ORIGINAL ARTICLE

Median network analysis of defectively sequenced entire mitochondrial genomes from early and contemporary disease studies

Hans-Jürgen Bandelt[1], Yong-Gang Yao[2], Claudio M Bravi[3], Antonio Salas[4] and Toomas Kivisild[5]

Sequence analysis of the mitochondrial genome has become a routine method in the study of mitochondrial diseases. Quite often, the sequencing efforts in the search of pathogenic or disease-associated mutations are affected by technical and interpretive problems, caused by sample mix-up, contamination, biochemical problems, incomplete sequencing, misdocumentation and insufficient reference to previously published data. To assess data quality in case studies of mitochondrial diseases, it is recommended to compare any mtDNA sequence under consideration to their phylogenetically closest lineages available in the Web. The median network method has proven useful for visualizing potential problems with the data. We contrast some early reports of complete mtDNA sequences to more recent total mtDNA sequencing efforts in studies of various mitochondrial diseases. We conclude that the quality of complete mtDNA sequences generated in the medical field in the past few years is somewhat unsatisfactory and may even fall behind that of pioneer manual sequencing in the early nineties. Our study provides a paradigm for an *a posteriori* evaluation of sequence quality and for detection of potential problems with inferring a pathogenic status of a particular mutation.

INTRODUCTION

Sequencing of the whole mitochondrial genome seems to be a routine exercise, deemed to be technically achievable by many medical labs nowadays—yet it is fraught with the very problems already known to affect minor-scaled projects, which are, regrettably, not always recognized in medical genetics. Samples can easily get mixed up or contaminated,[1–8] phantom mutations can plague sequencing results[9–13] and documentation errors or casual sequencing can distort the sequencing results considerably;[14–17] see Bandelt *et al.*[18] for an overview. In consequence, deficient sequencing efforts in medical research can miss the causal mutations, and this could in principle contribute to the negative results in seeking for the mutations responsible for a pathological phenotype. On the other hand, sequences of suboptimal quality could lead to false imputations of pathogenicity.

Here, we demonstrate that most of these problems could have been detected in time and thus avoided if an up-to-date knowledge of the global mtDNA phylogeny had been employed. The basal mtDNA phylogeny is described in terms of haplogroups encoded by letter–number codes and the mutations that distinguish them.[19] Allocating freshly obtained sequences to major haplogroups is the first step of a phylogenetic analysis of human mtDNA. The second step is to ascertain whether all haplogroup-specific mutations in the new sequences were actually observed, following the evolutionary pathway connecting the targeted sequence with the revised Cambridge reference sequence (rCRS: see Andrews *et al.*[20]) over their common root. The next step is to perform a fine-scale phylogenetic analysis of the targeted sequence together with the available sequences of the same sub-haplogroup it can be allocated to. An evaluation of the mutational events involved can be assisted by visualizing these few sequences in a median network that would highlight particular patterns of homoplasy, which are induced by sequencing problems or the natural cause of recurrent mutation.[21] We follow this strategy in re-analyzing some published complete mtDNA sequences and evaluate the causes for inferred back mutations.

MATERIALS AND METHODS

Global mtDNA phylogeny

The emerging worldwide mtDNA phylogeny can be viewed in the form of subcontinental trees, which capture the basal variation known to date and provide the information on haplogroups and reconstruct the mutational events along the evolutionary pathways.[19] For a first acquaintance one may visit the

exhibition of the major West and East Eurasian haplogroups as presented in the trees from Palanichamy et al.[22] and Kong et al.[23] The West Eurasian tree has been further refined and modified in peripheral parts by Achilli et al.,[24,25] Loogväli et al.,[26] Behar et al.,[27] Derenko et al.,[28] and Roostalu et al.[29]

To connect, for instance, some East Asian mtDNA sequence with the reference sequence, one would eventually pass through the root of haplogroup R through the roots (that is, ancestral haplotypes) of a number of nested haplogroups. If the sequence belongs to haplogroup M, then this pathway would first step down to the African root of haplogroup L3 and then move up until it reaches the rCRS:

$$M \leftarrow L3 \rightarrow N \rightarrow R \rightarrow R0 = \text{pre-HV} \rightarrow HV \rightarrow H \rightarrow H2 \rightarrow H2a \rightarrow H2a2 \rightarrow rCRS.$$

The rCRS is separated from the mosaic original CRS[30] by 11 'error' mutations.[20] Note that the names H2a and H2a2 for the nested sub-haplogroups of H2 harboring the rCRS have been introduced by Achilli et al.[24] and Roostalu et al.[29]

Networks
Networks constitute an ideal tool for exploring features of recurrent mutations in a data set.[10,21,31,32] Observed or partially reconstructed mutational patterns reflecting homoplasic changes can be represented by a median network provided that the observed or postulated mutations involve only two nucleotide states.[33] We propose to generate this kind of network for comparing a single sequence or very few sequences under examination to the closest relatives from the worldwide mtDNA database, together with the entire path of reconstructed ancestral haplotypes connecting the sequence(s) to the rCRS. In this way, one can evaluate the specific haplogroup allocation and detect any unusual combination of parallel mutations or reversals, which could signal problems in sequencing or documentation.

One can in principle use software such as Network 4.502 (available from http://www.fluxus-engineering.com/sharenet.htm) for generating median networks, provided that (i) the hypothesized evolution of a mutation recurrent along a postulated evolutionary pathway is formally encoded as hitting different sites[33] and (ii) the participating haplotypes are encoded in terms of all variable sites. In the particular situation that a single sequence is plotted against its reconstructed evolutionary pathway the median network is straightforward to construct by hand; it generically has the structure of a half-grid with three terminal links: one to the rCRS (and further extended to the CRS, if necessary with older data), one to the sequence under study and another one to its closest relative on the pathway (or to the root of the sub-haplogroup it is associated with). We let the mutations label the horizontal and vertical line segments. Then the pathway from the rCRS to the terminal haplotype (representing some published sequence or the inferred root of a sub-haplogroup) zigzags so that each horizontal segment confirms this part of the pathway for the sequence under examination, whereas each vertical segment involved in reticulations (signifying character conflicts) indicates potentially missed mutations. Similar diagrams have been used in the case of a recombinant sequence and its two constituents or their inferred close relatives.[8] In fact, missed mutations could be regarded conceptually as resulting from virtual recombination between the targeted sequence and the default rCRS.

'Algorithm'
The following procedure for evaluating new mtDNA sequences can supplement expert knowledge in complete mtDNA variation and direct the user to the correct classification of his/her mtDNA under consideration, so that (s)he could get alarmed whenever mutations are missing that would be expected from the mtDNA phylogeny.

(1) All mutations in the targeted sequence should be scored relative to the rCRS of Andrews et al.[20] by applying the conventional 'medical' notation adhered to in the present text. This, in particular, invokes the historic scoring of position 3106 as a gap in the rCRS (to maintain the old CRS numbering of nucleotides)—rather than 'N' as in the current MITOMAP version of the rCRS (J01415.2). The MITOMAP scoring constitutes an abuse of the code letter 'N' because, according to the IUPAC nucleotide code (http://www.bioinformatics.org/sms/iupac.html), 'N' designates 'any base' rather than a missing position. This has already led to some confusion: for example, the seven complete mtDNA sequences from Montiel-Sosa et al.[34] were all reported with 'N' at position 3107 (GenBank accession no. DQ156208–DQ156214) but were subsequently cited on a website by erroneously turning this information at 3107 into a 'mutation' C3107N (http://freepages.genealogy.rootsweb.com/~ncscotts/mtDNA/GenBank%20Mutation%20Lists/hg%20U/hg_U_mutation_list.htm). More confusingly, the sequences in GenBank (accession no. EF184580–EF184641) from Gonder et al.[35] were variably listed with 'C' (20 times) and with 'CN' (42 times) at positions 3106–3107 (see http://www.ianlogan.co.uk/lists/gonder.htm for a quick overview). We therefore advise downloading the MITOMAP rCRS and then replacing 'CN' by '–C' to conform with the originally proposed rCRS of Andrews et al.[20]

(2) Start sorting coding-region mutations from the target sequence into the slots of the chain $(L \rightarrow L1'5 \rightarrow L2'5 \rightarrow L2'6 \rightarrow L4'6 \rightarrow L3'7 \rightarrow) \ L3 \rightarrow N \rightarrow R \rightarrow R0 \rightarrow HV \rightarrow H \rightarrow H2 \rightarrow H2a \rightarrow H2a2 \rightarrow rCRS$. Check whether the batches of mutations filled in are as expected, without any mutation missing. Branch off from the haplogroup that was last hit in the chain to a nesting of sub-haplogroups. In case L3 is reached, decide whether one has either to ascend from L3 to M or some African branch of L3 (L3a, L3bcd, L3eix, L3f and L3h) or to descend from the root of L3 further down towards the root of L (see Figure 1 of Torroni et al.[19]).

(3) Google or Yahoo search any single mutation from the target sequence not yet assigned to the preceding basal pathway from the rCRS. For instance, in the case of mutation T669C, one would enter the query 'mtDNA T669C' (or the like; see Bandelt et al.[36,37]). Then typically one is directed to the website of Ian Logan (http://www.ianlogan.co.uk/mtdna.htm), which is most up-to-date in sorting the complete mtDNA sequences from GenBank into their haplogroup slots (best taken with a pinch of salt though). For T669C, the haplogroup in question is N1a (which it defines), along with the proper reference. Many other query results would point to published papers, mainly from the medical literature. In the case of T669C one would then learn that this mutation was most recently suspected of being putatively pathogenic. In contrast to this mutation, other mutations would often point to more than one haplogroup. Then one would resort to the haplogroup (and those particular lineages), which combines most of the mutations present in the target sequence.

(4) Next, it is worthwhile to check whether there are additional occurrences of the particular mutation in the somewhat older literature, by entering a query to MITOMAP (http://www.mitomap.org/). Moreover, one can search the mutation in the MITOMAP tree at the site (http://www.mitomap.org/mitomap-phylogeny.pdf). For example, no search result (as of 24 September 2008) for T669C, but well a place with the label N1a in the MITOMAP tree.

(5) Once a candidate haplogroup is found, perform a web-based search again, for instance, by entering 'mtDNA haplogroup N1a', or 'complete mtDNA haplogroup N1a' in Google for a sharper focus. As in this example, one would generally find most of the recent papers, which present tree views of that haplogroup in context.

(6) Compile a file with the target sequence and all closely related complete mtDNA sequences (of the same sub-haplogroup) from the previous searches.[37] Add to that file all the reconstructed ancestral sequences of the haplogroups along the entire pathway to the rCRS (as in Figures 1–3 below). Then build up the median network by hand (as done in this paper) or feed the file into the program Network 4.502. Find out which reticulation is added to the network by the target sequence and re-check the mutations involved in the excess reticulation.

RESULTS
West Eurasian mtDNA sequences
Complete sequencing of mtDNA performed in the early nineties was likely to fail in recording all variant nucleotides relative to CRS (or some partially corrected version of the CRS). At that time, the sequencing equipment and chemistry was somewhat inferior to what is available nowadays, and sequencing or documentation errors were hard to spot as there were only few related sequences around for comparison. A first systematic approach to analyze European mtDNA in a phylogenetic context was undertaken based on yet incomplete mtDNA information.[38] These data in connection with further control-region data and restriction fragment length polymorphisms then
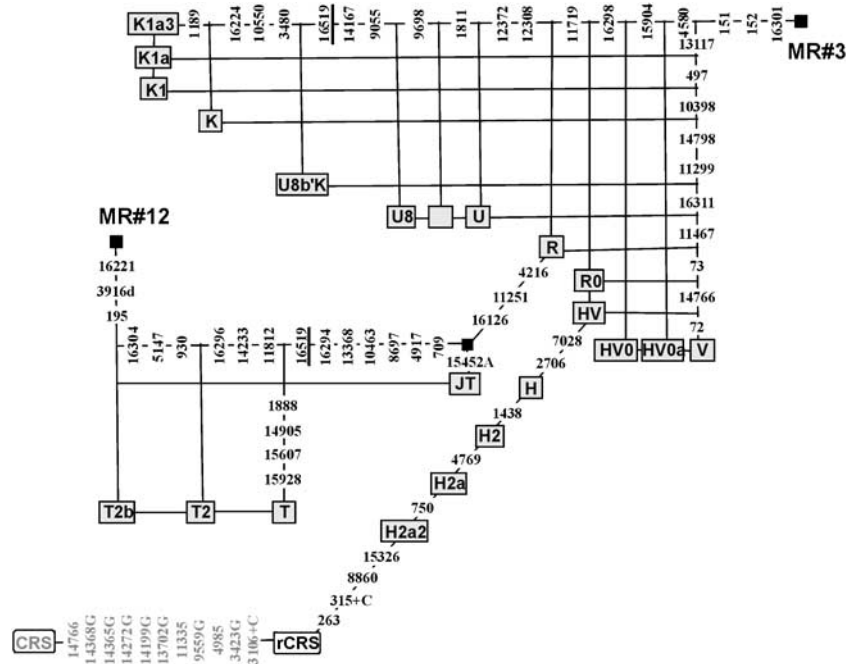
Figure 1 Median network of two complete mtDNA sequences (MR no. 3 and MR no. 12) from Rieder *et al.*[39] together with the evolutionary pathways of the West Eurasian mtDNA phylogeny[22] connecting related ancestral haplotypes to the rCRS.[20] Open squares indicate inferred ancestral haplotypes, with the names of the corresponding haplogroups they determine inscribed. Numbers designate mutations, which are transitions unless a suffix indicates a transversion (A, G, C or T), an insertion (+) or a deletion (d) relative to rCRS. Line segments opposite in a rectangle are associated with the same mutations. Recurrent mutations at position 16519 in the mtDNA phylogeny are highlighted by underlining.
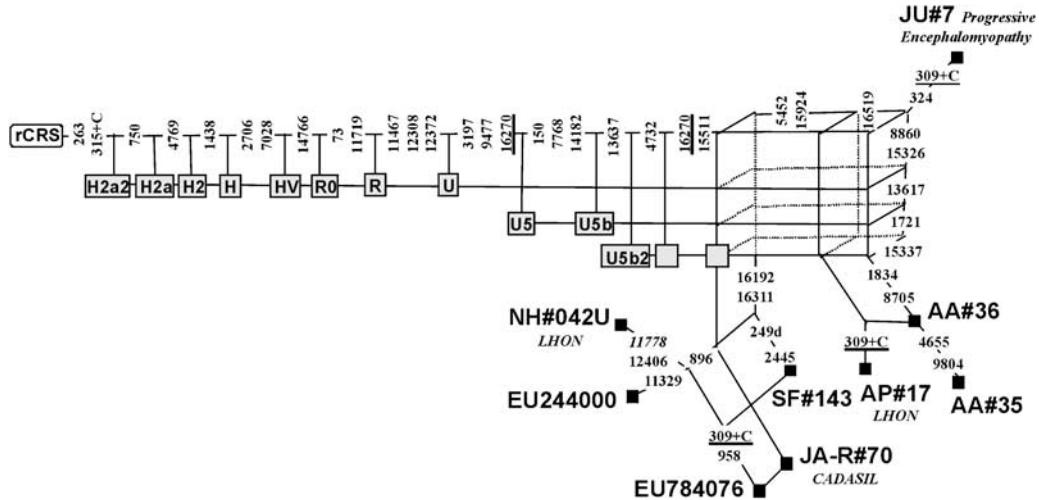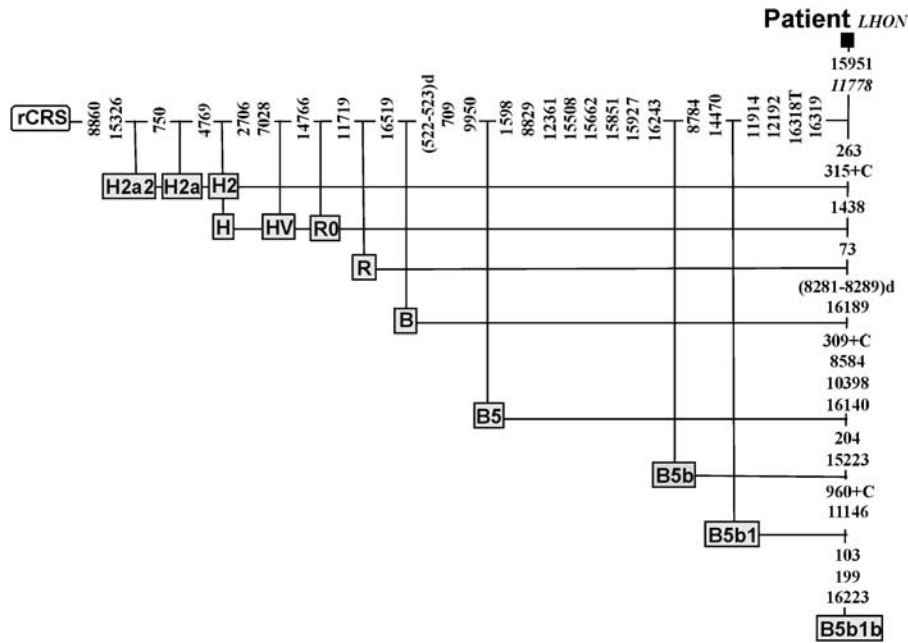

Figure 2 Median network displaying the evolutionary pathways between the rCRS and a patient's mtDNA sequence (JU no. 7) from Uusima *et al.*[49] together with related mtDNAs, namely SF no. 143 from Finnilä *et al.*,[50] NH no. 042U of a LHON patient from Howell *et al.*,[51] AA no. 35 and AA no. 36 from Achilli *et al.*,[25] AP no. 17 of a LHON patient from Puomila *et al.*,[52] and two further sequences from FTDNA deposited in GenBank (accession nos. EF244000 and EU784076). The latter mtDNA sequence differs only by one mutation from the coding-region sequence (JA-R no. 70) of a CADASIL patient from Annunen-Rasila *et al.*[74] the expected mutations in the control region have been hypothesized. The unreported states of indels at positions 309 and 315 in NH no. 042U were also hypothesized. Recurrent occurrences of the indel 309+C have been postulated beforehand. For symbols, see legend of Figure 1.

developed into the emerging tree of West Eurasian mtDNAs,[31] which set the agenda for subsequent studies of European mtDNAs.

Now, with a wealth of sequence information at hand, the early sequencing attempt of Rieder *et al.*[39] for example, appear in quite a critical light. We have selected two of their complete sequences, no. 3 and no. 12, for closer inspection (Figure 1). Sequence no. 3 is particularly problematic as it shows blocks of mutations that are associated with quite distant haplogroups, namely V and K1a3. The paths that connect the roots of K1a3 and V with the rCRS join at the root of haplogroup HV. Assigning the mutations observed in sequence

Figure 3 Median network representing the LHON mtDNA sequence from Mimaki et al.[60] and the evolutionary pathway between the rCRS and the root of haplogroup B5b1b. For symbols, see legend of Figure 1.

no. 3 to these two sub-paths, one sees that sequence no. 3 bears eight of the 21 mutations between K1a3 and HV, and at the same time, three of the four mutations between V and HV, plus the mutations between HV and the reference sequence. In addition, no. 3 carries three further unspecific control-region mutations (at positions 151, 152 and 16301), which could appear in either haplogroup (namely, these polymorphisms are not by themselves diagnostic for any basal haplogroup).

Similarly, the haplogroup T2b sequence no. 12 from Rieder et al.[39] lacks the transversion C15452A (characteristic of haplogroup JT) and four transitions (G1888A, G14905A, A15607G and G15928A) characteristic of haplogroup T. These features cannot be attributed to natural causes as is testified by numerous studies that do not show any evidence for such a large amount of concerted back (or parallel) mutations (for example, Herrnstadt et al.[40]). Instead, contamination, sample mix-up and possibly additional oversights of mutations are the more probable explanations for such a pattern. In fact, a single contamination or sample mix-up event (for example, one single PCR) would be nearly enough to create the mosaic sequence because four of these mutations belong to one common sequence fragment.

Furthermore, at least one case of a phantom mutation can be detected in the Rieder et al.[39] data as well. Namely, the otherwise unknown deletion 3916del was reported in as many as four out of 12 sequences, which can be allocated to haplogroups H3, H1c, H2a2 and T2b, respectively. Repeated occurrences of novel or very rare mutations on various branches of a tree inferred from a single mtDNA data set signpost biochemical problems with the electrophoresis.[3,10,13,41] The extensive grid-like structure of the median network representing sequences no. 3 and no. 12 together with the rCRS and the roots of haplogroups V, K1a3 and T2b (Figure 1) testifies to the mosaic pattern in the data. For this presentation, we disregarded any potential ambiguities incurred by an unknown reference sequence and we assumed by default that the mutations between rCRS and the root of haplogroup HV had been read correctly.

In a way similar to the instance presented in Figure 1, one can see that the haplogroup T2a sequence HCM P-9 from Ozawa[42–44] misses the mutations A11251G and C15452A (characteristic of haplogroup JT) as well as G13368A (characteristic of haplogroup T) and T13965C (characteristic of sub-haplogroup T2a); this sequence may in fact belong to a known (but unnamed) sub-haplogroup of T2a in view of the recorded mutation T2850C. The expected HVS-I mutation for T2, namely C16296T, is not present either, but this may very well constitute a natural back mutation as the mtDNA database testifies to such instances. Note that the later update of the sequence information in Ozawa[45] removed mutation C6521T as a potential correction but further dropped the A11812G mutation (characteristic of T2) by mistake (see Table 1). Whereas this sequence is thus obviously defective, the second mtDNA sequence, HCM P-8, of West Eurasian ancestry in the data set of Ozawa[42–45] seems to have been read perfectly: it bears all the mutations of haplogroup U8a1 and fits well into the current haplogroup U8a1 tree.[48]

In contrast to the rather early study of European mtDNA variation by Rieder et al.[39] the study by Uusima et al.[49] of the entire mitochondrial genome in 17 patients with mitochondrial encephalomyopathy resulted in mtDNA sequences that can well be accommodated to the current West Eurasian mtDNA tree—with one or two exceptions, however. First, it is clear that a partially corrected CRS still bearing the erroneous state at position 14766 was used consistently. The haplogroup J1c sequence of Patient 14 has the same nucleotide at position 7028 that is specific for haplogroup H, which would be quite unusual and might constitute an oversight.

The mtDNA of Patient 7 from Uusima et al.[49] clearly belongs to haplogroup U5b2 and, in particular, is related to eight mtDNA sequences;[25,50–52] GenBank sequences with accession nos. EU244000 and EU784076, which all share two specific coding-region mutations A4732G and T15511C. The median network generated from those nine sequences together with the pathway to rCRS (Figure 2) reveals a

Table 1 Comparison and evaluation of mutations in 10 mtDNAs from three papers by Ozawa and colleagues

Patient	Haplogroup[a]	Missed mutations			Mis-added mutations		
		Ozawa[42]	Ozawa[43,44]	Ozawa[45]	Ozawa[42]	Ozawa[43,44]	Ozawa[45]
HCM P-9	T2a	C15452A	C15452A	C15452A	C6521T	C6521T	
		A11251G	A11251G	A11251G			
		G13368A	G13368A	G13368A			
		G8697A	G8697A	G8697A			
		C16296T?	C16296T?	C16296T?			
				A11812G			
		T13965C	T13965C	T13965C			
HCM P-8	U8a1						
HCM P-3	B4a2	C15292T	C15292T		'C'15929T	'C'15929T	
HCM P-2	D4b1a	NA[b]	T15440C	T15440C	NA	C6455T	
			A14927G	A14927G			
MCM P-1	D4a1				C6455T	C6455T	
					C10202T	C10202T	
ID 119	D4a1	NA			NA	C6455T	'C'11447G
HCM P-10	M7b2	C4071T	C4071T	C4071T			
HCM P-5	M7a1b	T16324C?	T16324C?	T16324C?			
HCM P-1	M7a1a						
DCM P-2	M7a1	?	?	NA			NA

[a]Haplogroup allocation according to Kong *et al.*[23,46] and Tanaka *et al.*[47]
[b]NA=not applicable; variants with uncertainty are marked with a question mark.

number of incompatibilities that require some explanation. It is very likely that mutations T13617C and C1721T were overlooked since a natural back mutation at a pair of rather conservative positions is quite implausible; for example, in the combined data set of 518 sequences[41,53] there is no single recurrent change at positions 1721 and 13617. Finally, the role of T8705C remains obscure: Figure 1 of Uusima *et al.*[49] displays an empty row for this mutation, which in fact has been reported for the three closely related U5b2 sequences no. 35 and no. 36 from Achilli *et al.*[25] and no. 17 from Puomila *et al.*[52]

The mutation C5452T found in the Finnish LHON patient no. 17 of Puomila *et al.*[52] as well as Patient 7 from Uusima *et al.*[49] deemed to be pathogenic, is also recorded for the two U5b2 mtDNAs of an Italian male with fertility problems and a healthy Spanish control (no. 35 and no. 36 in Achilli *et al.*[25]). It is thus plausible that the transitions at 5452 and 15924 define a minor sub-haplogroup of U5b2. Therefore a direct involvement of C5452T in mitochondrial encephalomyopathy may not seem very likely, and verifying a secondary role would warrant further investigation of related haplogroup U5b2 mtDNAs. The fact that position 5452 showed some heteroplasmy in Patient 7 could also be interpreted in the way that the variant T at 5452 partially mutated back to C rather than the other way round. Possibly, there was some background noise in the sequence electropherogram (perhaps induced by contamination with DNA bearing the majority nucleotide at this position) that would then also be reflected by a few aberrant clones.

East Asian mtDNA sequences
The first (nearly) complete mtDNA sequences published by Ozawa and co-workers[43,44,54,55] had a considerable impact on our understanding of Eurasian mtDNA variation[56] and were discussed in the particular East Asian context by Kivisild *et al.*[57] where, unfortunately, the additional sequences from[55] were not integrated. A comparison with the haplogroup M10 sequence of the latter paper triggered a correction of an error in the original sequence YN163 of Kong *et al.*[46]

The two sequences from haplogroup F obtained by Sano *et al.*[58] are quite remarkable in regard to the sequence quality attained at the time: the mtDNA of Patient 1 bears all 13 known mutations distinguishing haplogroup F2 from the root of haplogroup R (10 mutations from F2 down to the root of R9 and then another three to the root of R) and, in addition, has G11150A, A13722G, C15714T, A16066G, C16192T, C16239T and C16355T as private mutations. This indicates that this mtDNA lineage may constitute a novel branch of haplogroup F2 (or F2a) not hitherto described.[59] The mtDNA of Patient 2 belongs to haplogroup F1b1a,[47] but lacks the mutations G16129A, T16189C, C16232A, T16311C and G14476A, likely because Sano *et al.*[58] listed only mutations that were 'infrequent' compared to controls.

The ten complete mtDNA sequences (besides the CRS) displayed by Ozawa,[43,44] eight of which were taken from Ozawa,[42] can be allocated to haplogroups T2a, B4a2, U8b, D4b1a, D4a1, D4a1, M7b2, M7a1b, M7a1a and M7a1, respectively (reading their diagram entries from left to right). Except for the two sequences (HCM P-8 and HCM P-9) of West Eurasian ancestry discussed above (Table 1), the remaining eight mtDNAs are typical members of Japanese haplogroups.[47] There are two obvious documentation errors in their diagram: the first concerns the mutation 'C15929T' in the B4a2 sequence (HCM P-3); in fact, the CRS has an A at 15929. As all B4a2 sequences from Tanaka *et al.*[47] bearing A9254G would also show C15292T, we infer that the correct number string '292' had inadvertently been inverted to '929'. The second problem is incurred by the wrong placement of the mutation C6455T in their diagram, which implied that the three sequences from two different branches of haplogroup D4 also had this mutation. These documentation errors were eventually corrected by Ozawa[45] where also the private mutation C10202T in MCM P1 was removed (Table 1).

The 11 mutations separating the four members of haplogroup R from the six members of haplogroup M were all reported by Ozawa[42–45] in his diagram. Thus, in view of the phylogenetic treat-

ment of the data, oversights of mutations were more likely to happen towards the periphery of the tree. Comparison with the more numerous sequences from Tanaka et al.[47] then suggests that mutations A14927G and T15440C were missed in the D4b1a sequence (HCM P-2) and C4071T in the M7b2 sequence (HCM P-10). The M7a1b sequence (HCM P-5) comes very close to sequence TC7 from Tanaka et al.[47] suggesting that the HVS-I mutation T16324C either had reverted naturally or was missed in HCM P-5. In conclusion, the complete mtDNA sequences as eventually displayed by Ozawa[45] had probably about one error per sequence on average. Furthermore, a seeming mutation, C11447G, slipped into sequence ID 119 from Ozawa[45] with a false reference nucleotide at 11447, which was evidently taken from a false 'corrected' reference sequence, such as the GenBank entry V00662. This reference sequence (or a related one) was also used by Hofmann et al.[38]

More recently, Mimaki et al.[60] claimed to have obtained the complete mtDNA sequence of a patient with LHON: this sequence appears to be closely related to the sequence published earlier by Shin et al.[61] because the latter covers all 29 nucleotide changes exhibited by Mimaki et al.[60] except for the (primary) LHON mutation G11778A and one further mutation (A15951G). The sequence from Shin et al.[61] was first displayed in its phylogenetic context in Figure 1 by Kivisild et al.[57] where the transversion at 16318 was misrepresented as a transition and the insertion of one C in the C stretch 955–960 could not be correctly identified (due to the limited information available at the time).

Phylogenetic assessment of these two sequences together with the related sequences JD33, JD58, KA81, ND56, ND179, TC14 and TC29 from the study of Tanaka et al.[47] all allocated to haplogroup B5b1b, reveals that at least the mutations A73G, A263G, 315+C, A1438G, (8281–8289)del (popularly known as the 9-bp deletion) and T16189C distinguishing the root of haplogroup B from rCRS and the mutations G8584A, A10398G and T16140C characteristic of haplogroup B5, the mutations T204C and C15223T characteristic of B5b, further 960+C and C11146T characteristic of B5b1 and G103A, T199C, 309+C and C16223T characteristic of B5b1b were likely all missed by Mimaki et al.[60] This can be traced in the median network of Figure 3 representing the particular LHON sequence together with the evolutionary path connecting the known sequences from haplogroup B5b1b with the rCRS. We conclude that more than one-third of all the variants expected in a B5b1b sequence relative to rCRS were actually left unrecorded by Mimaki et al.[60] Finally, the claim that 'the G12192A mutation caused cardiomyopathy as an additional symptom'[60] is not very convincing as this mutation has been found in different cohorts of patients and healthy individuals by Tanaka et al.,[47] namely in those seven mtDNAs belonging to haplogroup B5b1b and in three mtDNAs belonging to haplogroup G2a.

A recent study of Zhu et al.[62] has featured G7444A as a pathogenic mutation in aminoglycoside-induced and non-syndromic hearing loss — although G7444A occurs, for example, as a common variant within haplogroup V[17] and also recurrently on other haplogroup backgrounds; for example, in L1b.[53] Incidentally, this mutation has a long record since 1992 as a suspect for association with LHON[63–65] but fell out of favor already in the mid-nineties, probably because of its frequent occurrence within haplogroup V in Finland. Therefore, proving that G7444A serves as a secondary pathogenic mutation one would need to address this circumstance in a larger-scale study and not just through anecdotal findings. Zhu et al.[62] recorded this mutation in two patients, whose mtDNAs belong to haplogroups C4a1 (sample WZ201) and D4a (sample WZ202), respectively. Of seven mutations claimed to be 'novel', only one candidate (T6488C)

may really be new—which is not a rare phenomenon.[36,37] Besides misrepresentation of some nucleotide variants at positions 2226, 2706 and 4715, a considerable number of mutations were obviously overlooked: in WZ201 at positions 750, 2706, 7196, 11969, 14318, 14766, 14783 and 15204, and in WZ202 at 750, 3206, 14668 and 14766. The pattern and amount of likely errors in these data appear to be similar to those found in several earlier papers published by the same laboratory.[14,17,66]

An extremely incomplete listing of mutations can be seen in the family reported by Chen et al.[67] Misread nucleotides likely involve positions 490 and 751, with a +1 base shift (possibly triggered by the insertion 315+C). Then, of the 13 mutations listed, C298T and A13104G could point to haplogroup D4g2,[23] previously labeled as D4k3.[47,53] Further comparison with a possibly related D4g2 complete mtDNA sequence (PDsq0098) from Tanaka et al.[47] reveals that only one-third of the expected mutations have actually been detected. Alternatively, the single mutation G3421A might indicate haplogroup D4n membership. In any case, G3421A is thus not a novel mutation as claimed: besides its occurrence in haplogroup D4, it was previously observed once in a haplogroup L1c2 lineage.[40] This information can well be retrieved from the mtDB database (http://www.genpat.uu.se/mtDB/).

A most extreme case of incomplete sequencing constitutes the entire mtDNA from a patient's lymphocytes presented by Hattori et al.[68] Surprisingly, only a meager two homoplasmic mutations (C11215T and A15874G) and one heteroplasmic mutation (C3310T) were reported (as the 'three unique point mutations… in the protein-coding region'). The two homoplasmic mutations clearly indicate haplogroup D4e2 membership.[23] Thus many mutations, even in the protein-coding genes, must have been overlooked. On the other hand, a homoplasmic mutation C3310T was found by Starikovskaya et al.[69] on another haplogroup (A) background. It is then unclear whether C3310T is really pathogenic, and it cannot be excluded that a potentially pathogenic mutation might have been missed by the experimental assay of Hattori et al.[68]

DISCUSSION

Systematic comparison with the relevant mtDNA information available, mutation by mutation on known evolutionary pathways, is indispensable for putting freshly obtained complete mtDNA sequences into proper perspective. Even coarse phylogenetic screening could then quickly hit the target by highlighting a few complete mtDNA sequences as potentially related to the sequence under study. A detailed network analysis of these sequences together with the postulated pathway to the rCRS can then assist in pinpointing possibly discordant features of the published record and, more importantly, in discovering idiosyncratic features of the new sequence. To clarify the status of potentially missed mutations or putative phantom mutations, re-reading and re-sequencing parts of the mtDNA would be mandatory.

The results of some published complete sequencing efforts provide no more than rudimentary information and therefore lack any solid basis for inferring pathogenic status of a particular mutation. For the sake of comparison, it is instructive to count the number of reversed mutations along single pathways to the rCRS relative to the known mtDNA phylogeny, as inferred from systematic studies of mtDNA variation. In the East Eurasian data of Kong et al.[46] we are seeing on average 0.10 reversals of control-region mutations (hitting positions 146, 263, 16223 or 16304, but disregarding the 16519 polymorphism and length polymorphisms of the C stretches within regions 16184–16193 and 303–315 and the CA repeats in region 514–523, because recurrent mutations are extremely frequent for these polymorphisms)

180

along the inferred paths connecting the sequences each with the rCRS; for coding-region mutations the corresponding averaged value is even lower, namely 0.04 (reflecting two reversals at position 1438). The corresponding values for the mtDNAs of West Eurasian and South Asian ancestry sequenced by Palanichamy et al.[22] are 0.09 for the control region (involving positions 16234, 16266, 16292 and 16309) and 0.12 for the coding region (involving positions 2706, 4769, 8860 and 15326). Therefore, we would expect that with a future more fine-grained mtDNA tree no more than about 0.2 coding-region mutations per sequence would typically revert along the reconstructed pathway to the rCRS in the case of Eurasian mtDNA lineages.

In contrast, the number of reversed mutations observed in the mtDNA studies we have reanalyzed here are generally by one or two orders of magnitude larger than the naturally expected value, thus unmistakably pointing to incomplete sequencing. In such a situation, the chances are >5% that a true pathogenic mutation had actually been overlooked. If that really happened, then instead a rather innocent mutation defining a minor haplogroup might come into suspicion for pathogenicity. The clusters of mutations that would generate a pathogenic phenotype only in concert—but not separately—could go unnoticed because of incomplete sequencing. For example, G7444A was observed altogether three times[62,70] and in two instances together with A1811G. Suppose that the latter mutation was overlooked in the third instance, then a strong case could have been made for 'cosegregation' of this mutational pair (because position 1811 does not mutate frequently). The A1811G mutation is known to be a basal polymorphism in haplogroup U, and also the G7444A mutation might be an infrequent normal polymorphism, at least in the Brazilian population of mixed continental mtDNA ancestries.[71]

The complete mtDNA sequences obtained by Ozawa[45] that gradually evolved from the pioneering sequencing attempts of Yoneda et al.[72] and Ozawa et al.[54,55] came amazingly close to correct complete sequences compared to what is usually offered by contemporary sequencing attempts in medical genetics. It is instructive to learn where the few errors in Ozawa et al.'s sequences[42–45] (Table 1) are located: most of them cluster in a European haplogroup T2a sequence, appearing to be an absolute outlier in their Japanese mtDNA data set. On the other hand, the three members of the well represented haplogroup M7a1 can be regarded as error free. This obviously was the result of a phylogenetic approach that assisted the proofreading of the sequences (cf., their figure for the phylogenetic clustering). As a phylogenetic approach is no longer exercised by routine application of total mtDNA sequencing in medical genetics (notwithstanding exceptions such as the study by Hinttala et al.[73]), the sequencing results are therefore typically of rather poor quality.

CONCLUSION

Phylogenetic bookkeeping of mutations is an essential prerequisite for mtDNA disease studies that should not be missed out. Searching only for some key mutations that would allow gross allocation to major haplogroups does not yet shield against considerable mutation oversights (as for example, in the two mtDNA instances reported by Zhu et al.[62]). On the other hand, casual phylogenetic analysis could very well let incomplete sequences invade the study of entire mtDNA genomes; for example, the present MITOMAP tree incorporated the problematic sequences from Ozawa et al.,[43,54,55] Mimaki et al.[60] and Uusimaa et al.[49] discussed above. Employing our data mining strategies and network visualization tools could then help improving the quality of complete mtDNA sequences as well as avoiding premature conclusions regarding the pathogenicity status of a mutation deemed to be 'novel' (see Bandelt et al.[37]).

1   Bandelt, H.-J., Salas, A. & Bravi, C. Problems in FBI mtDNA database. Science 305, 1402–1404 (2004a).
2   Bandelt, H.-J., Salas, A. & Lutz-Bonengel, S. Artificial recombination in forensic mtDNA population databases. Int. J. Legal. Med. 118, 267–273 (2004b).
3   Bandelt, H.-J., Kong, Q.-P., Parson, W. & Salas, A. More evidence for non-maternal inheritance of mitochondrial DNA? J. Med. Genet. 42, 957–960 (2005b).
4   Salas, A., Carracedo, Á., Macaulay, V., Richards, M. & Bandelt, H.-J. A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. Biochem. Biophys. Res. Commun. 335, 891–899 (2005a).
5   Salas, A., Yao, Y.-G., Macaulay, V., Vega, A., Carracedo, Á. & Bandelt, H.-J. A critical reassessment of the role of mitochondria in tumorigenesis. PLoS Med. 2, e296 (2005b).
6   Salas, A., Bandelt, H.-J., Macaulay, V. & Richards, M. B. Phylogeographic investigations: The role of trees in forensic genetics. Forensic. Sci. Int. 168, 1–13 (2007).
7   Yao, Y.-G., Bandelt, H.-J. & Young, N. S. External contamination in single cell mtDNA analysis. PLoS ONE 2, e681 (2007).
8   Kong, Q.-P., Salas, A., Sun, C., Yao, Y.-G., Fuku, N., Tanaka, M. et al. Distilling artificial recombinants from large sets of complete mtDNA genomes. PLoS ONE 3, e3016 (2008).
9   Bandelt, H.-J. & Kivisild, T. Quality assessment of DNA sequence data: autopsy of a mis-sequenced mtDNA population sample. Ann. Hum. Genet. 70, 314–326 (2006).
10 Bandelt, H.-J., Quintana-Murci, L., Salas, A. & Macaulay, V. The fingerprint of phantom mutations in mtDNA data. Am. J. Hum. Genet. 71, 1150–1160 (2002).
11 Bandelt, H.-J., Yao, Y.-G., Salas, A., Kivisild, T. & Bravi, C. M. High penetrance of sequencing errors and interpretative shortcomings in mtDNA sequence analysis of LHON patients. Biochem. Biophys. Res. Commun. 352, 283–291 (2007).
12 Herrnstadt, C., Preston, G. & Howell, N. Errors, phantom and otherwise, in human mtDNA sequences. Am. J. Hum. Genet. 72, 1585–1586 (2003).
13 Brandstätter, A., Sänger, T., Lutz-Bonengel, S., Parson, W., Béraud-Colomb, E., Wen, B. et al. Phantom mutation hotspots in human mitochondrial DNA. Electrophoresis 26, 3414–3429 (2005).
14 Bandelt, H.-J., Achilli, A., Kong, Q.-P., Salas, A., Lutz-Bonengel, S., Sun, C. et al. Low 'penetrance' of phylogenetic knowledge in mitochondrial disease studies. Biochem. Biophys. Res. Commun. 333, 122–130 (2005a).
15 Yao, Y.-G., Macaulay, V., Kivisild, T., Zhang, Y.-P. & Bandelt, H.-J. To trust or not to trust an idiosyncratic mitochondrial data set. Am. J. Hum. Genet. 72, 1341–1346, 1348–1349 (reply) (2003).
16 Yao, Y.-G., Bravi, C. M. & Bandelt, H.-J. A call for mtDNA data quality control in forensic science. Forensic. Sci. Int. 141, 1–6 (2004).
17 Yao, Y.-G., Salas, A., Bravi, C. M. & Bandelt, H.-J. A reappraisal of complete mtDNA variation in East Asian families with hearing impairment. Hum. Genet. 119, 505–515 (2006).
18 Bandelt, H.-J., Kivisild, T., Parik, J., Villems, R., Bravi, C., Yao, Y.-G. et al. Lab-specific mutation processes In: Bandelt H-J, Macaulay V, Richards M (eds). Human Mitochondrial DNA and the Evolution of Homo sapiens, Springer-Verlag: Berlin-Heidelberg, pp 119–150 (2006a).
19 Torroni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H.-J. Harvesting the fruit of the human mtDNA tree. Trends. Genet. 22, 339–345 (2006).
20 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147 (1999).
21 Bandelt, H.-J., Macaulay, V. & Richards, M. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol. Phylogenet. Evol. 16, 8–28 (2000).
22 Palanichamy, M. g., Sun, C., Agrawal, S., Bandelt, H.-J., Kong, Q.-P., Khan, F. et al. Phylogeny of mtDNA macrohaplogroup N in India based on complete sequencing: implications for the peopling of South Asia. Am. J. Hum. Genet. 75, 966–978 (2004).
23 Kong, Q.-P., Bandelt, H.-J., Sun, C., Yao, Y.-G., Salas, A., Achilli, A. et al. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum. Mol. Genet. 15, 2076–2086 (2006).
24 Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R. et al. 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 (2004).
25 Achilli, A., Rengo, C., Battaglia, V., Pala, M., Olivieri, A., Fornarino, S. et al. Saami and Berbers—an unexpected mitochondrial DNA link. Am. J. Hum. Genet. 76, 883–886 (2005).
26 Loogväli, E.-L., Roostalu, U., Malyarchuk, B. A., Derenko, M. V., Kivisild, T., Metspalu, E. et al. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol. Biol. Evol. 21, 2012–2021 (2004).
27 Behar, D. M., Metspalu, E., Kivisild, T., Achilli, A., Hadid, Y., Tzur, S. et al. The matrilineal ancestry of Ashkenazi jewry: portrait of a recent founder event. Am. J. Hum. Genet. 78, 487–497 (2006).

28 Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M. et al. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. Am. J. Hum. Genet. **81,** 1025–1041 (2007).

29 Roostalu, U., Kutuev, I., Loogväli, E. L., Metspalu, E., Tambets, K., Reidla, M. et al. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. Mol. Biol. Evol. **24,** 436–448 (2007).

30 Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J. et al. Sequence and organization of the human mitochondrial genome. Nature **290,** 457–465 (1981).

31 Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V. et al. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am. J. Hum. Genet. **64,** 232–249; **64,** 918 (erratum) (1999a).

32 Macaulay, V., Richards, M. & Sykes, B. Mitochondrial DNA recombination—no need to panic. Proc. R. Soc. London. Ser. B. Biol. Sci. **266,** 2037–2039 (1999b).

33 Bandelt, H.-J., Forster, P., Sykes, B. C. & Richards, M. B. Mitochondrial portraits of human populations using median networks. Genetics **141,** 743–753 (1995).

34 Montiel-Sosa, F., Ruiz-Pesini, E., Enríquez, J. A., Marcuello, A., Díez-Sánchez, C., Montoya, J. et al. Differences of sperm motility in mitochondrial DNA haplogroup U sublineages. Gene **368,** 21–27 (2006).

35 Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A. & Tishkoff, S. A. Whole-mtDNA genome sequence analysis of ancient African lineages. Mol. Biol. Evol. **24,** 757–768 (2007).

36 Bandelt, H.-J., Salas, A. & Bravi, C. M. What is a 'novel' mtDNA mutation—and does 'novelty' really matter? J. Hum. Genet. **51,** 1073–1082 (2006b).

37 Bandelt, H.-J., Salas, A., Taylor, R. W. & Yao, Y.-G. The exaggerated status of 'novel' and 'pathogenic' mtDNA sequence variants due to inadequate database searches. Hum. Mutat. **30,** 191–196 (2009).

38 Hofmann, S., Jaksch, M., Bezold, R., Mertens, S., Aholt, S., Paprotta, A. et al. Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlations with D loop variants and association with disease. Hum. Mol. Genet. **6,** 1835–1846 (1997).

39 Rieder, M. J., Taylor, S. L., Tobe, V. O. & Nickerson, D. A. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic. Acids. Res. **26,** 967–973 (1998).

40 Herrnstadt, C., Elson, J. L., Fahy, E., Preston, G., Turnbull, D. M., Anderson, C. et al. 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) (2002).

41 Coble, M. D., Just, R. S., O'Callaghan, J. E., Letmanyi, I. H., Peterson, C. T., Irwin, J. A. et al. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int. J. Legal. Med. **118,** 137–146 (2004).

42 Ozawa, T. Mitochondrial cardiomyopathy. Herz **19,** 105–118 (1994).

43 Ozawa, T. Mitochondrial DNA mutations in myocardial diseases. Eur. Heart J. **16** (Suppl O), 10–14 (1995a).

44 Ozawa, T. Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. Biochem. Biophys. Acta. **1271,** 177–189 (1995b).

45 Ozawa, T. Genetic and functional changes in mitochondria. Physiol. Rev. **77,** 425–464 (1997).

46 Kong, Q.-P., Yao, Y.-G., Sun, C., Bandelt, H.-J., Zhu, C.-L. & Zhang, Y.-P. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am. J. Hum. Genet. **73,** 671–676; **75,** 157 (erratum) (2003).

47 Tanaka, M., Cabrera, V. M., González, A. M., Larruga, J. M., Takeyasu, T., Fuku, N. et al. Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res. **14,** 1832–1850 (2004).

48 González, A. M., García, O., Larruga, J. M. & Cabrera, V. M. The mitochondrial lineage U8a reveals a Paleolithic settlement in the Basque country. BMC Genomics **7,** e124 (2006).

49 Uusimaa, J., Finnilä, S., Remes, A. M., Rantala, H., Vainionpää, L., Hassinen, I. E. et al. Molecular epidemiology of childhood mitochondrial encephalomyopathies in a Finnish population: sequence analysis of entire mtDNA of 17 children reveals heteroplasmic mutations in tRNA^{Arg}, tRNA^{Glu}, and tRNA$^{Leu(UUR)}$ genes. Pediatrics **114,** 443–450 (2004).

50 Finnilä, S., Lehtonen, M. S. & Majamaa, K. Phylogenetic network for European mtDNA. Am. J. Hum. Genet. **68,** 1475–1484 (2001).

51 Howell, N., Oostra, R. J., Bolhuis, P. A., Spruijt, L., Clarke, L. A., Mackey, D. A. et al. Sequence analysis of the mitochondrial genomes from Dutch pedigrees with Leber hereditary optic neuropathy. Am. J. Hum. Genet. **72,** 1460–1469 (2003).

52 Puomila, A., Hämäläinen, P., Kivioja, S., Savontaus, M.-L., Koivumäki, S., Huoponen, K. et al. Epidemiology and penetrance of Leber hereditary optic neuropathy in Finland. Eur. Hum. Genet. **15,** 1079–1089 (2007).

53 Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K. et al. The role of selection in the evolution of human mitochondrial genomes. Genetics **172,** 373–387 (2006).

54 Ozawa, T., Tanaka, M., Ino, H., Ohno, K., Sano, T., Wada, Y. et al. 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 (1991a).

55 Ozawa, T., Tanaka, M., Sugiyama, S., Ino, H., Ohno, K., Hattori, K. et al. Patients with idiopathic cardiomyopathy belong to the same mitochondrial DNA gene family of Parkinson's disease and mitochondrial encephalomyopathy. Biochem. Biophys. Res. Commun. **177,** 518–525 (1991b).

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

57 Kivisild, T., Tolk, H.-V., Parik, J., Wang, Y., Papiha, S. S., Bandelt, H.-J. et al. The emerging limbs and twigs of the East Asian mtDNA tree. Mol. Biol. Evol. **19,** 1737–1751; **20,** 162 (erratum) (2002).

58 Sano, T., Ban, K., Ichiki, T., Kobayashi, M., Tanaka, M., Ohno, K. et al. Molecular and genetic analyses of two patients with Pearson's marrow-pancreas syndrome. Pediatr. Res. **34,** 105–110 (1993).

59 Kong, Q.-P., Yao, Y.-G., Sun, C., Zhu, C.-L., Zhong, L., Wang, C.-Y. et al. 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 (2004).

60 Mimaki, M., Ikota, A., Sato, A., Komaki, H., Akanuma, J., Nonaka, I. et al. A double mutation (G11778A and G12192A) in mitochondrial DNA associated with Leber's hereditary optic neuropathy and cardiomyopathy. J. Hum. Genet. **48,** 47–50 (2003).

61 Shin, W. S., Tanaka, M., Suzuki, J., Hemmi, C. & Toyo-Oka, T. A novel homoplasmic mutation in mtDNA with a single evolutionary origin as a risk factor for cardiomyopathy. Am. J. Hum. Genet. **67,** 1617–1620 (2000).

62 Zhu, Y., Qian, Y., Tang, X., Wang, J., Yang, L., Liao, Z. et al. Aminoglycoside-induced and non-syndromic hearing loss is associated with the G7444A mutation in the mitochondrial COI/tRNA$^{Ser(UCN)}$ genes in two Chinese families. Biochem. Biophys. Res. Commun. **342,** 843–850 (2006).

63 Brown, M. D., Voljavec, A. S., Lott, M. T., Torroni, A., Yang, C.-C. & Wallace, D. C. Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. Genetics **130,** 163–173 (1992a).

64 Brown, M. D., Yang, C.-C., Trounce, I., Torroni, A., Lott, M. T. & Wallace, D. C. A mitochondrial DNA variant, identified in Leber hereditary optic neuropathy patients, which extends the amino acid sequence of cytochrome c oxidase subunit I. Am. J. Hum. Genet. **51,** 378–385 (1992b).

65 Huoponen, K., Lamminen, T., Juvonen, V., Aula, P., Nikoskelainen, E. & Savontaus, J. L. The spectrum of mitochondrial DNA mutations in families with Leber hereditary optic neuroretinopathy. Hum. Genet. **92,** 379–384 (1993).

66 Wang, C.-Y., Kong, Q.-P., Yao, Y.-G. & Zhang, Y.-P. mtDNA mutation C1494T, haplogroup A, and hearing loss in Chinese. Biochem. Biophys. Res. Commun. **348,** 712–715 (2006).

67 Chen, F. L., Liu, Y., Song, X. Y., Hu, H. Y., Xu, H. B., Zhang, X. M. et al. A novel mitochondrial DNA missense mutation at G3421A in a family with maternally inherited diabetes and deafness. Mutat. Res. **602,** 26–33 (2006).

68 Hattori, Y., Takeoka, M., Nakajima, K., Ehara, T. & Koyama, M. A heteroplasmic mitochondrial DNA 3310 mutation in the ND1 gene in a patient with type 2 diabetes, hypertrophic cardiomyopathy, and mental retardation. Exp. Clin. Endocrinol. Diabetes **113,** 318–323 (2005).

69 Starikovskaya, E. B., Sukernik, R. I., Derbeneva, O. A., Volodko, N. V., Ruiz-Pesini, E., Torroni, A. et al. 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 (2005).

70 Yuan, H., Qian, Y., Xu, Y., Cao, J., Bai, L., Shen, W. et al. Cosegregation of the G7444A mutation in the mitochondrial COI/tRNA$^{Ser(UCN)}$ genes with the 12S rRNA A1555G mutation in a Chinese family with aminoglycoside-induced and nonsyndromic hearing loss. Am. J. Med. Genet. **138A,** 133–140 (2005).

71 Abreu-Silva, R. S., Lezirovitz, K., Braga, M. C. C., Spinelli, M., Pirana, S., Della-Rosa, V. A. et al. Prevalence of the A1555G (12S rRNA) and tRNASer(UCN) mitochondrial mutations in hearing-impaired Brazilian patients. Braz. J. Med. Biol. Res. **39,** 219–226 (2006).

72 Yoneda, M., Tanno, Y., Horai, S., Ozawa, T., Miyatake, T. & Tsuji, S. A common mitochondrial DNA mutation in the t-RNA(Lys) of patients with myoclonus epilepsy associated with ragged-red fibers. Biochem. Int. **21,** 789–796 (1990).

73 Hinttala, R., Smeets, R., Moilanen, J. S., Ugalde, C., Uusimaa, J., Smeitink, J. A. et al. Analysis of mitochondrial DNA sequences in children with isolated or combined oxidative phosphorylation system deficiency. J. Med. Genet. **43,** 881–886 (2006).

74 Annunen-Rasila, J., Finnilä, S., Mykkänen, K., Moilanen, J. S., Veijola, J., Pöyhönen, M. et al. Mitochondrial DNA sequence variation and mutation rate in patients with CADASIL. Neurogenetics **7,** 185–194 (2006).