

# Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations

Qing-Peng Kong<sup>1,2,3,\*</sup>, Hans-Jürgen Bandelt<sup>4</sup>, Chang Sun<sup>1,2,3</sup>, Yong-Gang Yao<sup>1</sup>, Antonio Salas<sup>5</sup>, Alessandro Achilli<sup>6</sup>, Cheng-Ye Wang<sup>1,2,3</sup>, Li Zhong<sup>2</sup>, Chun-Ling Zhu<sup>1</sup>, Shi-Fang Wu<sup>1</sup>, Antonio Torroni<sup>6</sup> and Ya-Ping Zhang<sup>1,2,\*</sup>

<sup>1</sup>Laboratory of Cellular and Molecular Evolution, and Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China, <sup>2</sup>Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China, <sup>3</sup>Graduate School of the Chinese Academy of Sciences, Beijing 100039, China, <sup>4</sup>Department of Mathematics, University of Hamburg, 20146 Hamburg, Germany, <sup>5</sup>Unidad de Genética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, 15782, and Centro Nacional de Genotipado (CeGen), Hospital Clínico Universitario, 15706 Galicia, Spain and <sup>6</sup>Dipartimento di Genetica e Microbiologia, Università di Pavia, 27100 Pavia, Italy

Received April 7, 2006; Revised May 8, 2006; Accepted May 14, 2006

GenBank accession nos<sup>†</sup>

**Knowledge about the world phylogeny of human mitochondrial DNA (mtDNA) is essential not only for evaluating the pathogenic role of specific mtDNA mutations but also for performing reliable association studies between mtDNA haplogroups and complex disorders. In the past few years, the main features of the East Asian portion of the mtDNA phylogeny have been determined on the basis of complete sequencing efforts, but representatives of several basal lineages were still lacking. Moreover, some recently published complete mtDNA sequences did apparently not fit into the known phylogenetic tree and conflicted with the established nomenclature. To refine the East Asian mtDNA tree and resolve data conflicts, we first completely sequenced 20 carefully selected mtDNAs—likely representatives of novel sub-haplogroups—and then, in order to distinguish diagnostic mutations of novel haplogroups from private variants, we applied a ‘motif-search’ procedure to a large sample collection. The novel information was incorporated into an updated East Asian mtDNA tree encompassing more than 1000 (near-) complete mtDNA sequences. A reassessment of the mtDNA data from a series of disease studies testified to the usefulness of such a refined mtDNA tree in evaluating the pathogenicity of mtDNA mutations. In particular, the claimed pathogenic role of mutations G3316A, T3394C, A4833G and G15497A appears to be most questionable as those initial claims were derived from anecdotal findings rather than e.g. appropriate association studies. Following a guideline based on the phylogenetic knowledge as proposed here could help avoiding similar problems in the future.**

## INTRODUCTION

Sequencing of entire molecules represents the ultimate approach to acquire information from the maternally inherited mitochondrial genome. With this approach in the past few

years, more and more data have been obtained for reconstructing the world mitochondrial DNA (mtDNA) phylogeny and for discerning the phylogenetic status of the (sub)continentally specific haplogroups (1–15). A well-reconstructed phylogeny helps to gain unique and valuable insights for elucidating

\*To whom correspondence should be addressed. Tel: +86 8715199030; Fax: +86 8715195430; Email: kongqp@yahoo.com.cn (Q.P.K.) or zhangyp1@263.net.cn (Y.P.Z.)

<sup>†</sup>GenBank accession nos for the mtDNA complete genome data: DQ272107–DQ272126.

human evolution and pioneer settlement patterns; for instance, complete mtDNA sequence data indicate that our ancestors adopted a *single* route [viz. the 'southern route'; (16)] to leave Africa and then migrated *rapidly* along the Asian coast (8,10,15,17,18). In mitochondrial disease studies, sequencing of the entire mitochondrial genome has become a routine method too. In this regard, the phylogenetic system of the world's mtDNAs has a great potential in assessing the pathogenicity of mtDNA mutations [e.g. see (19,20)].

In case-control studies, the attribution of a pathogenic or predisposing role to a specific variant is essentially based on the observation that the frequency of such a variant is significantly higher in the case subjects than in the normal controls; such results, however, could be easily affected by, e.g., population stratification (leading to false positive associations; type I error; 21–23). Even with well-matched controls, a significant association between the variant and the disease is not sufficient to prove its role as a disease-marker because of a number of reasons, including that some other mutation tightly linked with the nucleotide variant under study could be responsible for the significant association. Such a scenario is even more likely in a non-recombining molecule, such as mtDNA. Fortunately, the now emerging worldwide mtDNA phylogeny can provide a comprehensive understanding of where, when and how many times certain mtDNA mutations occurred, thus helping to evaluate the role of the variant under study.

To acquire a well-reconstructed mtDNA phylogeny, high-quality genome data are indispensable. Furthermore, to facilitate communication and avoid misunderstandings such a phylogeny has to be encoded by employing a consistent nomenclature. In this aspect, the recent inclusion in the East Asian mtDNA phylogeny of 672 new Japanese mtDNA genomes (24) should be treated with caution, as the genome data set bears a number of shortcomings, some of which are even reflected in the published summary trees (see Materials and Methods for further discussion); in addition, haplogroup designation partially conflicts with the already existing nomenclature and some haplogroup definitions seem to be problematic in a few instances. Meanwhile, with an increasing number of East Asian mtDNAs being analyzed, some novel (sub)haplogroups were recognized and awaited further characterization.

To address these issues, we reconstructed an updated East Asian mtDNA phylogeny based on the available complete sequences plus 20 newly sequenced ones. This resulting refined phylogeny was then employed to re-evaluate some previous disease studies that lacked a phylogenetic perspective. It appears that some mutations have repeatedly been the target of disease studies simply because they were not recorded in MITOMAP (<http://www.mitomap.org/>) although they clearly constitute polymorphisms in East Asia. With the phylogenetic bookkeeping of mutations at hand, such cases would demand a well-planned experimental design for extensive association studies (25). Moreover, the mtDNA phylogeny would also assist in the phylogenetic proof-reading of complete sequencing attempts (19) [see (26) for a response] as well as of packages of seemingly somatic mutations (20).

## RESULTS

### Reconciliation of previous nomenclature conflicts

A number of conflicts in the East Asian nomenclature have arisen among past studies, most of which were introduced because of the neglect of already existing haplogroup names or the delay of updating results in the light of most recent information. For example, the widely accepted haplogroup definitions M8a and M8 (5,27,28) were altered to S and CSZ, respectively, by Starikovskaya *et al.* (29). In the same study, the newly named haplogroup C2 is identical to the already defined haplogroup C4 (5). Table S1 (Supplementary Material) summarizes the conflicts as well as our corresponding reconciliation according to a chronological criterion (see Materials and Methods). Moreover, with our newly collected sequences, it was also possible to gain more insights into the phylogeny of haplogroup G, which was still poorly understood as for its entire variation.

As shown in Figure 1, the haplogroup G phylogeny, especially for G3 [defined by the control-region motif 16223-16274-16362 (30)], has been refined considerably. It is now evident that haplogroup G3 comprises two potential sister clades: G3a (further characterized by mutations at 143 and 15746) and G3b (recognizable by mutations at 13477, 14605 and 15927). Here we follow the convention that G3 is characterized by the 16274 transition (30), which, however, is now seen to be the sole mutation shared by both G3 branches. In retrospect, this definition was too optimistic in anticipating coding-region mutations accompanying the transition at 16274, which is a highly variable site. In haplogroup G2, a new branch appeared, named G2b, whose signature mutations (transitions at 3593, 4853, 8877, 11151, 16172 and a back mutation at 263) were identified by motif search (Supplementary Material, Table S2). The G sub-haplogroup 'G3', as identified by Tanaka *et al.* (24), is renamed here as G4 (Figure 1 or Supplementary Material, Table S1) and is determined by an array of mutations: the insertion 191 + A, transitions at 194, 4541, 5051, 5460, 6216, 7521, 7660, 9670, 11383 and 15940 and the 16114A transversion.

### Newly identified haplogroups

Our motif-search results (Supplementary Material, Table S2) generally support the definitions of some previously described (sub)haplogroups (5,24) for which only limited information was available, namely, G1a, M7b1, M7b2, F1a, F1c, F2a, R9b, B4a1, B5a and Y1. Our study also provides additional information for supporting several recently identified haplogroups. For instance, haplogroup M13 [here we adopted the nomenclature of (27) which was published online earlier than that of (24); cf. Supplementary Material, Table S1], originally identified by the transition motif 10411-16145-16188-16223 (27), is further characterized by eight mutations at sites 152, 3644, 5773, 6023, 6253, 6620, 10790 and 15924. The major sub-clade, M13a, of haplogroup M13 is defined by the coding-region transition at 13135. The lineage with the hypervariable segment (HVS) I motif 16086-16297-16324 that was suggested to belong to M7b and named M7b3 in Yao *et al.* (27), gets additional support from the coding region (Fig. 1). Similarly, the newly defined



haplogroup B6 (27) gets clear support from coding-region information (Fig. 2). A novel clade, which was characterized by mutations at 9968, 16261 and 16292, emerged in haplogroup B4 and is named B4g here. Furthermore, a new basal haplogroup, M12, could be identified. It shares the mutation 14569 with haplogroup G and was shown to be further defined by mutations at 4170, 5580, 12030, 12372, 14727, 15010, 16172, 16234 and 16290 (see Fig. 1 or Supplementary Material, Table S2).

### Updated East Asian mtDNA tree

The East Asian phylogenetic tree (Figs 1 and 2) was greatly refined on the basis of >1000 (near-)complete mtDNA sequences available so far: a large number of novel 'boughs' and 'twigs' (30) were recognized, thus largely broadening our understanding of the East Asian mtDNA phylogeny. In particular, a few artificial branches that were introduced in the past by relying predominantly on early (problematic) mtDNA data from disease studies (cf. 5,19,30) could now be eliminated. For example, the formerly defined haplogroup D4b1 turned out to be a sub-clade of what was then called haplogroup D4b2 (5), which was introduced because of the omission of variant 15440 in one haplogroup D4b sequence reported by Ozawa *et al.* (31). Subsequently, the definition of D4b was broadened as being based solely on the 8020 transition (24): it embraces sub-haplogroups D4b1 (identifiable by variants 10181, 15440, 15951 and 16319) and D4b2 (defined by 1382C, 8964 and 9824A) (Fig. 1).

### Pathogenic mutations in the mtDNA tree?

The bookkeeping of mtDNA mutations as organized in the current updated tree for East Asian mtDNAs permits the researcher to distinguish shared mutations of considerable age from private mutations of potentially much younger age. For instance, the substitution 3644 announced as 'uncharacterized' by Munakata *et al.* (32) is in fact characteristic of haplogroups M13 and D4h (Fig. 1) and thus is widely distributed across East Asia, although at a minor frequency. Then the mtDNAs of the other five cases also presented in Table 1 of Munakata *et al.* (32) can be tentatively allocated to East Asian haplogroups: the transition 13651 defines D4a1b; 12311 defines G2a3; 11084 defines M7a1a and 4705 defines F1b1a1.

Similarly, a number of earlier claims for pathogenic and somatic mtDNA mutations clearly need revision. For example, Chiu *et al.* (33) analyzed the entire mitochondrial genomes of 10 patients suffering from hydatidiform mole and one affected by choriocarcinoma, for which only the putatively 'new mtDNA polymorphisms' were reported. However, the vast majority of these mutations can be found in previously published complete mtDNA genome data. Specifically, mutations at 6392, 10535 and 13928 detected in patient M72 define haplogroup F2, those at 10031, 11061 and 13681 in patient M78 define haplogroup R11, and variants 5585, 5913, 6392 and 10320 in patient M89 are characteristic of haplogroup F3 (5). Moreover, Chiu *et al.* (33) claimed to have detected six somatic point mutations (*viz.* transitions at sites 153, 489, 1719, 16051, 16209 and 16519) in a single choriocarcinoma sample; however, these mutations have jointly

been observed in samples Hui45 [(27) and this study] and Bouyei50 (Supplementary Material, Table S2) which belong to haplogroup M9b, as revealed by our present complete sequencing efforts. Evidently, lab contamination or sample mix-up would easily explain the results, and the authors' claim that the high frequency of 'somatic' mutations is responsible for the gestational trophoblastic disease is therefore unsupported. This case would thus parallel a multitude of similar instances of sample mix-up, which appear to be predominant in the field of cancer research (20).

### DISCUSSION

In spite of having utilized more than 1000 complete genomes for phylogeny reconstruction, we cannot expect that our mtDNA tree covers absolutely all deep lineages in the East Asian matrilineal gene pool. In view of the enormous population size in East Asia (>1.4 billion), some basal additions and peripheral refinement will become necessary in the near future with more and more data available. This entails that not every novel combination of previously observed or hitherto unobserved mtDNA mutations found in some patient would necessarily point to mutations that were very recently acquired in the matriline. Only complete sequencing of the entire mtDNA and systematic searches for potentially related mtDNAs could then help to clarify the situation.

Knowledge about the mtDNA phylogeny is particularly necessary for disease studies when assessing the pathogenicity of certain mtDNA mutation(s). By elucidating the evolutionary relationships of certain mtDNA variants in a geographic context, we can (i) pinpoint the potential sequencing or documentation errors that could easily lead to spurious positive associations with diseases, (ii) anticipate the potential effect of population stratification, which would cause serious problems with case-control studies when ignored and (iii) distinguish the haplogroup-diagnostic variants from the private ones and determine whether the single mutation or a set of mutations or matrilineal backgrounds could play a role in the susceptibility to a disease. Unfortunately, clinical studies have hitherto shown a systematic lack of attention to the available mtDNA phylogenetic knowledge, which had some serious consequences (19,20,34-36).

### Effect of population stratification

The 3394 transition has a long record as a suspect for association with diseases, although its status as a pathogenic mutation is still 'unclear' according to MITOMAP. Johns *et al.* (37) had first brought the 3394 mutation into play in connection with patients of Leber's hereditary optic neuropathy (LHON). The mtDNAs of those patients harboring the 3394 change evidently belonged to haplogroup J1c1, for which this mutation is characteristic within haplogroup J1c (8). Quite a different scenario had been evoked in Japan, where this mutation was believed to contribute to hypertrophic cardiomyopathy (38). In East Asia, however, this mutation is a marker mainly for haplogroup M9a'b (besides D4g1a). In fact, the mtDNA of patient 6 described in Obayashi *et al.* (38) is a member of haplogroup M9a1. Later, this mutation was



assumed to contribute to the manifestation of non-insulin-dependent diabetes mellitus (NIDDM) (39–41) and was most recently observed in the case of carnitine palmitoyl-transferase II deficiency and ragged red fibers at muscle biopsy (42).

One of the arguments put forward by Hirai *et al.* (39) was that the mutation at 3394 was observed in 18/365 (4.9%) of the NIDDM patients but only in 3/225 (1.3%) of non-diabetic controls, which at face value could be regarded as significant. It is then interesting to compare these figures with the corresponding percentages in different mtDNA samples from Japan. In the data of Tanaka *et al.* (24), the frequency of the 3394 variant is 14/672 (2.1%), due to 11 M9a mtDNAs (contributing 1.6% alone) and 3 D4g1a mtDNAs. As judging from the HVS-I motifs, the haplogroup M9a contribution alone would, for example, be as high as 6/150 (4.0%) in Nishimaki *et al.* (43), take intermediate values 5/162 (3.1%) in the HVS-I&II data set of Imaizumi *et al.* (44) and 5/231 (2.2%) in the HVS-I data set of Tajima *et al.* (45), but merely 2/211 (0.9%) in the data from Maruyama *et al.* (46). Thus, by extrapolation from the Tanaka *et al.* sample (under the optimistic hypothesis that no other minor haplogroup would bear the 3394 mutation), we would expect the frequency of the 3394 transition to range freely between 1 and 5% in samples of size >150, depending on the sampling region, thus framing the frequencies found in both the patient and control cohorts by Hirai *et al.* (39). It is certainly not surprising that rather minor haplogroups, such as M9a, can show considerably different frequencies across a broad geographical area, which could turn out to be significantly different with large enough samples and thus positively testify to population substructure. Arbitrarily drawn control groups, constituting convenience samples, are never controlled for such hidden variables. As repeatedly demanded in the literature (21–23), control for population stratification is crucial in order to avoid spurious positive association in disease population-based studies, such as the classical case–control association design. The burden of proof for demonstrating that population structure did not influence the results rests on the researcher who has designed the study.

### Haplogroup-diagnostic mutation in disease study

An interesting case of a haplogroup-defining mutation that came up in connection with a number of diseases and mitochondrial disorders, mainly in patients from Japan [see (47) or MITOMAP for pertinent references], is represented by the 3316 transition, which defines haplogroup D4e1. This haplogroup encompasses the Native Northwest American haplogroup D2, which is, for instance, virtually fixed in the Aleuts of the Commander Islands (48). The 3316 mutation is, however, not restricted to haplogroup D4e1 but actually found in mtDNAs of all continental origins (cf. 3,49). Grazina *et al.* (47) reported the 3316 mutation along with the 3337 transition in a (Portuguese) patient with frontotemporal dementia (FTD). To our knowledge, there exists no record that would claim considerable frequency of FTD in the Aleuts. Therefore, the anecdotal finding of this mutation in a Portuguese patient cannot suggest potential disease

status of the 3316 variant, especially as another mutation (at 3337) was also observed in the patient.

To give another example, the 4833 transition was seen in connection with abnormal glucose tolerance (50), but it was not realized that this transition actually defines the entire haplogroup G and does not appear in any other part of the mtDNA phylogeny (among more than 2500 complete or coding mtDNA genomes scrutinized). Indeed, the 4833 change, recognizable by the gain of a *HaeII* site at 4830, originally defined the East Asian haplogroup G within the high-resolution system of restriction fragment length polymorphisms and has therefore been known for a long time.

The 15497 transition was claimed to be associated with obesity-related variables and lipid metabolism (51), but this mutation, in fact, defines haplogroup G1, which is mainly represented by sub-haplogroup G1a1 in Japan [with frequency 3.3%, according to Tanaka *et al.* (24)]. Another sub-haplogroup, G1b, is even the predominant haplogroup in Koryaks and Itel'men (52). The same mutation was also reported in a haplogroup H4 mtDNA from a 22-year-old male triathlete and thought to be connected with the development of exercise-induced paracrystalline inclusions (53). These authors erroneously asserted that this mutation would not occur among the 560 coding-region sequences provided by Herrstadt *et al.* (3,49). However, it does, viz. in sample no. 423, which is closely related to sample no. 11 of Achilli *et al.* (6). Both samples along with that described by Tarnopolsky *et al.* (53) belong to a common sub-haplogroup that is supported by transitions at sites 7581, 15497 and 15930 within haplogroup H4. In West Eurasia, the 15497 transition (recognizable by the gain of the *DdeI* restriction site at 15494) is also characteristic of a sub-haplogroup of haplogroup R1 (54). Tarnopolsky *et al.* (53) performed functional assays using cybrids (constructed by fusing the subject's platelets with U87MG glioblastoma and SH-SY5Y neuroblastoma rho-zero cells) and found that cybrids containing this mtDNA generated higher basal levels of reactive oxygen species but lower ATP production. Moreover, they were more sensitive to oxygen and glucose deprivation and peroxynitrite generation compared with control cybrids. It is unclear whether the 15497 transition is really responsible for these effects. Most problematically, this mutation, in the same individual, changed from heteroplasmic state (55) to homoplasmic state (53) in two studies performed by the same laboratory.

To put the number of dubious cases into relation to the total list of putative disease-associated mutations, we have taken a survey of mutations seen in mitochondrial encephalomyopathies (published in *Neuromuscular Disorders*, January 2004) by way of example. From a total of 97 substitutions listed there, six appear in our East Asian mtDNA tree: three are the 'secondary LHON mutations' (at 3394, 4216 and 13708) and one is the haplogroup G2a3 mutation at 12311 [first found by Hattori *et al.* (56) in chronic progressive external ophthalmoplegia] discussed earlier, and the other two are the 11696 transition [found in Leigh syndrome by De Vries *et al.* (57) and in LHON by Zhou *et al.* (58)] that defines haplogroup D4j and the 8296 transition [found in diabetes and deafness by Kameoka *et al.* (59), in myoclonus epilepsy with ragged-red fibers by Arenas *et al.* (60), in hypertrophic cardiomyopathy by Akita *et al.* (61) and in MELAS by

Sakuta *et al.* (62)] that defines haplogroup D4a2a. This could suggest that the number of haplogroup-defining mutations that came under suspicion of being associated with a disease is not too large, but any such mutation would be visited repeatedly in the context of different diseases, because, by chance, each cohort of patients could include a member of the respective haplogroups. Specifically, Sato and Hayasaka (63) already suggested in response to the Hattori *et al.* (56) study that the 12311 mutation may be a polymorphism. As for the 8296 mutation, Arenas *et al.* (60) judged it of unclear significance and a rare polymorphism. Indeed, it has been shown by Bornstein *et al.* (64,65) through cybrid analysis that ‘*the only effect detected in the A8296G mutation is a moderate decrease in the aminoacylation capacity, which does not affect mitochondrial protein biosynthesis*’.

### A phylogenetic solution

Before attributing a role in disease expression to a certain mtDNA haplogroup-diagnostic mutation, especially when it appears to be basal, great caution should be taken. As all the haplogroup-diagnostic variants shown deeply in the phylogenetic tree have been under selective pressure for thousands of years, their broad distribution among normal populations would suggest them to be rather benign. It is also possible that, in some cases, haplogroup-diagnostic variant(s) might have an additive (or multiplicative, by way of epistatic effects with some nuclear loci) effect to the susceptibility of diseases, as proposed by the common variant–common disease hypothesis; whereas in other cases, the variant(s) could be beneficial in certain environments but deleterious in others. However, to avoid spurious positive associations [as demonstrated in a recent study on East Asian families with hearing impairment (66)] in association studies, it is strongly recommended that, before drawing any hasty conclusion about the pathogenic role of such variant(s), a phylogenetic test should be performed. Any results which could not pass the test of phylogenetic analysis should be considered with major caution, and further studies with a much more refined experimental design should be carried out and an extended large data set (targeting the same population group) should be screened. Specifically, when a direct functional assay cannot be performed, a feasible solution based on phylogenetic knowledge could lend the extra aid for a proper evaluation of pathogenicity of certain haplogroup-specific mutation(s). First, such a study should have a phylogenetic focus and evaluate mtDNA lineages from a large number of well-matched patients and controls in different populations, in order to minimize and/or evaluate the effect of population stratification and efficiently determine whether the single mutation or a set of mutations (or matrilineal background) are associated with the disease. Secondly, the full motifs of the involved haplogroup and its subclade(s) should be taken into consideration according to the reconstructed phylogenetic tree, so as to rule out the possibility that the observed significant association of a certain haplogroup or haplogroup-diagnostic mutation is in fact attributed to its subclade(s) or the other tightly linked mutation(s). Thirdly, to further evaluate the pathogenicity of the diagnostic mutation, additional evidence could also come from the genetic

epidemiological analysis by screening the mutation or the haplogroup in larger samples from general populations, which then helps to reveal whether individuals bearing the mutation or belonging to the haplogroup are more susceptible to the disease. Finally, if possible, further evidence from studying the correlation of the mutation and the disease in other ethnic group(s) of different matrilineal background would be most helpful for a better understanding of the pathogenesis of the disease.

In this spirit, the detailed knowledge of the West Eurasian and East Asian mtDNA phylogenies opens a perspective to reassess the role of mtDNA backgrounds in both mitochondrial and complex disorders, for instance in LHON. In Europe and in populations of predominantly European ancestry, the two frequent primary LHON mutations in ND (NADH dehydrogenase) subunits (*viz.* 11778/ND4 and 14484/ND6) are consistently associated with haplogroup J (67–70). This suggests that the sequence variation within this haplogroup, in particular the transitions at 4216 (characteristic of haplogroups JT and R2), 10398 (back-mutated on haplogroup J but not on T), 13708 (characteristic of haplogroup J) and 3394 (characteristic of haplogroup J1c1), which cause amino acid changes in ND subunits, increase the penetrance of LHON (69,70). A recent study on LHON revealed that such a preferential association (with 11778/ND4) is attributable to sub-haplogroups J1c and J2b (71), thus well explaining the puzzling observation that this association is not confirmed for West Asian populations (72) and further indicating that only a complex pattern of several mutations combined has a notable influence on the etiology of LHON (71). In order to find out whether single occurrences or specific combinations of mutations play any role in this respect, one could envision large-scale studies that target other (sub)continents, e.g. East Asian-specific haplogroups bearing exactly one or two of those mutations. Specifically, the 4216 transition is part of the motif for haplogroup D5c, the 13708 transition is characteristic of haplogroups F2, D5a1, and D4b2a1 and 3394 of M9a'b and D4g1a; further, the ancestral state at site 10398 is present in virtually all macrohaplogroup M lineages as well as in particular sub-haplogroups of macrohaplogroup N that changed back at 10398 (as is the case for haplogroups R11, B5, B4c1c, F3b1 and Y). Such analyses, however, demand a reconstructed phylogeny with a much higher resolution. Therefore, the progress of mtDNA genomics studies should not stop at the identification of basal haplogroups, but aim at identifying more specific sub-haplogroups. This will not only help to reveal the subtle patterns in the past evolution of modern humans (*cf.* 6,7), but also play an important role in determining the involved mutation(s) or lineage(s) in the association studies (*cf.* 71).

Nevertheless, it is still possible that the same mtDNA variant (or haplotype) could contribute differently to the susceptibility of the disease in different human populations (73), mainly due to epistatic effects between e.g. some nuclear DNA factor and the mtDNA variant (haplotype) under study. Then, ideally, genotyping of an appropriate number (e.g. 40) of (apparently) neutral unlinked markers (e.g. SNPs) (the genomic control procedure) (74) in cases and controls would help to rule out the existence of a certain level of population stratification that could affect the

association study. The structure assessment method popularized by Pritchard *et al.* (75,76) could be equally helpful, but again, it requires to genotype autosomal unlinked genetic markers.

In summary, the combination of virtually all previously reported and our newly produced complete East Asian mtDNA sequences have helped to define a large number of novel (sub)haplogroups as well as further solidify previously reported haplogroups, which in turn have greatly refined the phylogenetic tree of mtDNAs found in East Asian populations. This updated phylogenetic tree with its standardized nomenclature resolves the confusion of nomenclature introduced by inadvertent neglect of the literature and provides an essential reference guide for disease, anthropological and forensic studies in East Asian populations.

## MATERIALS AND METHODS

### Data collection

To determine the full motifs of the newly emerging novel haplogroups, a total of 20 samples (GenBank accession nos. DQ272107–DQ272126), representing 20 different (sub)haplogroups, were selected from our mtDNA database (27,77,78) (authors' unpublished data) for complete sequencing. To identify the motifs of novel (sub)haplogroups, we systematically screened candidate mutations in control mtDNAs which appear to be closely related (as judged from the control-region and partial coding-region information) to complete mtDNA sequences constituting so far singular terminal branches of the estimated mtDNA tree, with the aim to distinguish private mutations from shared ones that are diagnostic for some twigs of the mtDNA tree ('motif search').

### Motif search

The motifs (strings of characteristic/diagnostic mutations shared by descent) of the most basal haplogroups could immediately be read off from the 'trunks' (30) or other major branches of the mtDNA tree, whereas the motifs of some rare basal haplogroups or minor lineages at the 'twig' level are more difficult to determine, as these haplogroups may be represented by only one complete sequence in the database. To identify the motifs of such lineages, the direct way is to completely sequence more representative samples. Another feasible way is, alternatively, to screen the polymorphisms of the representative sample allocated to the corresponding (terminal) branch of the tree in other mtDNAs (*viz.* controls) that appear to be closely related or (near-)matched according to available control-region and/or partial coding-region information. Then the motif of the haplogroup (*viz.* the shared variants) can be easily distinguished from the private polymorphisms of the representative sample. Such a motif-search strategy can thus serve as a useful auxiliary method of mtDNA genome study for the phylogeny reconstruction. As a result, a total of 59 control samples, which were selected from our mtDNA database (27,28,77–80) (authors' unpublished data), were analyzed for short coding-region segments. The amplification and sequencing

procedures performed in the present study have been fully described elsewhere (5,8,15), and the resulting sequences were edited with DNASTAR software (DNASTAR Inc.). Mutations were scored relative to the revised Cambridge reference sequence (81). To ensure high quality of our sequencing results, we adopted previously described quality-control measures (5) and some suggestions (82–84) for data generation and editing.

### Annotated tree reconstruction

Besides our newly produced data, a total of 1001 reported near-)complete mtDNA sequences belonging to East Asian haplogroups (1–5,10,12,14,18,19,24,29,30,48,85–88) were reanalyzed in order to further annotate the available East Asian mtDNA phylogeny (5). To reconcile the nomenclature conflicts between different studies, a chronological criterion was employed in the present study, that is, the haplogroup name or definition, which was harmonious with the previous haplogroup system and was reported first, was preferred.

Several prerequisites needed to be in place before entering the published sequences for tree reconstruction: (i) the 13 sequences from Tanaka *et al.* (24) that were very likely affected by artificial recombination (authors' unpublished data) were excluded from the analysis. Some of the potential recombinants have even been reflected by the summary trees offered by the authors (24); for instance, the seemingly 'private' variants 6455, 12406, 12771, 12882, 13759, 13928C and 14002 of ND168 (haplogroup M7a2) in their Figure 1a are in fact faithfully 'borrowed' from the haplogroup F1a1b branch. (ii) Several obvious editing errors, such as 16569+G and 16569+GATCACAG, were corrected. (iii) In consideration of the relatively high frequency of coding-region homoplasmy in those 672 sequences as well as the possibility of some potential recombinants undetected which involved only one mutation, we adopted a more conservative approach here for the tree reconstruction: any observed novel intermediate node that was introduced by the presence or absence of one or two adjacent mutations, when supported by only one sequence, was not considered. (iv) Errors in the three complete mtDNAs of Zhao *et al.* (88) were eliminated according to Yao *et al.* (66).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

## ACKNOWLEDGEMENTS

We thank the two anonymous referees for their valuable comments. The work was supported by grants from the Natural Science Foundation of Yunnan Province, Natural Science Foundation of China (NSFC, No. 30021004), the Chinese Academy of Sciences (KSCX2-SW-2010), Fondo Investimenti Ricerca di Base 2001 and Fondazione Cariplo (to A.T.), the National Key Technologies R&D Program of China and the Science and Technology Committee of Yunnan Province (2005NG07).

*Conflict of Interest statement.* None declared.

## REFERENCES

- Ingman, M., Kaessmann, H., Pääbo, S. and Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**, 708–713.
- Finnilä, S., Lehtonen, M.S. and Majamaa, K. (2001) Phylogenetic network for European mtDNA. *Am. J. Hum. Genet.*, **68**, 1475–1484.
- 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. *et al.* (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. [Erratum **71**, 448–449.] *Am. J. Hum. Genet.*, **70**, 1152–1171.
- Ingman, M. and Gyllensten, U. (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.*, **13**, 1600–1606.
- Kong, Q.-P., Yao, Y.-G., Sun, C., Bandelt, H.-J., Zhu, C.L. and Zhang, Y.-P. (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. [Erratum **75** 157.] *Am. J. Hum. Genet.*, **73**, 671–676.
- Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V. *et al.* (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am. J. Hum. Genet.*, **75**, 910–918.
- Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S., Magri, C., Scozzari, R., Babudri, N., Santachiara-Benerecetti, A.S. *et al.* (2005) Saami and Berbers—an unexpected mitochondrial DNA link. *Am. J. Hum. Genet.*, **76**, 883–886.
- Palanichamy, M.G., Sun, C., Agrawal, S., Bandelt, H.-J., Kong, Q.-P., Khan, F., Wang, C.Y., Chaudhuri, T.K., Palla, V. and Zhang, Y.-P. (2004) Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am. J. Hum. Genet.*, **75**, 966–978.
- Friedlaender, J., Schurr, T., Gentz, F., Koki, G., Friedlaender, F., Horvat, G., Babb, P., Cerchio, S., Kaestle, F., Schanfield, M. *et al.* (2005) Expanding southwest Pacific mitochondrial haplogroups P and Q. *Mol. Biol. Evol.*, **22**, 1506–1517.
- Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F. *et al.* (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, **308**, 1034–1036.
- Merrifether, D.A., Hodgson, J.A., Friedlaender, F.R., Allaby, R., Cerchio, S., Koki, G. and Friedlaender, J.S. (2005) Ancient mitochondrial M haplogroups identified in the Southwest Pacific. *Proc. Natl Acad. Sci. USA*, **102**, 13034–13039.
- Trejaut, J.A., Kivisild, T., Loo, J.H., Lee, C.L., He, C.L., Hsu, C.J., Li, Z.Y. and Lin, M. (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol.*, **3**, e247.
- Behar, D.M., Metspalu, E., Kivisild, T., Achilli, A., Hadid, Y., Tzur, S., Pereira, L., Amorim, A., Quintana-Murci, L., Majamaa, K. *et al.* (2006) The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. *Am. J. Hum. Genet.*, **78**, 487–497.
- Kivisild, T., Shen, P., Wall, D.P., Do, B., Sung, R., Davis, K., Passarino, G., Underhill, P.A., Scharfe, C., Torroni, A. *et al.* (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**, 373–387.
- Sun, C., Kong, Q.-P., Palanichamy, M.G., Agrawal, S., Bandelt, H.-J., Yao, Y.-G., Khan, F., Zhu, C.-L., Chaudhuri, T.K. and Zhang, Y.-P. (2006) The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.*, **23**, 683–690.
- Stringer, C. (2000) Palaeoanthropology Coasting out of Africa. *Nature*, **405**, 24–25.
- Palanichamy, M.G., Agrawal, S., Yao, Y.-G., Kong, Q.-P., Sun, C., Khan, F., Chaudhuri, T.K. and Zhang, Y.-P. (2006) Comment on 'Reconstructing the origin of Andaman Islanders'. *Science*, **311**, 470.
- Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A.G., Singh, V.K., Rasalkar, A.A. and Singh, L. (2005) Reconstructing the origin of Andaman Islanders. *Science*, **308**, 996.
- Bandelt, H.-J., Achilli, A., Kong, Q.-P., Salas, A., Lutz-Bonengel, S., Sun, C., Zhang, Y.-P., Torroni, A. and Yao, Y.-G. (2005) Low 'penetrance' of phylogenetic knowledge in mitochondrial disease studies. *Biochem. Biophys. Res. Commun.*, **333**, 122–130.
- Salas, A., Yao, Y.-G., Macaulay, V., Vega, A., Carracedo, A. and Bandelt, H.-J. (2005) A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med.*, **2**, e296.
- Marchini, J., Cardon, L.R., Phillips, M.S. and Donnelly, P. (2004) The effects of human population structure on large genetic association studies. *Nat. Genet.*, **36**, 512–517.
- Devlin, B., Bacanu, S.A. and Roeder, K. (2004) Genomic Control to the extreme. *Nat. Genet.*, **36**, 1129–1130.
- Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H.E. *et al.* (2005) Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.*, **37**, 1243–1246.
- Tanaka, M., Cabrera, V.M., González, A.M., Larruga, J.M., Takeyasu, T., Fuku, N., Guo, L.J., Hirose, R., Fujita, Y., Kurata, M. *et al.* (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.*, **14**, 1832–1850.
- Howell, N., Elson, J.L., Chinnery, P.F. and Turnbull, D.M. (2005) mtDNA mutations and common neurodegenerative disorders. *Trends Genet.*, **21**, 583–586.
- Da Pozzo, P. and Federico, A. (2005) Commentary to mtDNA research for the pathogenesis of mitochondrial disorders. *Biochem. Biophys. Res. Commun.*, **336**, 1003–1004.
- Yao, Y.-G., Kong, Q.-P., Wang, C.-Y., Zhu, C.-L. and Zhang, Y.-P. (2004) Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Mol. Biol. Evol.*, **21**, 2265–2280.
- Yao, Y.-G., Kong, Q.-P., Bandelt, H.-J., Kivisild, T. and Zhang, Y.-P. (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.*, **70**, 635–651.
- Starikovskaya, E.B., Sukernik, R.I., Derbeneva, O.A., Volodko, N.V., Ruiz-Pesini, E., Torroni, A., Brown, M.D., Lott, M.T., Hosseini, S.H., Huoponen, K. *et al.* (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann. Hum. Genet.*, **69**, 67–89.
- Kivisild, T., Tolk, H.-V., Parik, J., Wang, Y., Papiha, S.S., Bandelt, H.-J. and Villems, R. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. [Erratum **20**, 162.] *Mol. Biol. Evol.*, **19**, 1737–1751.
- Ozawa, T. (1995) Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. *Biochim. Biophys. Acta*, **1271**, 177–189.
- Munakata, K., Tanaka, M., Mori, K., Washizuka, S., Yoneda, M., Tajima, O., Akiyama, T., Nanko, S., Kunugi, H., Tadokoro, K. *et al.* (2004) Mitochondrial DNA 3644T→C mutation associated with bipolar disorder. *Genomics*, **84**, 1041–1050.
- Chiu, P.M., Liu, V.W.S., Ngan, H.Y.S., Khoo, U.S. and Cheung, A.N.Y. (2003) Detection of mitochondrial DNA mutations in gestational trophoblastic disease. *Hum. Mutat.*, **22**, 177.
- Bandelt, H.-J., Kong, Q.-P., Parson, W. and Salas, A. (2005) More evidence for non-maternal inheritance of mitochondrial DNA? *J. Med. Genet.*, **42**, 957–960.
- Bandelt, H.-J., Yao, Y.-G. and Kivisild, T. (2005) Mitochondrial genes and schizophrenia. *Schizophr. Res.*, **72**, 267–269.
- Yao, Y.-G., Kong, Q.-P. and Zhang, Y.-P. (2002) Mitochondrial DNA 5178A polymorphism and longevity. *Hum. Genet.*, **111**, 462–463.
- Johns, D.R., Neufeld, M.J. and Park, R.D. (1992) An ND-6 mitochondrial DNA mutation associated with Leber hereditary optic neuropathy. *Biochem. Biophys. Res. Commun.*, **187**, 1551–1557.
- Obayashi, T., Hattori, K., Sugiyama, S., Tanaka, M., Tanaka, T., Itoyama, S., Deguchi, H., Kawamura, K., Koga, Y., Toshima, H. *et al.* (1992) Point mutations in mitochondrial DNA in patients with hypertrophic cardiomyopathy. *Amer. Heart J.*, **124**, 1263–1269.
- Hirai, M., Suzuki, S., Onoda, M., Hinokio, Y., Ai, L., Hirai, A., Ohtomo, M., Komatsu, K., Kasuga, S., Satoh, Y. *et al.* (1996) Mitochondrial DNA 3394 mutation in the NADH dehydrogenase subunit 1 associated with non-insulin-dependent diabetes mellitus. *Biochem. Biophys. Res. Commun.*, **219**, 951–955.
- Chen, Y., Liao, W.X., Roy, A.C., Loganath, A. and Ng, S.C. (2000) Mitochondrial gene mutations in gestational diabetes mellitus. *Diabetes Res. Clin. Pract.*, **48**, 29–35.
- Ohkubo, K., Yamano, A., Nagashima, M., Mori, Y., Anzai, K., Akehi, Y., Nomiya, R., Asano, T., Urae, A. and Ono, J. (2001) Mitochondrial gene mutations in the tRNA(Leu(UUR)) region and diabetes: prevalence and clinical phenotypes in Japan. *Clin. Chem.*, **47**, 1641–1648.

42. Sciacco, M., Prelle, A., Fagioli, G., Bordoni, A., Crimi, M., Di Fonzo, A., Ciscato, P., Lamperti, C., D'Adda, E., Jann, S. *et al.* (2005) A case of CPT deficiency, homoplasmic mtDNA mutation and ragged red fibers at muscle biopsy. *J. Neurol. Sci.*, **239**, 21–24.
43. Nishimaki, Y., Sato, K., Fang, L., Ma, M., Hasekura, H. and Boettcher, B. (1999) Sequence polymorphism in the mtDNA HV1 region in Japanese and Chinese. *Legal Med.*, **1**, 238–249.
44. Imaizumi, K., Parsons, T.J., Yoshino, M. and 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.
45. Tajima, A., Hayami, M., Tokunaga, K., Juji, T., Matsuo, M., Marzuki, S., Omoto, K. and Horai, S. (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J. Hum. Genet.*, **49**, 187–193.
46. Maruyama, S., Minaguchi, K. and 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.
47. Grazina, M., Silva, F., Santana, I., Santiago, B., Mendes, C., Simões, M., Oliveira, M., Cunha, L. and Oliveira, C. (2004) Frontotemporal dementia and mitochondrial DNA transitions. *Neurobiol. Dis.*, **15**, 306–311.
48. Derbeneva, O.A., Sukernik, R.I., Volodko, N.V., Hosseini, S.H., Lott, M.T. and 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.
49. Herrnstadt, C., Preston, G. and Howell, N. (2003) Errors, phantoms and otherwise, in human mtDNA sequences. *Am. J. Hum. Genet.*, **72**, 1585–1586.
50. Ohkubo, E., Aida, K., Chen, J., Hayashi, J.I., Isobe, K., Tawata, M. and Onaya, T. (2000) A patient with type 2 diabetes mellitus associated with mutations in calcium sensing receptor gene and mitochondrial DNA. *Biochem. Biophys. Res. Commun.*, **278**, 808–813.
51. Okura, T., Koda, M., Ando, F., Niino, N., Tanaka, M. and 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.
52. Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B. and 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.
53. Tarnopolsky, M.A., Simon, D.K., Roy, B.D., Chorneyko, K., Lowther, S.A., Johns, D.R., Sandhu, J.K., Li, Y. and Sikorska, M. (2004) Attenuation of free radical production and paracrystalline inclusions by creatine supplementation in a patient with a novel cytochrome *b* mutation. *Muscle Nerve*, **29**, 537–547.
54. Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. and Torroni, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. [Erratum **64**, 918–919.] *Am. J. Hum. Genet.*, **64**, 232–249.
55. Tarnopolsky, M.A. (2002) When fatigue is more than simple exertion. *Phys. Sportsmed.*, **30**, Available at Internet: [http://www.physsportsmed.com/issues/2002/06\\_02/tarno.htm](http://www.physsportsmed.com/issues/2002/06_02/tarno.htm).
56. Hattori, Y., Goto, Y., Sakuta, R., Nonaka, I., Mizuno, Y. and Horai, S. (1994) Point mutations in mitochondrial tRNA genes: sequence analysis of chronic progressive external ophthalmoplegia (CPEO). *J. Neurol. Sci.*, **125**, 50–55.
57. De Vries, D.D., Went, L.N., Bruyn, G.W., Scholte, H.R., Hofstra, R.M., Bolhuis, P.A. and van Oost, B.A. (1996) Genetic and biochemical impairment of mitochondrial complex I activity in a family with Leber hereditary optic neuropathy and hereditary spastic dystonia. *Am. J. Hum. Genet.*, **58**, 703–711.
58. Zhou, X., Wei, Q., Yang, L., Tong, Y., Zhao, F., Lu, C., Qian, Y., Sun, Y., Lu, F., Qu, J. *et al.* (2006) Leber's hereditary optic neuropathy is associated with the mitochondrial ND4 G11696A mutation in five Chinese families. *Biochem. Biophys. Res. Commun.*, **340**, 69–75.
59. Kameoka, K., Isotani, H., Tanaka, K., Azukari, K., Fujimura, Y., Shiota, Y., Sasaki, E., Majima, M., Furukawa, K., Haginomori, S. *et al.* (1998) Novel mitochondrial DNA mutation in tRNA(Lys) (8296A→G) associated with diabetes. *Biochem. Biophys. Res. Commun.*, **245**, 523–527.
60. Arenas, J., Campos, Y., Bornstein, B., Ribacoba, R., Martin, M.A., Rubio, J.C., Santorelli, F.M., Zeviani, M., DiMauro, S. and Garesse, R. (1999) A double mutation (A8296G and G8363A) in the mitochondrial DNA tRNA (Lys) gene associated with myoclonus epilepsy with ragged-red fibers. *Neurology*, **52**, 377–382.
61. Akita, Y., Koga, Y., Iwanaga, R., Wada, N., Tsubone, J., Fukuda, S., Nakamura, Y. and Kato, H. (2000) Fatal hypertrophic cardiomyopathy associated with an A8296G mutation in the mitochondrial tRNA(Lys) gene. *Hum. Mutat.*, **15**, 382.
62. Sakuta, R., Honzawa, S., Murakami, N., Goto, Y. and Nagai, T. (2002) Atypical MELAS associated with mitochondrial tRNA(Lys) gene A8296G mutation. *Pediatr. Neurol.*, **27**, 397–400.
63. Sato, W. and Hayasaka, K. (1994) A T-to-C substitution at nucleotide 12311 in tRNA(Leu(CUN)) gene may be a mtDNA polymorphism. *J. Neurol. Sci.*, **127**, 236.
64. Bornstein, B., Mas, J.A., Fernández-Moreno, M.A., Campos, Y., Martin, M.A., del Hoyo, P., Rubio, J.C., Arenas, J. and Garesse, R. (2002) The A8296G mtDNA mutation associated with several mitochondrial diseases does not cause mitochondrial dysfunction in cybrid cell lines. *Hum. Mutat.*, **19**, 234–239.
65. Bornstein, B., Mas, J.A., Patrono, C., Fernández-Moreno, M.A., González-Vioque, E., Campos, Y., Carrozzo, R., Martín, M.A., del Hoyo, P., Santorelli, F.M. *et al.* (2005) Comparative analysis of the pathogenic mechanisms associated with the G8363A and A8296G mutations in the mitochondrial tRNA(Lys) gene. *Biochem. J.*, **387**, 773–778.
66. Yao, Y.-G., Salas, A., Bravi, C.M. and Bandelt, H.-J. (2006) A reappraisal of complete mtDNA variation in East Asian families with hearing impairment. *Hum. Genet.*, **119**, 505–515.
67. Howell, N., Kubacka, I., Halvorson, S., Howell, B., McCullough, D.A. and Mackey, D. (1995) Phylogenetic analysis of the mitochondrial genomes from Leber hereditary optic neuropathy pedigrees. *Genetics*, **140**, 285–302.
68. Howell, N., Oostra, R.J., Bolhuis, P.A., Spruijt, L., Clarke, L.A., Mackey, D.A., Preston, G. and Herrnstadt, C. (2003) Sequence analysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy. *Am. J. Hum. Genet.*, **72**, 1460–1469.
69. Brown, M.D., Sun, F. and Wallace, D.C. (1997) Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage. *Am. J. Hum. Genet.*, **60**, 381–387.
70. Torroni, A., Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P. *et al.* (1997) Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am. J. Hum. Genet.*, **60**, 1107–1121.
71. Carelli, V., Achilli, A., Valentino, M.L., Rengo, C., Semino, O., Pala, M., Olivieri, A., Mattiuzzi, M., Pallotti, F., Carrara, F. *et al.* (2006) Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees. *Am. J. Hum. Genet.*, **78**, 564–574.
72. Houshmand, M., Sharifpanah, F., Tabasi, A., Sanati, M.-H., Vakilian, M., Lavasani, S.H. and Jougehhdoust, S. (2004) Leber's hereditary optic neuropathy: the spectrum of mitochondrial DNA mutations in Iranian patients. *Ann. NY Acad. Sci.*, **1011**, 345–349.
73. Ioannidis, J.P., Ntzani, E.E. and Trikalinos, T.A. (2004) 'Racial' differences in genetic effects for complex diseases. *Nat. Genet.*, **36**, 1312–1318.
74. Devlin, B., Roeder, K. and Wasserman, L. (2000) Genomic control for association studies: a semiparametric test to detect excess-haplotype sharing. *Biostatistics*, **1**, 369–387.
75. Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
76. Pritchard, J.K., Stephens, M., Rosenberg, N.A. and Donnelly, P. (2000) Association mapping in structured populations. *Am. J. Hum. Genet.*, **67**, 170–181.
77. Kong, Q.-P., Yao, Y.-G., Liu, M., Shen, S.-P., Chen, C., Zhu, C.-L., Palanichamy, M.G. and Zhang, Y.-P. (2003) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum. Genet.*, **113**, 391–405.
78. Kong, Q.-P., Yao, Y.-G., Sun, C., Zhu, C.-L., Zhong, L., Wang, C.-Y., Cai, W.-W., Xu, X.-M., Xu, A.-L. and Zhang, Y.-P. (2004) Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic. *J. Hum. Genet.*, **49**, 414–423.

79. Yao, Y.-G., Nie, L., Harpending, H., Fu, Y.-X., Yuan, Z.-G. and Zhang, Y.-P. (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am. J. Phys. Anthropol.*, **118**, 63–76.
80. Yao, Y.-G., Kong, Q.-P., Man, X.-Y., Bandelt, H.-J. and Zhang, Y.-P. (2003) Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol. Biol. Evol.*, **20**, 214–219.
81. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.*, **23**, 147.
82. Bandelt, H.-J., Lahermo, P., Richards, M. and Macaulay, V. (2001) Detecting errors in mtDNA data by phylogenetic analysis. *Int. J. Legal Med.*, **115**, 64–69.
83. Yao, Y.-G., Macaulay, V., Kivisild, T., Zhang, Y.-P. and Bandelt, H.-J. (2003) To trust or not to trust an idiosyncratic mitochondrial data set. *Am. J. Hum. Genet.*, **72**, 1341–1346. (response **72**, 1348–1349).
84. Yao, Y.-G., Bravi, C.M. and Bandelt, H.-J. (2004) A call for mtDNA data quality control in forensic science. *Forensic Sci. Int.*, **141**, 1–6.
85. Ozawa, T., Tanaka, M., Ino, H., Ohno, K., Sano, T., Wada, Y., Yoneda, M., Tanno, Y., Miyatake, T., Tanaka, T. *et al.* (1991) Distinct clustering of point mutations in mitochondrial DNA among patients with mitochondrial encephalomyopathies and with Parkinson's disease. *Biochem. Biophys. Res. Commun.*, **176**, 938–946.
86. Maca-Meyer, N., González, A.M., Larruga, J.M., Flores, C. and Cabrera, V.M. (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.*, **2**, 13.
87. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A.G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M.D. *et al.* (2003) Natural selection shaped regional mtDNA variation in humans. *Proc. Natl Acad. Sci. USA*, **100**, 171–176.
88. Zhao, L., Young, W.-Y., Li, R., Wang, Q., Qian, Y. and Guan, M.-X. (2004) Clinical evaluation and sequence analysis of the complete mitochondrial genome of three Chinese patients with hearing impairment associated with the 12S rRNA T1095C mutation. *Biochem. Biophys. Res. Commun.*, **325**, 1503–1508.