Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 141 (2004) 1–6
Forensic Science International
www.elsevier.com/locate/forsciint

A call for mtDNA data quality control in forensic science

Yong-Gang Yao[a,1], Claudio M. Bravi[b], Hans-Jürgen Bandelt[c,*]

[a]*Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China*

[b]*Centro de Investigaciones en Genetica Basica y Aplicada, School of Veterinary Sciences, National University of La Plata, La Plata, Argentina*

[c]*Fachbereich Mathematik, Universität Hamburg, Bundesstr. 55, 20146 Hamburg, Germany*

Received 3 October 2003; received in revised form 14 November 2003; accepted 10 December 2003

Abstract

There is increasing evidence that many of the mitochondrial DNA (mtDNA) databases published in the fields of forensic science and molecular anthropology are flawed. An a posteriori phylogenetic analysis of the sequences could help to eliminate most of the errors and thus greatly improve data quality. However, previously published caveats and recommendations along these lines were not yet picked up by all researchers. Here we call for stringent quality control of mtDNA data by haplogroup-directed database comparisons. We take some problematic databases of East Asian mtDNAs, published in the *Journal of Forensic Sciences* and *Forensic Science International*, as examples to demonstrate the process of pinpointing obvious errors. Our results show that data sets are not only notoriously plagued by base shifts and artificial recombination but also by lab-specific phantom mutations, especially in the second hypervariable region (HVR-II).

Keywords: mtDNA typing; Hypervariable regions; Phylogeny; Haplogroup; Quality control

1. Introduction

DNA typing is the most important advance in forensic science and is very useful in criminal prosecutions. Among the DNA markers (loci) employed in the field, mitochondrial DNA (mtDNA) is chosen for its specific characteristics, such as maternal inheritance, absence of recombination, high divergence rate, and high copy number per cell. MtDNA databases of various populations offer valuable information for estimating a chance matching probability when a forensic stain and a suspect share a sequence as well as for inferring (sub)continental origin of an mtDNA lineage. Unfortunately, many published mtDNA data are not sufficiently reliable [1–3], thus strongly limiting their forensic use.

There are five major and common types of errors, namely, base shifts, reference bias, phantom mutations, base mis-scoring, and artificial recombination, observed in published mtDNA control region sequence data [1]. These errors can be detected by phylogenetic analysis and comparison with closely related sequences from other databases. This approach has been used to pinpoint errors in mtDNA control region [2], coding region [4–6] as well as in ancient DNA data [7], which led to a number of caveats for data generation and compilation. However, many researchers do not perform sufficiently stringent a posteriori quality control for their data. The recent report of 105 Chinese Han mtDNA control region sequences [8] constitutes a case in this regard. In what follows, we took this data set as well as the data from Tsai et al. [9] and Koyama et al. [10] for exemplifying the phylogenetic strategy of pinpointing potential errors in mtDNA sequences.

2. Database comparisons

2.1. Haplogroups

The East Asian mtDNA phylogeny is now quite well worked out [11–13], and this reference system has recently

* Corresponding author. Fax: +49-40-42838-5190.
E-mail address: bandelt@math.uni-hamburg.de (H.-J. Bandelt).
[1] Present address: The Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

0379-0738/$ – see front matter
doi:10.1016/j.forsciint.2003.12.004

entered the forensic field as well [14]. The basal mutations on each branch of the phylogeny define an mtDNA haplogroup, with a more or less restricted geographic distribution. Normally, the haplogroup-specific mutation(s) of one haplogroup do not co-occur with the haplogroup-specific mutation(s) of another haplogroup, except for singular recurrent events. If, in a sequence composed of several independently generated mtDNA fragments, chunks of mutations from one segment signifying one haplogroup co-occur with chunks of mutations from another segment that define another haplogroup, then one would see clear evidence for artificial recombination between two generated mtDNA lineages. By comparison with complete sequence data and a near-matching analysis, one can quite reliably assign partial control-region sequences to the haplogroups they belong to. The reference to haplogroups, coded as alternating strings of letters and numbers (see Kong et al. [13] for the most recent East Asian mtDNA nomenclature), thus constitutes a convenient shortcut of a broad phylogenetic analysis (covering the worldwide database). Consequently, if needed, explicit estimation of phylogenetic trees can be confined to the (sub)haplogroup level.

2.2. Fast transitions

It is important in this context to be aware of the mutational rate spectrum in the two hypervariable regions, HVR-I and HVR-II, of the mtDNA control region [2,15,16]. Note that an early attempt by Meyer et al. [17] was problematic for both methodology and database [2,18]. For instance, site 247 signposts this failure since this site appears to be a top hotspot in Figure 2 in Meyer et al. [17], whereas it is found mutated in only three sequences (yielding two distinct haplotypes from haplogroup D5 and N9a, respectively) in ~1500 HVR-II sequences from East Asia [6,9–12,14,19–24]. Rather than distinguishing between fast- and slow-mutating sites, we discriminate between fast transitions and slower mutations. The latter thus comprise the remaining transitions as well as all transversions, deletions, and insertions, except for length polymorphisms of C-stretches and flanking short A-stretches that are generally disregarded for phylogenetic analysis. For HVR-II, we regard the transitions at the following ten sites as fast: 146, 150, 151, 152, 185, 189, 195, 199, 204, and 207. These 10 sites occur among the top 12 in the HVR-II hit lists of both Allard et al. [15] and Malyarchuk et al. [16], although these authors did not distinguish the types of mutations hitting a site.

2.3. Indels and transversions

Deletions and insertions outside the C-stretches in regions 303–309 and 311–315 are very rare in HVR-II: Yao et al. [25] found only one more instance of 249d in a haplogroup A lineage besides the two prominent occurrences of 249d in the East Asian mtDNA phylogeny that are characteristic of haplogroup CZ and haplogroup F3, respectively [13], and furthermore found three independent insertions of C at site 44 in more than 1100 mtDNA samples from China. In 373 Japanese HVR-II sequences [14,21], the following additional indels were detected: the double-deletion 290d-291d (characteristic of the Asian/Native American haplogroup C1; see [5]), the haplogroup-specific insertion 191.1A (defining a D4 subhaplogroup with ancestral sequence 16245-16362-73-191.1A-194-199-207-263 [14,20,21,23,26,27]), and yet another occurrence of 191.1A in a haplogroup G lineage and a private event 56d-71.1G in a haplogroup B4b1 lineage. Real (that is, not artificial) transversions may be even rarer, as judged from those Japanese sequences: only 58A is reported. The single shared HVR-II transversion in East Asia exhibited so far is the T to A transversion at site 146, which defines a subhaplogroup of haplogroup M7c that lacks the HVR-I motif 16223 [9–13].

3. Results and discussion

3.1. Haplogrouping

Tables 1 and 2 of Rao et al. [8] compile 105 Chinese Han mtDNA lineages based on HVR-I and HVR-II typing. More than 95% of these mtDNA lineages can actually be allocated to specific mtDNA haplogroups according to their mutation motifs (Table 1). For instance, since the sequence 16185-16223-16260-16298-73-152-249d-263 (transitions and indels relative to the revised Cambridge reference sequence [28]; length polymorphisms of C-stretches disregarded) is inferred to be the ancestral sequence of haplogroup Z, sample nos. 61, 78, and 94 should be members of this haplogroup. Note that the sequence 16223-16298-73-249d-263 is ancestral to the super-haplogroup CZ (that embraces Z and its sister haplogroup C), whereas the seemingly similar sequence 16298-16362-73-249d-263 is ancestral to a phylogenetically distant haplogroup, now referred to as F3. Haplogroup F3 comprises two subhaplogroups, F3a (formerly called R9a [11,12]) with the additional transition at sites 16355, and F3b with the transversion 16220C. For the latter subhaplogroup, only control-region information is available so far, and it has been found in seven different (South)-East Asian datasets [29–35]. Here we can assign sample nos. 17, 47, and 70 to F3b and sample no. 84 to F3a. Note that sample no. 84 also bears a transition at the (highly variable) site 16260 that would otherwise be expected in haplogroup Z sequences, but inasmuch as it lacks the 16185-16223-152 mutations, Z membership is very unlikely. Therefore, despite the fact that the partial mutation motif 16298-249d is shared by haplogroups CZ and F3 as a parallelism, confusion is unlikely to arise because of additional characteristic mutations for the particular haplogroups C, Z, F3a, and F3b. Comparing the 16298 transition as a single ''allele'' across different mtDNA pools, however, as done in Table 3 of Nishimaki et al. [36], is misleading because not only in East Asia this mutation hints at three

Table 1
Haplogroup assignment of the samples from Rao et al. [8]

Haplogroup	Ancestral sequence (HVR-I and HVR-II)[a]	Sample number[b]
M/N	16223-73-263	15, ***22***, 90
M7b	16223-16297-73-150-199-263	11, 14, 27, 35, ***80***, 91, 105
M7c	16223-73-146-199-263	77
M7c1	16223-16295-73-146-199-263	88
M8a	16184-16223-16298-16319-73-263	74, ***98***, 104
C	16223-16298-16327-73-249d-263	***13***, ***32***, 66
Z	16185-16223-16260-16298-73-152-249d-263	61, ***78***, ***94***
M9a	16223-16234-16316-(16362)-73-153-263	***100***
M10	16223-16311-73-263	1
M11	16223-73-215-263-318-326	***34***
G1a	16223-16325-16362-73-150-263	31
G2a	16223-16227-16278-16362-73-263	***59***
D4	16223-16362-73-263	16, 28, ***44***, 49, 53, 76, 79, 86, 89, 95
D5	16189-16223-16362-73-150-263	***9***, 30, 42, 43, 45, 96
D5a	16189-16223-16266-16362-73-150-263	2, 39, 65, 83
A	16223-16290-16319-73-152-235-263	***7***, 20, 25, ***26***, 41, 57, 73
N9a	16223-16257A-16261-73-150-263	69, 81
Y	16126-16231-73-263	***36***, 63
Y1	16126-16231-16266-73-(146)-263	***62***
B	16189-73-263	10, 38
B4	16189-16217-73-263	8, 21, 33, 46, 52, 54, ***67***, ***71***, ***72***, 97, 101
B4b1	16136-16189-16217-73-263	6, ***48***, 55
B5a	16140-16189-16266R-73-263	***5***, 29
B5b	16140-16189-16243-73-263	64, 75, 82
F	16304-73-249d-263	3, 12, ***18***, ***50***, ***56***, 58, 92, 103
F1a	16129-16172-16304-73-249d-263	19, 23, 24, 37, ***60***, 87
F1b	16189-16304-73-249d-263	40, 85, 93, ***102***
F2a	(16291)-16304-73-249d-263	***4***, 68, 99
F3a	16298-16355-16362-73-249d-263	84
F3b	16220C-16298-16362-73-249d-263	17, 47, ***70***
R11	16189-16311-73-185-189-263	51

[a] Length polymorphisms of C-stretches are disregarded; suffix C, A or R indicates transversion and d indicates deletion; uncertain mutations are given in parentheses.

[b] Samples with potential errors, such as artificial recombination (crossover between different lineages), mutation oversight (reference bias), base shift (typically by one position), and phantom mutation are highlighted in bold italic.

very different haplogroups (M7b2, CZ, and F3), but also in West Eurasia this mutation typically points to yet another haplogroup, pre-V [37]. The proper phylogenetic position of an mtDNA lineage often requires more than just one characteristic control-region mutation.

3.2. Base shifts and mistyping

Note that site 95 with G, site 205 with T, site 206 with G, site 16175 with C, and site 16318 with G, as asserted in the two tables of Rao et al. [8], do not exist in the (revised) Cambridge reference sequence [28]. By comparison with related sequences from East Asia, we can infer that the recorded sites 95 and 205 likely refer to sites 94 and 204 instead, whereas sites 206 and 16318 refer to 207 and 16319.

In each table, Rao et al. [8] intended to highlight sequences that bear a deletion in the corresponding hypervariable segment by adding a suffix ‘‘d’’ to the sample number. While for Table 1 this is consistently performed, there are inconsistencies for Table 2: sample nos. 13, 32, and 103 harbor a deletion but lack the ‘‘d’’ suffix. On the other hand, sample nos. 70d and 78d lack any deletions in their HVR-II sequences. Incidentally, no. 70d and 78d belong to haplogroups F3b and Z, respectively, in which otherwise a deletion at site 249 (249d) is always present [13]. In the worldwide database, a deletion at site 249 seems to be a rare event, essentially confined to haplogroups CZ and F. The deletion at this site was poorly reported by Rao et al. [8] anyway: sample nos. 4, 18, 60, very likely belonging to haplogroup F, all lack 249d. Remarkably, sample nos. 59 and 62, which clearly belong to haplogroups G2a and Y1, respectively, show a deletion at site 249 in their second hypervariable segments. The most plausible explanation for this oddity is artificial recombination of the two segments in these instances!

3.3. Phantom transversions

Phantom mutations of the kind discussed by Bandelt et al. [1,2] and Herrnstadt et al. [38] seem to occur as well in the Rao et al. [8] data. The variation at sites 311 and 317 is particularly alarming. These two sites are invariable in >1000 mtDNA samples from East Asia that seem to be not affected by phantom mutations [6,11,12,14,19,21, 22,24], but in the data of Rao et al. [8] we are seeing four independent mutational events at site 311 in B4*, B4a, B4b, and F1b lineages as well as two independent transversions at site 317, one jointly with the insertion 310 + T and the transition 320 (in sample no. 22), and another one jointly with 311 mutated (in sample no. 48). Interestingly, the latter sample has a question mark after the sample name in their Table 2. A possible explanation is that the HVR-II sequence of this sample was not of good quality and the authors might have used a question mark for reminder.

The HVR-II sequences published by Koyama et al. [10] contain some suspicious transversions and a yet unobserved deletion (at site 85). Whereas a transversion 146A has been reported before in haplogroup M7c, the transversion 114G seems to be novel and yet it occurs here in two lineages from different haplogroups, and also 190G has been not observed so far. The first sequence listed in Table 1 of Koyama et al. [10] even shows two transversions, namely, 189C and 190A. This lineage, however, bears additional mutations at 16245-16362-73-199-207-246-263 and thus qualifies as a member of the Japanese-specific D4 subhaplogroup with the salient insertion of an A at site 191. This insertion was probably miscoded by shifting back the preceding CA by one position and thereby compressing the two A nucleotides at sites 188 and 189 to one. In total, we then count five lineages with private transversions or indels in the 50 HVR-II sequences of Koyama et al. [10], but find only a single lineage with private transversions/indels among 373 other samples from Japan [14,21].

The 155 mtDNAs reported by Tsai et al. [9] also contain several unusual transversions to G not observed in those other East Asian mtDNA data sets, namely, at sites 125, 162, 170, 198, 224, and 253. Three of these transversions even occur in one lineage (from haplogroup F3a). Transversions at site 162 are seen in four mtDNAs belonging to different haplogroups (F1ac, F3a, M10, and M7b1). In three instances the 162G transversion occurs together with a transition at site 253 and in one instance with the 170G transversion. Transitions at site 253 re-occur in four additional lineages from haplogroups D4, M9a, A, and N9a, although this site is not polymorphic in the other East Asian mtDNA data sets. The repeated occurrence of tandem mutations 162G-253 is reminiscent of the recurrent tandem mutations 16239G-16242 or 16242-16248 in the HVR-I data of Seo et al. [20] (see [2]). The explanation for this phenomenon is that a small portion (<10%) of the data set was prone to phantom mutations, which for some biochemical reasons would hit a very limited number of sites with a high probability.

4. Conclusion

In view of an increasing number of forensic labs that start to work on mtDNA, we feel it urgent to reiterate the need for data quality control. We entreat new mtDNA researchers to check for the sources of error summarized in Bandelt et al. [1]. Extreme caution should be exercised at all stages of data collection and proof-reading processes. Reliance on one strand alone can easily lead to phantom mutations [2]. The recommendation is thus clear: ''Both strands of the amplified product must be sequenced to reduce ambiguities in sequence determination'' [39]. Reference to a worked-out (continental) phylogeny and comparison with published data sets of similar geographic/ethnic origins are also indispensable for identifying potential errors [1,11,24]. It is somewhat unsatisfactory that such a posteriori analysis is not (yet) considered to be part of a routine quality assurance/quality control [39,40]. One should also bear in mind that a final check of the data matrix before submission as well as at proof-reading stage will help to eliminate trivial editing errors.

For some reasons, sequence data published in forensic science journals are not required to be deposited in public databanks, such as GenBank, EMBL, or DDBJ. Some authors, enforced by a requirement of Lincoln and Carracedo [41], claimed that ''the sequence data are available from the authors on request'', but actually they do not always reply to the request for the data. Or, the e-mail address of the corresponding author does not seem to exist. All these circumstances definitely hamper the evaluation of the original data obtained by other labs. The easiest way to settle the problem of original data access is to deposit sequences into GenBank or to put them on the Web for general access. The ''D-Loop-BASE'' project initiated by Wittig et al. [42] is problematic in this respect, as ''individual sequences will not be published'' [43].

In summary, we still have every reason to question the quality of the mtDNA data published in forensic science. It does not seem that the alarming news of 2001 have led to a noticeable improvement of data quality so far. For instance, we expect that more than 20% of the 105 mtDNA sequences published by Rao et al. [8] harbor obvious errors, such as trivial editing errors (resulting in base shifts) and artificial recombination (crossover between different lineages). It is axiomatic that such data are rather useless for forensic purposes. Therefore, researchers who provide data, peer reviewers, and journal editors are urged to take the possibility of serious errors into consideration and to subject the data a thorough a posteriori analysis.

References

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

[2] H.-J. Bandelt, L. Quintana-Murci, A. Salas, V. Macaulay, The fingerprint of phantom mutations in mtDNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.
[3] P. Forster, To err is human, Ann. Hum. Genet. 67 (2003) 2–4.
[4] Y.-G. Yao, V. Macaulay, T. Kivisild, Y.-P. Zhang, H.-J. Bandelt, To trust or not to trust an idiosyncratic mitochondrial data set, Am. J. Hum. Genet. 72 (2003) 1341–1346.
[5] H.-J. Bandelt, C. Herrnstadt, Y.-G. Yao, Q.-P. Kong, T. Kivisild, C. Rengo, R. Scozzari, M. Richards, R. Villems, V. Macaulay, N. Howell, A. Torroni, Y.-P. Zhang, Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats, Ann. Hum. Genet. 67 (2003) 512–524.
[6] Q.-P. Kong, Y.-G. Yao, M. Liu, S.-P. Shen, C. Chen, C.-L. Zhu, M.G. Palanichamy, Y.-P. Zhang, Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China, Hum. Genet. 113 (2003) 391–405.
[7] Y.-G. Yao, Y.-P. Zhang, Pitfalls in the analysis of ancient human mtDNA, Chinese Sci. Bull. 48 (2003) 826–830.
[8] L. Rao, M.Y. Wu, W.B. Liang, L. Zhang, Sequence polymorphisms of the mitochondrial DNA control region in 105 Chinese Han population, J. Forensic Sci. 48 (2003) 891–895.
[9] L.C. Tsai, C.Y. Lin, J.C.I. Lee, J.G. Chang, A. Linacre, W. Goodwin, Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population, Forensic Sci. Int. 119 (2001) 239–247.
[10] H. Koyama, M. Iwasa, Y. Maeno, T. Tsuchimochi, I. Isobe, Y. Seko-Nakamura, J. Monma-Ohtaki, T. Matsumoto, S. Ogawa, B. Sato, M. Nagao, Mitochondrial sequence haplotype in the Japanese population, Forensic Sci. Int. 125 (2002) 93–96.
[11] Y.-G. Yao, Q.-P. Kong, H.-J. Bandelt, T. Kivisild, Y.-P. Zhang, Phylogeographic differentiation of mitochondrial DNA in Han Chinese, Am. J. Hum. Genet. 70 (2002) 635–651.
[12] T. Kivisild, H.-V. Tolk, J. Parik, Y. Wang, S.S. Papiha, H.-J. Bandelt, R. Villems, The emerging limbs and twigs of the East Asian mtDNA tree, Mol. Biol. Evol. 19 (2002) 1737–1751 and 20 (2003) 162 (erratum).
[13] Q.-P. Kong, Y.-G. Yao, C. Sun, H.-J. Bandelt, C.-L. Zhu, Y.-P. Zhang, Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences, Am. J. Hum. Genet. 73 (2003) 671–676.
[14] S. Maruyama, K. Minaguchi, N. Saitou, Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population, Int. J. Legal Med. 117 (2003) 218–225.
[15] M.W. Allard, K. Miller, M. Wilson, K. Monson, B. Budowle, Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences, J. Forensic Sci. 47 (2002) 1215–1223.
[16] B.A. Malyarchuk, I.B. Rogozin, V.B. Berikov, M.V. Derenko, Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region, Hum. Genet. 111 (2002) 46–53.
[17] S. Meyer, G. Weiss, A. von Haeseler, Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA, Genetics 152 (1999) 1103–1110.
[18] E. Árnason, Genetic heterogeneity of Icelanders, Ann. Hum. Genet. 67 (2003) 5–16.
[19] H. Pfeiffer, R. Steighner, R. Fisher, H. Mörnstad, C.-L. Yoon, M.M. Holland, Mitochondrial DNA extraction and typing from isolated dentin—experimental evaluation in a Korean population, Int. J. Legal Med. 111 (1998) 309–313.
[20] Y. Seo, B. Stradmann-Bellinghausen, C. Rittner, K. Takahama, P.M. Schneider, Sequence polymorphism of mitochondrial DNA control region in Japanese, Forensic Sci. Int. 97 (1998) 155–164.
[21] K. Imaizumi, T.J. Parsons, M. Yoshino, M.M. Holland, A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals, Int. J. Legal Med. 116 (2002) 68–73.
[22] S.D. Lee, Y.S. Lee, J.B. Lee, Polymorphism in the mitochondrial cytochrome B gene in Koreans: an additional marker for individual identification, Int. J. Legal Med. 116 (2002) 74–78.
[23] S.D. Lee, C.H. Shin, K.B. Kim, Y.S. Lee, J.B. Lee, Sequence variation of mitochondrial DNA control region in Koreans, Forensic Sci. Int. 87 (1997) 99–116.
[24] Y.-G. Yao, Q.-P. Kong, X.-Y. Man, H.-J. Bandelt, Y.-P. Zhang, Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA, Mol. Biol. Evol. 20 (2003) 214–219.
[25] Y.-G. Yao, Q.-P. Kong, C. Sun, Y.-P. Zhang, Can the occurrence of rare insertion/deletion polymorphisms in human mtDNA be verified from phylogeny? Chinese Sci. Bull. 48 (2003) 663–667.
[26] T. Ozawa, M. Tanaka, H. Ino, K. Ohno, T. Sano, Y. Wada, M. Yoneda, Y. Tanno, T. Miyatake, T. Tanaka, S. Itoyama, S. Ikebe, N. Hattori, Y. Mizuno, Distinct clustering of point mutations in mitochondrial DNA among patients with mitochondrial encephalomyopathies and with Parkinson's disease, Biochem. Biophys. Res. Commun. 176 (1991) 938–946.
[27] H. Oota, N. Saitou, T. Matsushita, S. Ueda, A genetic study of 2000-year-old human remains from Japan using mitochondrial DNA sequences, Am. J. Phys. Anthropol. 98 (1995) 133–145.
[28] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.
[29] S. Horai, K. Hayasaka, Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA, Am. J. Hum. Genet. 46 (1990) 828–842.
[30] J.K. Lum, R.L. Cann, MtDNA lineage analyses: origins and migrations of Micronesians and Polynesians, Am. J. Phys. Anthropol. 113 (2000) 151–168.
[31] T. Melton, S. Clifford, J. Martinson, M. Batzer, M. Stoneking, Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes, Am. J. Hum. Genet. 63 (1998) 1807–1823.
[32] B. Sykes, A. Leiboff, J. Low-Beer, S. Tetzner, M. Richards, The origins of the Polynesians: an interpretation from mitochondrial lineage analysis, Am. J. Hum. Genet. 57 (1995) 1463–1475.
[33] A. Tajima, C.-S. Sun, I.-H. Pan, T. Ishida, N. Saitou, S. Horai, Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese, Hum. Genet. 113 (2003) 24–33.
[34] T. Yoshii, E. Takeda, K. Akiyama, I. Ishiyama, Sequence polymorphism of mitochondrial DNA and its forensic

application (in Japanese), Nippon Hoigaku Zasshi 49 (1995) 242–250.

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

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

[37] A. Torroni, H.-J. Bandelt, V. Macaulay, M. Richards, F. Cruciani, C. Rengo, V. Martinez-Cabrera, R. Villems, T. Kivisild, E. Metspalu, J. Parik, H.-V. Tolk, K. Tambets, P. Forster, B. Karger, P. Francalacci, P. Rudan, B. Janicijevic, O. Rickards, M.-L. Savontaus, K. Huoponen, V. Laitinen, S. Koivumäki, B. Sykes, E. Hickey, A. Novelletto, P. Moral, D. Sellitto, A. Coppa, N. Al-Zaheri, A.S. Santachiara-Benerecetti, O. Semino, R. Scozzari, A signal, from human mtDNA, of post-glacial recolonization in Europe, Am. J. Hum. Genet. 69 (2001) 844–852.

[38] C. Herrnstadt, G. Preston, N. Howell, Errors, phantom and otherwise, in human mtDNA sequences, Am. J. Hum. Genet. 72 (2003) 1585–1586.

[39] SWGDAM, Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation, Forensic Sci. Commun. 5 (2003) [Online]. Available: www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm.

[40] B. Budowle, M.W. Allard, M.R. Wilson, R. Chakraborty, Forensics and mitochondrial DNA: applications, debates, and foundations, Annu. Rev. Genomics Hum. Genet. 4 (2003) 119–141.

[41] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, Forensic Sci. Int. 110 (2000) 3–5.

[42] H. Wittig, C. Augustin, A. Baasner, U. Bulnheim, N. Dimo-Simonin, J. Edelmann, S. Hering, S. Jung, S. Lutz, M. Michael, W. Parson, M. Poetsch, P.M. Schneider, G. Weichhold, D. Krause, Mitochondrial DNA in the Central European population: human identification with the help of the forensic mtDNA D-loop-base database, Forensic Sci. Int. 113 (2000) 113–118.

[43] H. Wittig, M. Koecke, K.-U. Sattler, D. Krause, D-Loop-BASE is online now: Central European database of mitochondrial DNA, Progress in Forensic Genetics 9 (International Congress Series 1239) (2003) 505–509.