earlier study (Lee *et al*. 1997). In the mitochondrial cytochrome *b* gene, these nine Korean mtDNAs (Lee *et al*. 2002) are characterized by three transitions that occurred at nucleotides 15323, 15497, and 15860 – notwithstanding three instances of unrecorded mutations at 15323 and 15497. This association between control region motifs and cytochrome *b* motifs was not disrupted by artificial recombination that affected other sequences from this data set (for example, sequences H81, H98, F531.2). These three mutations in the cytochrome *b* gene (*MT-CYB*) are also jointly observed in seven mtDNAs sampled from Japan (Fuku *et al*. 2002; Tanaka *et al*. 2002; note that both studies overlap in 96 Cyt*b* sequences of young obese adults, without cross-reference). Further analysis of the G haplotype (16223–16325–16362) from Inner Mongolia, China, confirms the presence of the three *MTCYB* mutations (Table 2). Two of these mutations (at nt 15323 and 15497), but not the third (15860), also appear in the complete mtDNA sequence designated as 7g of Mishmar *et al*. (2003), as well as in two *MTCYB* sequences reported in Fuku *et al*. (2002) and Tanaka *et al*. (2002), two of which additionally share a transition at nt 14490. The former complete mtDNA sequence carries the transitions at nt 16017 and 16129 previously detected in haplogroup G mtDNAs of Koryaks and Itel'men (Schurr *et al*. 1999). We therefore propose to extend the definition of haplogroup G1 so that it embraces the branches G1a (15860–16325–150) and G1b (16017–16129). Then in Japan, there is at least one further branch of G1 in which nt 14490 is mutated. An additional mtDNA with the 16325 mutation belonging to haplogroup M but not to D (16093–16223–16278–16325–16362; sample 8.22 of Kolman *et al*. 1996) could be a member of haplogroup G1a (or G2 on the basis of a transition at nt 16278). In the absence of any partial RFLPs and HVS-II, it is difficult to decide whether a HVS-I sequence with the motif 16223–16325–16362, and additional mutations (such as one Vietnamese mtDNA from Lum *et al*. 1998), is a

© University College London 2003

member of G1a. This motif is, not unexpectedly, found twice more in 150 Japanese HVS-I sequences, but it is absent in 120 Chinese sequences of Nishimaki *et al.* (1999), although their data have to be considered with caution in view of some likely errors (Bandelt *et al.* 2001).

The picture is somewhat complicated by the fact that the 16325 polymorphism is also observed, though only very rarely, in other sub-clades of haplogroup M from east Asia. The Japanese sample of Imaizumi *et al.* (2002) contains two sequences with the following control region markers: 94–194–16093–16176–16223–16325–16362 and 146–16223–16224–16245–16292–16325–16362. Although both mtDNAs carry the HVS-I motif 16223–16325–16362, a comparison with the data from Yao *et al.* (2002) reveals that both of those sequences must belong to haplogroup D, because they are closely related to representatives of two particular sub-clades of haplogroup D4, as represented by sequences LN7575 and QD8137 and by sequence QD8131. Since these latter sequences lack the 16325 mutation, we conclude that 16325 has mutated in parallel in those two Japanese sequences.

Haplogroup Specific Markers and Motifs in Native American mtDNAs

Native American haplogroup A mtDNAs apparently have two characteristic transitions in the coding region, at nucleotides 8027 and 12007, plus one at nt 16111 in HVS-I (Forster *et al.* 1996; Schurr *et al.* 1999), and as many as three transitions in HVS-II (at nucleotides 64, 146, and 153). This distinction is also confirmed by three further complete mtDNA sequences, two from Asia and one from North America (Mishmar *et al.* 2003). It should be noted, however, that the polymorphisms at nt 12007 and those in the HVS-I and -II regions appear prone to recurrent mutational events, thereby complicating their analysis. Sequences 244 and 331 of Herrnstadt *et al.* (2002), originally classified as Asian (see their figure 2), both harbour the full array of "Beringian" HVS-I and -II mutations, as well as a coding region mutation that is shared with two other sequences that were labelled Native American. Therefore, those mtDNA sequences should be assigned instead to the Native American mtDNA pool. Asian haplogroup A mtDNA sequences rarely harbour any of the 64, 146, 153, and 16111 mutations, but they typically carry the HVS-II transition at nt 152 (for example, Imaizumi *et al.* 2002; Yao *et al.* 2002), except for the A5 sub-branch (Kivisild *et al.* 2002). This result underscores the necessity of incorporating control region polymorphisms into the assignment process. The combination of several highly variable sites can be highly informative in certain instances, such as in those mtDNAs that carry the control region motif 64–146–153–16111. By searching for this motif, one can rapidly and accurately identify any haplogroup A mtDNA of Native American origin on the basis of HVS-I and HVS-II sequence data. For example, the Catalonian mtDNA set of Crespillo *et al.* (2000) clearly harbours a haplogroup A2 sequence. Because the back mutations at sites 64, 146, 153, 12007 and 16111 in haplogroup A mtDNAs appear to be essentially random, it is impossible, even using complete mtDNA sequences, to identify unambiguously Native American haplogroup A founders that differ from the haplogroup A2 root by one of these mutations (compare the discussion of Forster *et al.* 1997 on the occurrence of the 16519 polymorphism in Native American haplogroups).

It is possible that all Native American haplogroup B mtDNAs can be assigned to the B2 branch (characterized as a particular branch of B4b by mutations at nt 3547, 4977, 6473, 9950, and 11177) as defined by Herrnstadt *et al.* (2002). The two Native American mtDNA sequences reported by Ingman *et al.* (2000) and Mishmar *et al.* (2003) also contain these five characteristic changes. The 16111 polymorphism occurs among very few Native American mtDNAs other than those that belong to haplogroup A2, and thus the criterion employed by Herrnstadt *et al.* (2002) for accepting only three of the haplogroup B2 mtDNAs as being of Native American origin is invalid. It is likely that all B2 sequences are Native American, although the HVS-I motifs alone are not sufficient to prove this association. For instance, sequence 263 of Herrnstadt *et al.* (2002) matches a Kazakh mtDNA (Comas *et al.* 1998) in the HVS-I region. On the other hand, sequence 419 (previously labelled Asian) has a pronounced HVS-I motif (16189–16217–16249–16312–16344) that is found in northern Brazil (Alves-Silva *et al.* 2000). Nevertheless, through extensive screening one might find B4b

 © University College London 2003

mtDNAs in Asia that carry some of the characteristic B2 mutations, because it would be surprising if all five mutations arose during or after the migration to America.

The situation with Native American haplogroup C mtDNAs is also somewhat intricate, because they do not carry any mutations that clearly distinguish them from all of their Asian counterparts. In contrast to most Asian haplogroup C mtDNAs, however, these Native American sequences carry an HVS-I transition at nt 16325 and a deletion of the HVS-II nucleotide pair 290–291, which define a sub-branch of haplogroup C, here designated as C1. In fact, all haplogroup C sequences that were labelled Native American in Figure 2 of Herrnstadt *et al.* (2002) belong to haplogroup C1, whereas the putative Asian sequences 46 and 467 carry the 16234 and 16288 transitions instead, that were previously observed in Central Asian mtDNAs (Comas *et al.* 1998). One of the two coding region mutations (at nt 7694 and 10454) that define this specific Asian sub-clade of haplogroup C has been observed in the Asian haplogroup C sequence designated as5c of Mishmar *et al.* (2003). Interestingly, haplogroup C1 occurs in Asia as well, although rather infrequently: one mtDNA sampled in the Daur population from Inner Mongolia (China) belongs to haplogroup C1 (Table 2). This HVS-I and -II sequence perfectly matches the Japanese mtDNA jp019 from Imaizumi *et al.* (2002) and is two mutational steps from jp122 of the same data set. These three HVS-I and -II sequences share the transitions at nucleotides 16356 and 93. In the worldwide HVS-I database, the motif 16223–16298–16325–16327–16356 also occurs in the mtDNA of two Mongolian individuals (Kolman *et al.* 1996), one Kirghiz individual (Comas *et al.* 1998), and in one Japanese individual (Horai *et al.* 1996). A potential Asian C1 HVS-I haplotype, which also includes the nt 16362 mutation but not the 16356 transition, has been detected in the Bashkirs (Bermisheva *et al.* 2002). We are thus forced to conclude that there exists an Asian subbranch of C1 (in contrast to the claim of Eshleman *et al.* 2003 that the "characteristic Native American form of haplogroup C" is absent in Asia). The ancestral HVS-I motif of C1 (i.e. without the 16356 transition) is shared by two mtDNAs from Iceland (Helgason *et al.* 2000) and by one sequence each from Catalonia (Crespillo *et al.* 2000) and Tenerife (Rando *et al.* 1999), which are likely to be of Native American ancestry, particularly as the latter two data sets each contain an A2 mtDNA.

Another complicating factor is that some Native American haplogroup C mtDNAs share the 15930 transition with some Asian C mtDNAs, so that more than one branch of haplogroup C could be labelled "Native American" if the assignment process is limited solely to coding region polymorphisms. For example, Asian haplogroup C mtDNA LN7710 (Yao *et al.* 2002) carries the 15930 mutation, but neither the 16325 transition nor the 290–291 deletion. Because the 15930 transition occurs in combination with both the 16325 transition and the rare 290–291 HVS-II deletion as well as without them, we are led to conclude that the 15930 site has mutated at least twice in haplogroup C. This polymorphism seems prone to recurrence because it has also arisen in a haplogroup B4 mtDNA (sequence WH6892 in table 1), in haplogroups H (at least twice) and U5 (Herrnstadt *et al.* 2002), and in haplogroup Lod (Mishmar *et al.* 2003).

Native American haplogroup D mtDNAs belong to the major Asian branch D4 of haplogroup D, and they split into branches D1 (found in Native Americans, characterized by transitions at nucleotides 2092 and 16325) and D2 (restricted to Na-Dene, Aleuts, and Eskimo, and characterized by several mutations; Derbeneva *et al.* 2002). It should be noted that the single haplogroup D1 sequence D^{CA} reported by Derbeneva *et al.* (2002) harbours the extremely rare pathogenic mutation at 14459 (associated with both Leber hereditary optic neuropathy and dystonia), and most likely constitutes a correction of the complete mtDNA sequence obtained by Jun *et al.* (1994). The Native American mtDNA sequence na2d from Mishmar *et al.* (2003) is identical to D^{CA} except for the mutation at 13879 (which however is also missing in Fig. 4 of Derbeneva *et al.* 2002). This sequence may thus represent just another variant of the Jun *et al.* (1994) sequence. It is not clear yet whether the 16325 mutation has reverted in some Native American haplogroup D mtDNAs (such as the "Cayapa" haplotype BR53 of Alves-Silva *et al.* 2000), or whether we have to reckon with more than one Native American founder from this haplogroup. On the other hand, the HVS-I motif 16223–16325–16362 is found in Asia (mainly in conjunction with the transition at nt 150 in HVS-II), but as we have seen above,

© University College London 2003

most of these mtDNAs belong to haplogroup G1a and a few to other minor sub-clades of haplogroup M. Summarizing these results, there is no evidence at this time for the occurrence of haplogroup D1 mt-DNAs in Asia, in contrast with the situation for haplogroup C1 mtDNAs.

In addition to the major east Asian haplogroups A–D, the minor west Eurasian haplogroup X is also observed among Native (North) American mtDNAs (Forster *et al.* 1996; Scozzari *et al.* 1997; Brown *et al.* 1998; Smith *et al.* 1999). From the Native American haplogroup X data of Brown *et al.* (1998) that combine HVS-I and HVS-II with RFLPs, we infer that there are two major branches of the Native American X tree, one given by cluster V and the other represented by clusters III and IV in Fig. 1 of Brown *et al.* (1998). Hitherto, only one complete mtDNA sequence from Native American haplogroup X has been available (Mishmar *et al.* 2003). Therefore, we have completely sequenced a haplogroup X mtDNA from the Ojibwa that belongs to cluster III of Brown *et al.* (1998); see the footnote to Table 3. A comparison of this sequence with those haplogroup X mtDNA sequences of European ancestry that were published by Finnilä *et al.* (2001) and by Herrnstadt *et al.* (2002) shows that this Native American mtDNA carries all of the mutations that define haplogroup X.

Table 3 Key to Native American haplogroups (highlighted in boldface) and to some related Asian haplogroups embedded in a nesting of Eurasian mtDNA haplogroups[1]

	Characteristic mutations	
Haplogroup nesting	Coding region	Control region
N	8701, 9540, 10398, 10873, 15301	–
A	663, 1736, 4248, 4824, 8794	16290, 16319, 235
A4	–	16362
A2	8027, 12007	16111, 64, 146, 153
R	12705	16223
B	8281–8289del	16189
B4	–	16217
B4b	827, 4820, 13590, 15535	499
B4b1	–	16136
B2	3547, 4977, 6473, 9950, 11177	–
B4c	1119, 15346	–
X	1719, 6221, 6371, 13966, 14470	16189, 16278, 153, 195
X2a	8913, 12397, 14502	16213, 200
M	10400, 14783, 15043	489
M8	4715, 7196A, 8584, 15487T	16298
CZ	–	249del
C	3552A, 9545, 11914, 13263, 14318	16327
C1	–	16325, 290–291del
D	4883, 5178A	16362
D4	3010, 8414, 14668	–
D1	2092	16325
D2	3316, 7493, 8703, 9536, 11215, 11959	16129, 16271

[1]Mutations are transitions unless specified by suffixes and refer to changes from the African L3 root mtDNA. Length polymorphisms of C tracts (and compensating mutations), as well as the 16519 polymorphism, are disregarded. The haplogroup notation of Kivisild *et al.* (2002) is used for Asian mtDNAs. Native American haplogroups A2 (Forster *et al.* 1996), B2 (Herrnstadt *et al.* 2002), D1 and D2 (Forster *et al.* 1996; Saillard *et al.* 2000; Derbeneva *et al.* 2002) were previously defined, while haplogroup C1 is newly defined here. In this notational system, we thus allow the inclusions of A2 in A4, B2 in B4b, and D1, D2 in D4. The distinction of haplogroups A4 and B4 from haplogroups A and B, respectively, is problematic because the sites 16362 and 16217 may be prone to recurrent mutation. Some sites in HVS-II (such as nucleotides 64, 146, 153, 195) that belong to the corresponding consensus motif are also mutationally unstable. The Native American clade of haplogroup X is here termed X2a, and its characteristic coding region mutations are inferred from the mtDNA sequence that bears the following mutations, relative to the revised CRS, at nucleotides 73, 143, 153, 195, 200, 263, 309+C, 315+C, 522–523del, 750, 1438, 1719, 2706, 3552, 4732, 4769, 5147, 6221, 6371, 7028, 8860, 8913, 11719, 12397, 12705, 13966, 14470, 14502, 14766, 15326, 16093, 16183del, 16189, 16213, 16223, 16278, 16519.

 © University College London 2003

However, it shares three additional coding region mutations (at nt 8913, 12397, and 14502) with the other Native American X (see Table 3) that are not found in the haplogroup X mtDNAs of European origin. For these three polymorphisms, extensive surveys of haplogroup X mtDNAs are required to differentiate the polymorphisms that are Old World founder mutations from those that arose *in situ* in Beringia or North America. With those results, it will then become clear whether the transition at nt 16213 is indeed characteristic of X2a, and whether back mutations have actually occurred at these sites. Furthermore, it remains to be clarified whether the atypical Ojibwa X haplotype NA22 mtDNA (Brown *et al.* 1998) represents the descendant of a second haplogroup X founder sequence, or whether it has simply back-mutated at HVS-II site 200.

Errors in Coding Region Sequences

Unfortunately, the data reported by Silva *et al.* (2002) cannot provide an unambiguous portrayal of the Native South American mtDNA pool because the mtDNA sequence variation surveyed within the 8.8 kb fragment (nucleotides 7148–15946) appears to be inaccurately or incompletely reported (see Yao *et al.* 2003b). First, as many as 8 of the 23 basal polymorphisms in the mtDNA region sequenced that distinguish haplogroups A, B4b, C, and D4 from each other have been undetected or unreported. In addition, polymorphisms at a further 9 sites (7196A, 8584, 9540, 10398, 10400, 10873, 11914, 15043, and 15301) do not appear to have been reported correctly. Second, the original CRS (Anderson *et al.* 1981) or only a partially corrected CRS (compare with the revised CRS; Andrews *et al.* 1999) evidently served as a reference sequence, so that some polymorphisms (viz. 8860, 11335, 11719, and 15326), which distinguish the CRS from the root of haplogroup R, were misinterpreted. Third, there are also a number of phantom mutations (such as the T to G transversions at nt 11617, 12292, and 12314) as well as some erroneous missense mutations (such as rare deletions of a nucleotide at 8047, 10238, 15802, and 15848) that would likely be lethal for a living cell. Fourth, the polymorphism at nt 8805 displayed in the data matrix (Figure 2 of Silva *et al.* 2002) is not supported by the original sequences; other polymorphic sites identified by Silva *et al.* (2002) appear to have been numbered incorrectly: site 15043 is consistently referred to as 15543, and a T at site 11344 in the CRS does not exist and in fact refers to 11335, which represents an error in the original CRS as shown by Andrews *et al.* (1999). Fifth, the data matrix omits informative sites (for instance, 13803, 14318, 15535 and 15784), but, on the other hand, lists a number of uninformative sites (7626, 10792, 10793, 10804, 11654, and 11722). Fortunately, most of these problems now seem to have been resolved with a first revision of this data set (Silva *et al.* 2003), although it seems to list the polymorphic sites rather than the (parsimoniously) informative sites as claimed. There remains a suspiciously large number of recurrent mutations at the most basal polymorphisms, e.g. one at 8701, one at 14783, and as many as (at least) five independent mutations at nt 10400 (Yao *et al.* 2003c). Only one of the three phantom transversions at nt 11617, 12292, and 12314 that each occurred three times in the original data set have now disappeared completely in the revised data set: one instance of 11617G is still left and so are all three instances of 12292G. The fact that at each of the three sites the nucleotide change from T to G has flanking nucleotides T to the left and G to the right suggests that a fundamental sequencing problem has not yet been resolved completely by Silva *et al.*

The variability of the eight sites that separate haplogroups M and N (all of which fall into this 8.8kb fragment) can be evaluated using the coding region sequences published by Ingman *et al.* (2000), Finnilä *et al.* (2001), Maca-Meyer *et al.* (2001), Torroni *et al.* (2001), Herrnstadt *et al.* (2002), Derbeneva *et al.* (2002) and Mishmar *et al.* (2003), and the data clearly show a different pattern of polymorphism at these sites. Thus, we can estimate from this combined dataset that sites 8701, 9540, 10873, 10400 and 14783 have mutated only once, although site 15043 has mutated four times (with parallel mutations in haplogroups A, N1, and T2), and site 15301 has mutated six times (including an early mutation leading to super-haplogroup L2′3, and in haplogroups L1c, B4c, U5, and K1). Site 10398 is known to have undergone many recurrent mutations, but site 15301 may have undergone almost as many (see Table 1 for another parallel mutation at this site within haplogroup B4b1). It is thus not surprising to find site 15043 mutated in haplogroup B4b (Silva *et al.* 2002,

© University College London 2003

2003). On the other hand, the fact that the former five sites did not mutate in these over 900 published coding region sequences does of course not preclude the possibility of (single) rare recurrent events. Fuku *et al.* (2002), for example, found 8701 mutated in a specific (but as yet unknown) sub-clade of haplogroup M (comprising five mtDNAs) that is associated with a transition at site 15518.

The revised data from South America (Silva *et al.* 2003) do not point to more, or other, founder mtDNAs than the ones we have described, except that the 11177 mutation is only shared by two out of 10 haplogroup B2 sequences. As only one of these is from the original data set (Silva *et al.* 2002), one might suspect a sequencing problem at this site (as with 10400). Because no other haplogroup B sequences were found among Native South American mtDNAs, this finding supports the conjecture that all Native American haplogroup B mtDNAs belong to haplogroup B2 (whether or not site 11177 is among the defining mutations).

While errors in the scoring of basal polymorphisms could cause problems in the phylogenetic analysis of the mtDNA sequences, and eventually in the identification of founder mtDNAs, errors in detecting private mutations would lead to biased estimation times for founder events. Phantom mutations (Bandelt *et al.* 2002) would inflate the age estimates. It turned out that a few of the coding region sequences published by Herrnstadt *et al.* (2002) were affected by artificial C to G transversions in G-rich regions of the molecule, and by errors arising during database construction (Herrnstadt *et al.* 2003). An extreme instance is the Native American A haplogroup sequence no. 171, which carried C to G transversions at sites 7927, 7985, 14227, and 14385. The revised sequences are made available on the website of MitoKor (www.mitokor.com/science/560mtdnasrevision.php).

Conclusions

With the larger and larger mtDNA sequence sets that are being analyzed, it is inevitable that the number of errors will also increase and that these errors will confound analyses. To make future sequencing efforts more accurate, it is strongly recommended that quality assurance tests, such as independent sequencing of the same mtDNA, are performed (for example, Herrnstadt *et al.* 2002). It is also important to compare newly obtained sequences with "benchmark" data, especially at those polymorphic sites that are mutated in interior branches of the phylogeny and that are not highly variable (Bandelt *et al.* 2002). If errors are detected at such sites, then one also has to consider overlooked or misrecorded private mutations in the periphery of the mtDNA phylogeny. We would expect some oversights in the 30 complete sequences obtained by Maca-Meyer *et al.* (2001) through manual sequencing. Indeed, the sequences from haplogroups C, X, and one from H (referred to as H_2 by Maca-Meyer *et al.* 2001) each lack one of the defining mutations (viz. 15487T, 13966, and 2706, respectively).

It is axiomatic that erroneous sequence data cannot provide a reliable estimate for the time of the first human settlement of America (or Beringia), even when the errors did not affect the recognition of founder mtDNAs. Leaving aside problems inherent in the calibration of the mtDNA substitution rate (for example, Torroni *et al.* 2001), methodological shortcomings may further exacerbate the estimation of the ages of mtDNA founder clades. There is a long tradition in using nucleotide diversity π ("sequence divergence") for estimating "divergence" times (e.g., Table 4 of Chen *et al.* 2000); unfortunately, Silva *et al.* (2002, 2003) have followed this tradition. This sort of estimation, however, depends heavily on the unknown and presumably complex demographic history of the population groups and is bound to deliver biased haplogroup ages (see e.g. Forster *et al.* 1996).

Founder analysis should ideally be performed on complete mtDNA sequences and not just on coding region sequences. Indeed, in the case of the analyses of haplogroup C mtDNAs, one would have made a case for at least two distinct mtDNA founders if the control region sequence data were disregarded. On the other hand, it is clear that HVS-I data alone cannot provide sufficient evidence for founder status in view of a number of highly mutable sites, as emphasized by Malhi *et al.* (2002), although their suggestion that the entire region of HVS-I might be hypervariable has not been substantiated in our studies. With the complete mtDNA sequence information currently at hand, we fail to see solid evidence for more than four founder

 © University College London 2003

mtDNAs entering Beringia before the LGM (one each from A, B, C, and D), with the additional possibility of one from haplogroup X (Brown *et al.* 1998). This conjecture is unspectacular in view of its agreement with the very first results from RFLP studies (Torroni *et al.* 1993). One can speculate whether there might be another minor founder that differs from a major founder at a mutational hot spot, such as nt 16111 (as assumed by Forster *et al.* 1996). However, the fact remains that, as long as one can demonstrate with a complete mtDNA phylogeny that several back mutations in the respective haplogroup must have occurred at a particular site, there is insufficient evidence for founder status.

There is yet another caveat: we should not simplify the "migration" to the Americas as a single arrow drawn on a geographical map with an age attached. The process could well have been a staggered one, reaching Beringia and eventually America (perhaps more than once) before the Last Glacial Maximum (LGM). A question of equal interest is when and where the ancestral populations on the move to Beringia gradually separated from the interior Siberian and coastal east Asian peoples. Certainly, "it seems unreasonable to identify any extant group in the Old World, itself a product of many millennia of history, as the single parent population" (Eshleman *et al.* 2003). It is noteworthy that Native American haplogroup A and D mtDNAs coalesce with their Asian counterparts just at the respective root types of haplogroups A4 and D4. In the case of haplogroup B4b sequences, we currently lack sufficient information from the Asian mtDNA pool to draw an unambiguous conclusion. These results suggest a massive population increase and spread in east Asia, not long after the emergence of the A4 and D4 haplogroups, in the course of which the mitochondrial genetic roots of the later Beringians were shaped. This early period could have been during some of the short-lived climatic ameliorations, between 35,000 and 26,000 years ago (as inferred from the climatic data derived from Greenland ice cores; Daansgaard *et al.* 1993), which may have triggered the settlement of new areas in interior parts of east Asia and along the NE coast. In fact, the period ~32,000–26,000 years ago, which for Siberia is referred to as the warm "Lipovka–Novoselovo phase of the Karginsky mega-interstadial" (Vasil'ev, 2000), testifies to a rich archaeological record in the river valleys of southern Siberia. Contemporary to this assemblage, Upper Paleolithic adaptations are also manifest in sites of China (Gao & Norton, 2002). The controversial interpretations of the Y chromosomal record for the origins of Native Americans (Lell *et al.* 2002; Tarazona-Santos & Santos, 2002) possibly reflect the mosaic origin of the Upper Paleolithic population groups on the fringes of eastern Asia, rather than two or more independent migrations into North America. The ages of mtDNA haplogroups A, C, and D4 are estimated to be >30,000 years. Because we observe 2, 0, and 1 characteristic coding region mutations for haplogroups A2, C1, and D1, respectively, the transition period of gradual separation took only a few thousand years, so that the Beringian mtDNA diversification probably took place during the Lipovka–Novoselovo phase. The drop in temperature after the end of this phase could have enforced the eventual southeastward movements of some Beringian groups, which could have proceeded along the coast (Dalton, 2003). The approximate timing of the southward spread into the Americas will only become apparent through complete sequencing of numerous Native American mtDNAs from North, Central and South America. Subsequent to this initial migration process, further mtDNA sequences (from haplogroup D, which subsequently gave rise to haplogroup D2) were introduced to Northwest America from an evolved and modified Beringian mtDNA pool after the LGM, possibly as late as the beginning of the Holocene (Saillard *et al.* 2000; Derbeneva *et al.* 2002).

A number of important questions about the diversification and spread of Asian and Native American peoples still lack definitive answers. However, the increasing numbers of complete mtDNA sequences and their analysis are bringing the picture into sharper focus (see also Richards & Macaulay, 2001).

Acknowledgements

Y-GY, Q-PK, and Y-PZ are supported by the Natural Sciences Foundation of Yunnan Province and Natural Sciences Foundation of China (NSFC); VM is a Wellcome Trust Research Career Development Fellow; NH acknowledges research support from the National Science Foundation; AT received support from Telethon-Italy E.0890, the Italian Ministry of the University

© University College London 2003

(Progetti Ricerca Interesse Nazionale 2002) and Progetto CNR-MIUR Genomica Funzionale-Legge 449/97.

References

Alves-Silva, J., da Silva Santos, M., Guimarães, P.E., Ferreira, A.C., Bandelt, H.-J., Pena, S.D. & Prado, V.F. (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* **67**, 444–461, 775 (erratum).

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. & Young, I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & 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.

Bermisheva, M.A., Tambets, K., Villems, R. & Khusnutdinova, E.K. (2002) Diversity of mitochondrial DNA haplogroups in ethnic populations of the Volga–Ural region. *Molec Biol* **36**, 802–812.

Brown, M.D., Hosseini, S.H., Torroni, A., Bandelt, H.-J., Allen, J.C., Schurr, T.G., Scozzari, R., Cruciani, F. & Wallace, D.C. (1998) mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am J Hum Genet* **63**, 1852–1861.

Chen, Y.-S., Olckers, A., Schurr, G., Kogelnik, A.M., Huoponen, K. & Wallace, D. (2000) mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am J Hum Genet* **66**, 1362–1383.

Comas, D., Calafell, F., Mateu, E., Pérez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., Pettener, D. & Bertranpetit, J. (1998) Trading genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* **63**, 1824–1838.

Crespillo, M., Luque, J.A., Paredes, M., Fernández, R., Ramírez, E. & Valverde, J.L. (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. *Int J Legal Med* **114**, 130–132.

Dalton, F. (2003) The coast road. *Nature* **422**, 10–12.

Dansgaard, W., Johnsen, S.J., Clausen, H.B., Dahl-Jensen, D., Gundestrup, N.S., Hammer, C.U., Hvidberg, C.S., Steffensen, J.P., Sveinbjörnsdottir, A.E., Jouzel, J. & Bond, G. (1993) Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220.

Derbeneva, O.A., Sukernik, R.I., Volodko, N.V., Hosseini, S.H., Lott, M.T. & Wallace, D.C. (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.

Eshleman, J.A., Malhi, R.S. & Smith, D.G. (2003) Mitochondrial DNA studies of Native Americans: conceptions and misconceptions of the population prehistory of the Americas. *Evol Anthropol* **12**, 7–18.

Finnilä, S., Lehtonen, M.S. & 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.

Forster, P., Harding, R., Torroni, A. & Bandelt, H.-J. (1997) Reply to Bianci and Bailliet. *Am J Hum Genet* **61**, 246–247.

Fuku, N., Oshida, Y., Takeyasu, T., Guo, L.-J., Kurata, M., Yamada, Y., Sato, Y. & Tanaka, M. (2002) Mitochondrial ATPase subunit 6 and cytochrome *b* gene polymorphisms in young obese adults. *Biochem Biophys Res Comm* **290**, 1199–1205.

Gao, X. & Norton, C.J. (2002) A critique of the Chinese 'Middle Palaeolithic'. *Antiquity* **76**, 397–412.

Helgason, A., Sigurðardóttir, S., Gulcher, J.R., Ward, R. & Stefánsson, K. (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* **66**, 999–1016.

Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E. & 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).

Herrnstadt, C., Preston, G. & Howell, N. (2003) Errors, phantom and otherwise, in human mtDNA sequences. *Am J Hum Genet* **72**, 1585–86.

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

Imaizumi, K., Parsons, T.J., Yoshino, M. & Holland, M.M. (2002) A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *Int J Legal Med* **116**, 68–73.

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

 © University College London 2003

Jun, A.S., Brown, M.D. & Wallace, D.C. (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.

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

Kolman, C., Sambuughin, N. & Bermingham, E. (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* **142**, 1321–1334.

Koyama, H., Iwasa, M., Maeno, Y., Tsuchimochi, T., Isobe, I., Seko-Nakamura, Y., Monma-Ohtaki, J., Matsumoto, T., Ogawa, S., Sato, B. & Nagao, M. (2002) Mitochondrial sequence haplotype in the Japanese population. *Forensic Sci Int* **125**, 93–96.

Lee, S.D., Lee, Y.S. & Lee, J.B. (2002) Polymorphism in the mitochondrial cytochrome B gene in Koreans: an additional marker for individual identification. *Int J Legal Med* **116**, 74–78.

Lee, S.D., Shin, C.H., Kim, K.B., Lee, Y.S. & Lee, J.B. (1997) Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Sci Int* **87**, 99–116.

Lell, J.T., Sukernik, R.I. & Wallace, D.C. (2002) Reply to Tarazona-Santos and Santos. *Am J Hum Genet* **70**, 1380–1381.

Lum, J.K., Cann, R.L., Martinson, J.J. & Jorde, L.B. (1998) Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. *Am J Hum Genet* **63**, 613–624.

Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C. & Cabrera, V.C. (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* **2**, 13.

Malhi, R.S., Eshleman, J.A., Greenberg, J.A., Weiss, D.A., Schultz Shook, B.A., Kaestle, F.A., Lorenz, J.G., Kemp, B.M., Johnson, J.R. & Smith, D.G. (2002) The structure of diversity within New World mitochondrial DNA haplogroups: implications for the prehistory of North America. *Am J Hum Genet* **70**, 905–919; **71**, 215 (erratum).

Mishmar, D., Ruiz–Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D., Sukernik, R.I., Olckers, A. & Wallace, D.C. (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* **100**, 171–176.

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

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

Pfeiffer, H., Steighner, R., Fisher, R., Mörnstad, H., Yoon, C.-L. & Holland, M.M. (1998) Mitochondrial DNA extraction and typing from isolated dentin—experimental evaluation in a Korean population. *Int J Legal Med* **111**, 309–313.

Rando, J.C., Cabrera, V.M., Larruga, J.M., Hernández, M., González, A.M., Pinto, F. & Bandelt, H.-J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* **63**, 413–428.

Richards, M. & Macaulay, V. (2001) The mitochondrial gene tree comes of age. *Am J Hum Genet* **68**, 1315–1320.

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

Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B. & Wallace, D.C. (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea–Bering Sea region during the Neolithic. *Am J Phys Anthropol* **108**, 1–39.

Scozzari, R., Cruciani, F., Santolamazza, P., Sellitto, D., Cole, D.E.C., Rubin, L.A., Labuda, D., Marini, E., Succa, V., Vona, G. & Torroni, A. (1997) mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. *Am J Hum Genet* **60**, 241–244.

Silva, Jr., W.A., Bonatto, S.L., Holanda, A.J., Ribeiro-Dos-Santos, A.K., Paixão, B.M., Goldman, G.H., Abe-Sandes, K., Rodriguez-Delfin, L., Barbosa, M., Paçó-Larson, M.L., Petzl-Erler, M.L., Valente, V., Santos, S.E.B. & Zago, M.A. (2002) Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. *Am J Hum Genet* **71**, 187–192.

Silva, Jr., W.A., Bonatto, S.L., Holanda, A.J., Ribeiro-Dos-Santos, A.K., Paixão, B.M., Goldman, G.H., Abe-Sandes, K., Rodriguez-Delfin, L., Barbosa, M., Paçó-Larson, M.L., Petzl-Erler, M.L., Valente, V., Santos, S.E.B. & Zago, M.A. (2003) Correction: mitochondrial DNA variation in Amerindians. *Am J Hum Genet* **72**, 1346–1348.

Smith, D.G., Malhi, R.S., Eshleman, J., Lorenz, J.G. & Kaestle, F.A. (1999) Distribution of mtDNA haplogroup X among Native North Americans. *Am J Phys Anthropol* **110**, 271–284.

Snäll, N., Savontaus, M.-L., Kares, S., Lee, M.S., Cho, E.K., Rinne, J.O. & Huoponen, K. (2002) A rare mitochondrial DNA haplotype observed in Koreans. *Hum Biol* **74**, 253–262.

Tanaka, M., Fuku, N., Takeyasu, T., Guo, L.-J., Hirose, R., Kurata, M., Borgeld, H.-J.W., Yamada, Y., Maruyama, W., Arai, Y., Hirose, N., Oshida, Y., Sato, Y., Hattori, N.,

© University College London 2003

Mizuno, Y., Iwata, S. & Yagi, K. (2002) Golden mean to longevity: rareness of mitochondrial cytochrome *b* variants in centenarians but not in patients with Parkinson's disease. *J Neurosci Res* **70**, 347–355.

Tarazona-Santos, E. & Santos, F.R. (2002) The peopling of the Americas: a second major migration? *Am J Hum Genet* **70**, 1377–1380.

Torroni, A., Rengo, C., Guida, V., Cruciani, F., Sellitto, D., Coppa, A., Luna-Calderon, F., 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.

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M. & Wallace, D.C. (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* **53**, 563–590.

Tsai, L.C., Lin, C.Y., Lee, J.C.I., Chang, J.G., Linacre, A. & Goodwin, W. (2001) Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Sci Int* **119**, 239–247.

Vasil'ev, S.A. (2000) The Siberian mosaic: Upper Palaeolithic adaptions and change before the Last Glacial Maximum. In: Roebroeks, W., Mussi, M., Svoboda, J., Fennema, K. (eds) *Hunters of the golden age*. University of Leiden, Leiden, pp 173–195.

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

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., Lü, X.-M., Luo, H.-R., Li, W.-H. & Zhang, Y.-P. (2000) 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., 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. & 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.

Received: 29 November 2002
Accepted: 8 April 2003

 © University College London 2003