

DATABASES

Exaggerated Status of “Novel” and “Pathogenic” mtDNA Sequence Variants Due to Inadequate Database Searches

Hans-Jürgen Bandelt,¹ Antonio Salas,² Robert W. Taylor,³ and Yong-Gang Yao^{4,5*}

¹Department of Mathematics, University of Hamburg, Hamburg, Germany; ²Unidade de Xenética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, Galicia, Spain; ³Mitochondrial Research Group, The Medical School, Newcastle University, Newcastle upon Tyne, United Kingdom; ⁴Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; ⁵State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

Communicated by A. Jamie Cuticchia

Given its relative ease, screening the entire mitochondrial DNA (mtDNA) for heteroplasmic or novel homoplasmic mutations has become part of the routine diagnostic workup for the molecular geneticist confronted with a disease case exhibiting clinical and biochemical features of mitochondrial dysfunction. “Novelty” of a given mtDNA variant is most often equated with nonregistration in the extensive MITOMAP database (www.mitomap.org). This practice has led to a number of spurious findings and wrong conclusions concerning the pathogenic status of specific mtDNA mutations, especially in the absence of proper evaluation and pathogenicity scoring. We demonstrate by way of real cases targeting the mt-tRNA^{Cys} (MT-TC) gene and a stretch within the MT-ND3 gene, that a straightforward Google search can identify twice as many previously observed mutations than any MITOMAP query could achieve. Further, we reassess the recent rediscovery of m.15287T>C by listing all known occurrences and, where possible, providing the haplogroup context, shedding new light on the potential pathogenicity status of m.15287T>C. *Hum Mutat* 0,1–7, 2008. © 2008 Wiley-Liss, Inc.

KEY WORDS: MITOMAP; Google; mtDNA polymorphism; pathogenic mutation; MT-ND3; MT-TC

INTRODUCTION

The mitochondrial DNA (mtDNA) is limited in size, therefore—after two decades of intensive mtDNA research—the discovery of really new, hitherto unobserved mtDNA mutations should have become a somewhat less frequent occurrence. Nevertheless, the number of mtDNA mutations with claimed “novelty” seem to grow considerably. This is mainly due to the fact that “novelty” status is exaggerated, so that an infrequent mutation may be discovered *de novo* and consequently regarded as “pathogenic” multiple times, without proper functional analysis. Despite the availability of many online resources, a number of scholars of human mitochondrial genetic disorders still confine their search to a single Web resource—the MITOMAP database (www.mitomap.org) [Brandon et al., 2005], without recourse to a large number of complete mtDNAs from published sources (that are not necessarily indexed by MITOMAP). The phylogenetic content of >4,300 mtDNA entire coding-region sequences, mainly (but not exclusively) drawn from healthy individuals worldwide, is then often largely left untapped. However, as emphasized in Bandelt et al. [2006b], it would take only a few Internet searches to obtain pertinent information about any targeted variation.

There are many lacunae in the MITOMAP presentation of mtDNA variation [Bandelt et al., 2006b, 2008]. As an extreme example, the deletion 2395del in the MT-RNR2 gene (MIM#

561010), which is characteristic of the major African haplogroup L1c [Quintana-Murci et al., 2008], is represented by >1% of the lineages in the mtDB database (www.genpat.uu.se/mtDB) as well as on Ian Logan’s website (www.ianlogan.co.uk/mtDNA.htm). Thus, this mutation would be hard to miss when analyzing the entire mtDNA of, say, individuals from the U.S. Nonetheless, this mutation is not yet registered in MITOMAP, although the huge MITOMAP tree (www.mitomap.org/mitomap-phylogeny.pdf) includes this deletion along the branch to L1c. One cannot expect, however, that a clinician with little interest or insight into the mtDNA phylogeny would take notice of this tree, let alone consult this tree in detail. On the other hand, the information recorded in that tree is not fully reliable and would need cross-comparisons with the original sources [Bandelt et al., 2008]. For example, the

Received 6 April 2008; accepted revised manuscript 22 May 2008.

Grant sponsors: Wellcome Trust; Newcastle upon Tyne Hospitals Foundation, National Health Service (NHS) Trust; NHS National Commissioning Group (NCG) Rare Mitochondrial Disorders of Adults and Children; Chinese Academy of Sciences.

*Correspondence to: Yong-Gang Yao, Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, 650223 China. E-mail: ygyaozh@yahoo.com

DOI 10.1002/humu.20846

Published online in Wiley InterScience (www.interscience.wiley.com).

haplogroup L2 mutation m.2416T>C has also been shunted into a specific haplogroup L0a2a lineage by misediting of the information provided in Figure 2 of Torroni et al. [2001]. It seems that supplementary Internet searches may help to balance the information retrieved from MITOMAP.

In this article, we show, using specific examples, how systematic literature searches and phylogenetic considerations can guide the judgment of potential pathogenicity for mtDNA mutations and pave the ground for subsequent focused studies and functional analyses.

METHODS AND RESULTS

Variation in a Stretch of the *MT-ND3* Gene (MIM# 516002)

The recent rediscovery of the m.10197G>A mutation [Chae et al., 2007] and its iterated labeling as “novel” [Sarzi et al., 2007] constitutes a paradigmatic case. m.10197G>A has long since been known as a rare private mutation, possibly first discovered in an African haplogroup L3h lineage (GenBank accession number AF347000), published in 2000 [Ingman et al., 2000]; see Bandelt et al. [2006a] for a phylogenetic display of Ingman et al. [2000] data. Moreover, this mutation also occurs in a haplogroup V lineage (GenBank accession number AY195781), published in 2003 [Mishmar et al., 2003]. Then, in 2004, this mutation was first mentioned in the context of human mitochondrial disease [Kirby et al., 2004].

The mutation m.10197G>A was originally dismissed as a pathogenic mutation because it was seemingly shared by a whole cohort of 22 patients of Polynesian origin whose mtDNAs bore the so-called Polynesian motif (haplogroup B) [Kirby et al., 2004]. Using current terminology, these lineages were thus asserted to belong to the “16247 branch” of haplogroup B4a1a1, which is characterized by m.14022A>G [Trejaut et al., 2005]. The fact that m.12239C>T and m.15746A>G (both signifying haplogroup B4a1a) were also observed in all those mtDNA lineages further confirmed this haplogroup affiliation. However, none of the currently available 12 complete mtDNA sequences having the m.14022A>G and m.16247A>G variants bear m.10197G>A, which can all be retrieved from the human mitochondrial genome database mtDB. Thus, with the phylogenetic knowledge of today, the findings of Kirby et al. [2004] would warrant re-examination. In fact, the positive signal for m.10197G>A in 21 out of 22 Polynesian samples turned out to be a PCR-RFLP artifact (David Thorburn, personal communication).

Interestingly, MITOMAP queried for “G10197A” does not provide any information but does for “10197G>A” (in the notation following the HGVS guidelines; www.hgvs.org/mutnomen/refseq.html#mtDNA) or “10197” (i.e., without the nucleotide circumfix). Three references are reported for “10197” but only a subset of two for “10197G>A.” This would suggest that there is not a fully consistent recording of mtDNA mutations in MITOMAP yet (see also below). MITOMAP classifies m.10197G>A as a “confirmed” Leigh dystonia mutation, listing later papers [Chae et al., 2007; Sarzi et al., 2007] for reference but not the earlier sources. Examples of such patchy citation outcomes have been reported before concerning MITOMAP searches [Bandelt et al., 2006b]. The unbalanced citation scheme can also be seen in context with the mutations m.10197G>A, m.12239C>T, and m.15746A>G discussed by Kirby et al. [2004]. The reference to Sarzi et al. [2007] is given in all three cases, although the latter two mutations get only mentioned through citation of the Kirby et al. [2004] work and the statement

that those two mutations were not observed in the three families under investigation [Sarzi et al., 2007]. However, the original work by Kirby et al. [2004] is not cited on MITOMAP. On the other hand, searching with “15746” also retrieves a reference to Brandstätter et al. [2005], in which this mutation does not occur at all. Note that queries of the form “12239C>T” or “15746A>G” do not meet success.

Entering the query “G10197A mtDNA” to Google yields five hits, viz. to Chae et al. [2007] (only through MITOMAP), Kirby et al. [2004], Sarzi et al. [2007], and Mishmar et al. [2003] (twice). In contrast, the query “10197 mtDNA” results in 152 multiple citations, whereas the query “10197G>A” is somewhat more specific, providing only 61 multiple hits. These queries were all performed on 20 March 2008. Now, two months later, at revision, we retrieved more than 600 and 400 citations, respectively. For a new Google search with “G10197A mtDNA” we were provided with two additional references, one to Ingman et al. [2000] (through Ian Logan’s website) and another one to Wang et al. [2006]. Thus, it clearly pays off to update searches for specific mutations of interest.

Google searches slightly depend on the versions used; country-specific Google engineering can yield different outcomes for the same query [Bandelt et al., 2006b]. Moreover, one should be aware of the syntactic rules of Google searches that would essentially give the same results for queries “10197G>A mtDNA,” “10197G > A mtDNA” (with spacing), “10197G A mtDNA,” and “10197G mtDNA.” These queries then provide further references: one review article [Wong, 2007] and two research abstracts [Steffann et al., 2007; Tchikviladzé et al., 2007]. Google Scholar searches do not provide any further information: with “G10197A mtDNA” we are led to Chae et al. [2007] and Kirby et al. [2004]; with “10197G>A mtDNA” we get Chae et al. [2007], Sarzi et al. [2007], and Tchikviladzé et al. [2007], whereas the somewhat too unspecific query “10197 mtDNA” is not useful in the case of Google Scholar.

To test how efficient a straightforward Google search is in detecting publications that report some (partial) variation, we entered queries of the type “G10197A mtDNA” for all possible mutations in a 33-bp fragment centered at 10197. We distinguished whether the Google search led to mtDNA lineages from population genetics databases or from some disease studies. Moreover, we checked the mtDB and MITOMAP for corresponding information. In the case of MITOMAP, we entered a query in two ways, with and without the circumfix of nucleotides, such as “G10197A” and alternatively as “10197.” As can be seen from Table 1 (summarizing the search results from 20 March 2008), MITOMAP is only half as effective compared to Google in tracing publications mentioning a particular mutation. When taking the mtDB and MITOMAP databases together as the only sources of information, 4 out of 17 mutations remain undiscovered, whereas the standard Google search misses just a single instance.

Variation in the mt-tRNA^{Cys} (*MT-TC*) Gene (MIM# 590020)

The mt-tRNA^{Cys} (*MT-TC*) gene (5761–5826), encoded on the H-strand, bears more than two dozen variants (Table 2), most of which occur in the general population, as testified by mtDNA population studies. Since the gene is encoded on the H-strand, there is a risk of confusion whether the nucleotides are read with respect to the L-strand or the H-strand. This, in particular, concerns some of the “popular” mutations such as “T5814C,” which is more often recorded as “A5814G”: a Google search for

TABLE 1. Variation in a 33-bp Fragment (10181–10213) of the *MT-ND3* Gene Seen Through the Lenses of Web-Based Searches*

Variant for query	Google: population context	Google: disease context	mtDB	MITOMAP
C10181T	Yes	Yes	Yes	Yes
C10184T	Yes	No	No	No
C10184G	No	No	No	Yes
T10187C	Yes	No	Yes	No
A10188G	Yes	No	No	No ^a
A10190G	Yes	No	Yes	No ^a
T10191C	No	Yes	No	Yes
C10192T	Yes	Yes	Yes	Yes
C10192A	Yes	No	Yes	No ^a
C10196T	Yes	No	No	No
G10197A	Yes	Yes	Yes	Yes
G10197C	No	Yes	No	Yes
C10199T	Yes	No	Yes	No ^a
C10202T	Yes	No	Yes	No ^a
G10203A	No	Yes	No	No
T10208C	Yes	No	Yes	No ^a
C10211T	Yes	Yes	Yes	Yes

*The variants were listed in a format for Google search, e.g., mutation C10181 T should be presented as 10181C>T and m.10181C>T according to the "traditional" and "approved" formats for mutation nomenclature, respectively (HGVS nomenclature guidelines: www.hgvs.org/mutnomen/refseq.html#mtDNA). For mtDB query, one only needs the position, e.g., query 10197 for mutation m.10197G>A. For MITOMAP query, we performed all kinds of queries as exemplified by mutation m.10197G>A in the text. The searches were performed on 20 March 2008.

^aInaccessible through the query option of MITOMAP but included in the MITOMAP tree.

TABLE 2. Variation in the 66-bp *MT-TC* Gene (5761–5826) Through the Lenses of Web-Based Searches*

Variant for query	Google: population context	Google: disease context	mtDB	MITOMAP
G5772A	Yes	No	Yes	No ^a
G5773A	Yes	Yes	Yes	Yes
T5774C	Yes	Yes	Yes	Yes
T5774A	Yes	No	Yes	No ^a
G5777A	Yes	No	Yes	No ^a
G5780A	No	Yes	No	Yes
G5783A	Yes	Yes	Yes	Yes
T5785C	No ^b	No	No	Yes
T5786C	Yes	No	Yes	No ^a
T5788C	Yes	No	Yes	Yes
C5790A	Yes	No	Yes	No
T5794C	No	Yes	No	No
A5805G	Yes	No	No	No ^a
T5806C	Yes	No	Yes	Yes
G5809A	Yes	No	Yes	No ^a
A5811G	Yes	No	Yes	Yes
A5813G	Yes	No	No	No ^a
T5814C	Yes	Yes	Yes	Yes
T5814A	Yes	No	Yes	No ^a
A5816G	No	Yes	No	Yes
C5817T	Yes	No	Yes	No
G5820A	No	Yes	No	Yes
G5821A	Yes	Yes	Yes	Yes
G5822A	Yes	No	Yes	No ^a
A5823G	Yes	Yes	Yes	No ^a
G5824A	Yes	No	Yes	Yes
T5826C	Yes	No	Yes	Yes

*The searches were performed on 20 March 2008; see the note in Table 1 for additional information regarding the mutation nomenclature and the format for mutation query.

^aInaccessible through the query option of MITOMAP but included in the MITOMAP tree.

^bRetrievable via the query "5785T>C mtDNA."

the former gave 90 results whereas a Google search for the latter gave as many as 216 results. The corresponding two queries to MITOMAP both yielded references, albeit disjoint ones: Sternberg et al. [2001] for the former and Manfredi et al. [1996] and Santorelli et al. [1997] for the latter.

The total results of Google, mtDB, and MITOMAP searches for variants (scored with respect to the L-strand) in the *MT-TC* gene (performed 20 March 2008) are shown in Table 2. Only about half of all variants were registered in MITOMAP and, in contrast, all but one ("T5785C") published variant were recognized by Google. If "5785 mtDNA" had been entered instead, then this particular case would also have been retrieved from Google (with reference to MITOMAP). The population database hits have all been obtained through the website of Ian Logan, to which one is directed via Google.

There are three frequent variants shared by major haplogroups in the mtDNA phylogeny: m.5773G>A (in the L-strand origin), m.5814T>C, and m.5821G>A. None of these mutations is well conserved in mammals; see Figure 2 of Kivisild et al. [2006]. Each of these three variants occur at least 10 times in the entire coding-region sequences registered in mtDB. In fact, those three mutations appear multiple times in the human mtDNA phylogeny and, in particular, enter the characteristic motif of several major haplogroups, such as the African haplogroups L2b (m.5814T>C) and L3b (m.5773G>A), the East Asian haplogroups A5a1a, D4i1, and M13 (m.5773G>A), and the Native American haplogroup D1a (m.5821G>A). For most tRNA genes only one site, such as m.5773G>A was seen to be polymorphic in the Finnish population, without disease status [Vilmi et al., 2005].

In MITOMAP, both m.5773G>A and m.5821G>A are listed under the heading "Coding & RNA Polymorphisms" with the appropriate references. m.5814T>C is found in the list "Reported Mitochondrial DNA Base Substitution Diseases: rRNA/tRNA mutations" and, surprisingly, regarded as a confirmed disease mutation for mitochondrial encephalopathy, with reference to the three studies [Santorelli et al., 1997; Sternberg et al., 1998, 2001]. Somewhat confusingly, the same mutation is also listed in "Clinical Phenotypes (non-LHON) Associated with mtDNA rRNA & tRNA Mutations" with provisional status ("not yet confirmed by multiple labs"), with two references [Manfredi et al., 1996; Sternberg et al., 1998].

The m.5773G>A mutation seems to have been reported first by Jaksch et al. [2001] and Maca-Meyer et al. [2001]. Subsequently, it was rediscovered as a "novel" pathogenic mutation in Sato et al. [2003]. Later, this mutation was declared a "novel polymorphism" albeit with unclear functional significance [Jacobs et al., 2005]. Those disease studies did not disclose to which haplogroup the m.5773G>A carriers belonged. In Europe, mtDNAs with m.5773G>A could well be members of some small sub-branch of either haplogroup J1c1 (also represented by one patient from Sonnenschein [2006]) or K1a2.

The m.5821G>A mutation was first (according to MITOMAP) observed in a case (HCM P-2) of hypertrophic cardiomyopathy by Ozawa et al. [1991]. This mtDNA lineage belongs to haplogroup D4b1a and shares m.5821G>A with another D4b1a lineage NDsq0101 from Tanaka et al. [2004], from which this variant might define a minor branch of D4b1a. This mutation was found several times by Wong et al. [2002], mostly in the heteroplasmic state; preliminary results had already been announced at a meeting in 1998 (www.lecb.ncifcrf.gov/~zullo/mitominiDB/1998NIHMito.html), in which pathogenicity of m.5821G>A was explicitly suggested. Later, this mutation was detected in the mtDNA of Patient BJ105 [Zhao et al., 2005], which is closely related to a particular haplogroup C lineage found by Kong et al. [2003] (see also Yao et al. [2006]). It was claimed in Young et al. [2006] that the presence of this mutation (as well as others) would influence the penetrance of hearing loss.

Is m.15287T>C Mildly Deleterious?

Most recently, the mutation m.15287T>C in the cytochrome *b* (*MT-CYB*; MIM# 516020) gene has been declared to be “novel” [Ballana et al., 2008]. The query “T15287C” to MITOMAP leads to no reference but “15287” does: this mutation is judged as a coding-region polymorphism with reference to Janssen et al. [2006], in which this mutation was found on a haplogroup I background in a family with diabetes and deafness (GenBank accession number AY245555). A closely related coding-region sequence (accession number DQ112955) has been published by Kivisild et al. [2006]. The mutations in the coding region shared by both haplogroup I lineages are m.3645T>C, m.3915G>A, m.6116A>G, m.7804A>G, and m.15287T>C, which thus seem to define a novel infrequent sub-branch of haplogroup I (Fig. 1) [Derenko et al., 2007].

Interestingly, m.15287T>C was also found earlier in a patient with a mitochondrial disorder, together with m.9205-9206del, m.9335C>T, m.11362A>G, m.12822A>G, and m.15705T>C [Chrzanowska-Lightowlers et al., 2004]. The three variants m.9335C>T, m.11362A>G, and m.12822A>G (deemed in Chrzanowska-Lightowlers et al. [2004] to be novel at the time) point to haplogroup N1b1 [Maca-Meyer et al., 2001; Mishmar et al., 2003]. The entire mtDNA variation of this lineage is displayed in Figure 1. In contrast, the occurrence of m.15287T>C in the affected Spanish family in Ballana et al. [2008] was in a particular branch of haplogroup H6a [Achilli et al., 2004], but the pathogenicity status of m.15287T>C was rather left open [Ballana et al., 2008]. Four more mtDNA sequences also bear m.15287T>C, which all belong to haplogroup M5a (DQ112875 [Kivisild et al., 2006], AY922289 [Sun et al., 2006], and EF583176 and PL173 [Malyarchuk et al., 2008]).

The mutation m.15287T>C was also found in a child (Patient E) with oxidative phosphorylation (OXPHOS) deficiency in skeletal muscle [Hinttala et al., 2006; Hinttala, 2007]. The corresponding mtDNA lineage belongs to a specific branch of haplogroup U6a2, which has otherwise been sampled only in East Africa [Olivieri et al., 2006]. Interestingly, the patient’s mtDNA bears only a single coding-region mutation, viz. m.15287T>C, not recorded in the closest East African near-match. Since the mtDNAs investigated in Hinttala et al. [2006] and Hinttala [2007] were all collected at the University Medical Centre in Nijmegen (The Netherlands) without indicating ethnic background, Patient E may well have an African matrilineal ancestry. It is not surprising that a rare variant such as m.15287T>C found mainly outside Europe (e.g., in the Indian haplogroup M5a1), whether it is judged a benign polymorphism or not, was not found in “the 617 control sequences belonging to European haplogroups” [Hinttala et al., 2006; Hinttala, 2007]. A similar caveat holds for the mutation m.15851A>G found in Patient A (with an mtDNA belonging to haplogroup K1) [Hinttala, 2007; Hinttala et al., 2006]: this mutation is otherwise known to be characteristic of the entire East Asian haplogroup B5b.

Curiously, 4 out of the 5 independent occurrences of m.15287T>C are coupled with a synonymous mutation at position 3915 or 3921. In these patients’ cases, mutations other than m.15287T>C could contribute to the severity of the disease. For instance, on haplogroup H6a1 the additional mtDNA background of two patients (one with and one without m.15287T>C; Fig. 1) encompasses the very rare mt-tRNA^{Leu} variant m.4314T>C, which otherwise seems to have been observed only once [Fraumene et al., 2006].

At present, one cannot firmly decide whether or not m.15287T>C is a benign polymorphism. At least the occurrence

as a mutation characteristic of haplogroup-M5a1 and in a branch of haplogroup I rules out a role as a high penetrance pathogenic mutation. A general slight impoverishment of OXPHOS activity incurred by this mutation alone (as might be suggested by the case from Hinttala [2007] and Hinttala et al. [2006]) would still need confirmation in healthy individuals from different haplogroups bearing this mutation. To clarify whether, under diverse nuclear backgrounds, mtDNAs belonging to these branches are impoverished in their function or not, further functional analyses would be required.

CONCLUSION

There is a considerable temptation to label a mutation found in some patients as “novel” if it is yet to be registered in MITOMAP. This runs the risk of reinventing the wheel, since many mutations repeatedly recorded in population or disease studies have not yet found their way into this database. Extrapolating from our screening of mutations in the *MT-TC* and a short fragment of the *MT-ND3* genes, the chances may be as low as ~50% that a particular mutation featured in some study gets actually recognized by MITOMAP. These shortcomings would not constitute a real problem if MITOMAP was just consulted as one convenient information source out of many. But it seems that most authors as well as reviewers would typically use MITOMAP as their only Web resource, rather than in combination with other search strategies. Claims of “novelty” of a mutation should thus always be treated with caution. Once the mutation has been published, a later rediscovery should no longer attach the label “novel” or “new” to the mutation, even when it occurs in a different disease context for the first time. Therefore, the title of the study by Sarzi et al. [2007] is misleading, despite the correct citation of the earlier work of Kirby et al. [2004].

The traditional means of dealing with “novel mtDNA mutations” may even promote scientific malpractice, in that a laboratory could succeed in publishing case studies by deliberately suppressing information that was available in the laboratory and public databases at the time but not yet present in MITOMAP [Bandelt et al., 2008]. A focus of seeming “novelty” also bears the risk of disregarding the information on known mutations seen in patients. This is most unfortunate because the whole mtDNA haplotype could matter and should be taken into consideration inasmuch as one could envision a more complex interplay between mutations in the phenotypic expression of a disease. Haplotypes are classified according to haplogroup status, which reflects their position in the mtDNA phylogeny. Note that the nomenclature concerning haplogroups is constantly evolving by refinement and modification of earlier nomenclature, as more and more complete mtDNA sequences are becoming available for phylogenetic comparison. Previous and novel information about the classification of mtDNA lineages can also be retrieved from the Web by searching a targeted geographic region or specific haplogroup name [Bandelt et al., 2006b, 2008].

We would suggest the following simple guideline in querying the status of an mtDNA variant. First, search the available Web resources such as MITOMAP and mtDB for the presence of this variant. Second, perform the absolutely minimum search strategy by entering the targeted mutation plus a key word such as “mtDNA” into an Internet search engine, e.g., Google [Bandelt et al., 2006b]. Thereby all possible ways of denoting the mutation should be used in the queries, following the strategies outlined above in the examples from the literature. Third, classify the mtDNA lineage and perform a phylogenetic analysis for all

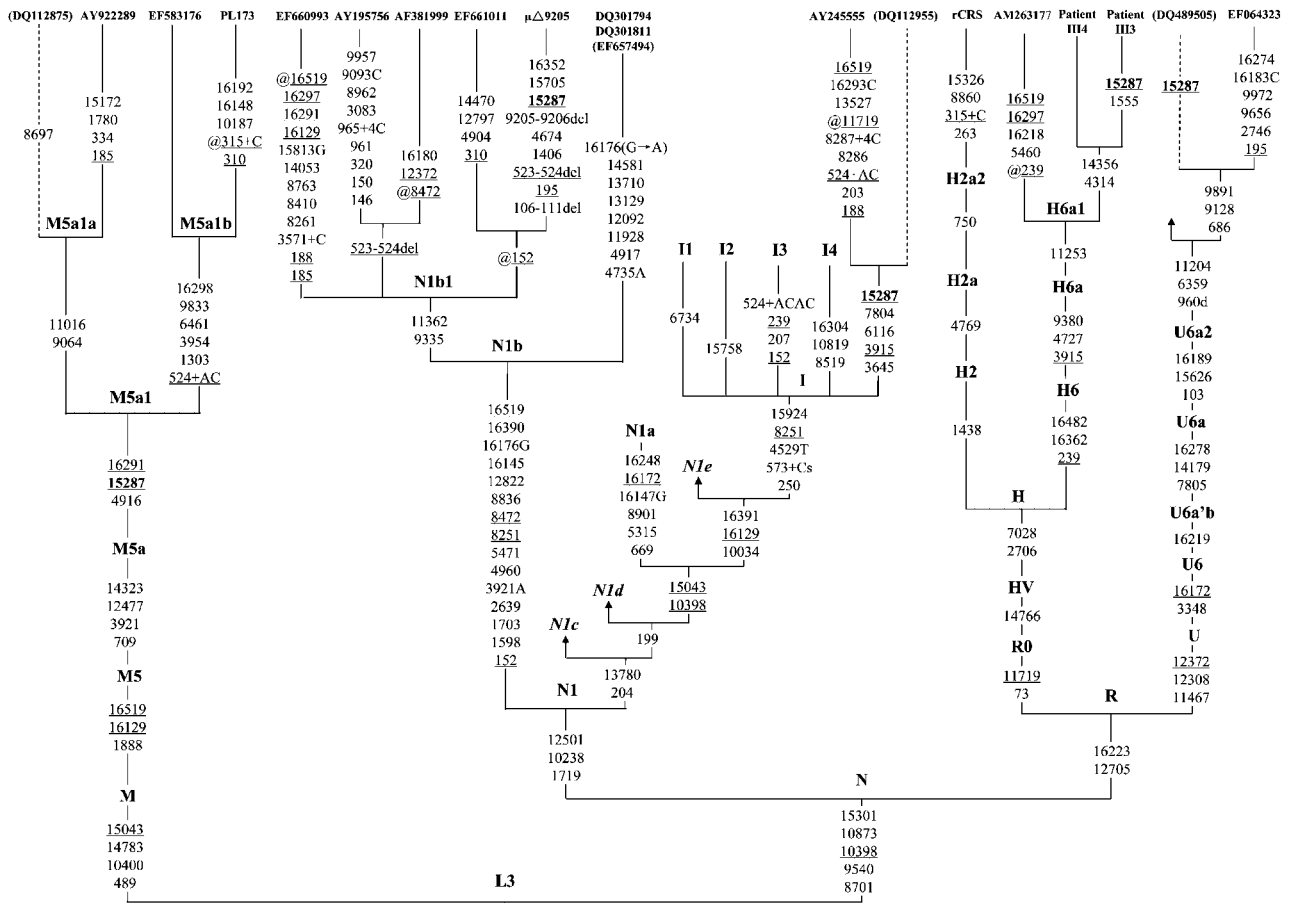


FIGURE 1. The mtDNA phylogeny of mtDNA lineages bearing m.15287T>C and their closest relatives lacking this variant. All mutations are transitions relative to rCRS, unless suffixes (A, G, C, T, or +, del, respectively) indicate transversions or indels; 573 + Cs means a variable number of cytosines can be inserted at this site, such as 309 + C, 309 + CC, is not recorded; all recurrent mutations are underlined and any back mutation is prefixed with @. Patients III3 and III4 are from Ballana et al. [2008]; sequence $\mu\delta$ 9205 is from Chrzanowska-Lightowlers et al. [2004] and contains a 2-bp microdeletion of mtDNA ($\mu\delta$) that removed the two residues at or between the ends of *MT-ATP6* and the *MT-CO3* gene. Lineages EF657494, DQ112875, DQ112955, and DQ489505 miss the control-region information, and therefore control-region mutations from sister branches may well occur in those mtDNA lineages as well. The names of haplogroups are given at the branching points or along single branches.

reported mtDNAs with the queried mutation, taking our investigation of m.15287T>C as a guide. It is strongly recommended that authors, reviewers, and editors take a minute to perform such necessary searches. In view of the immense number of novel mtDNA studies and data that are being published every month, one should not expect that a single database source can serve all the needs for the working medical geneticist. Consequently, an understanding of available and pertinent search strategies might be considered far more important than any single reference database.

ACKNOWLEDGMENTS

We thank Dr. Z. Chrzanowska-Lightowlers for sharing previously unpublished data and Dr. David Thorburn for useful information on an earlier case study. We also thank the anonymous reviewers for helpful comments on the early version of the manuscript. R.W.T. thanks the Wellcome Trust, the Newcastle upon Tyne Hospitals Foundation NHS Trust, and the NHS NCG Rare Mitochondrial Disorders of Adults and Children Service) for their continuing financial support. Y.G.Y. was supported by the

“Century Program” (or Hundreds-Talent Program) of the Chinese Academy of Sciences.

REFERENCES

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon J-M, Roostalu U, Loogväli E-L, Kivisild T, Bandelt H-J, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A. 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.

Ballana E, Govea N, de Cid R, Garcia C, Arribas C, Rosell J, Estivill X. 2008. Detection of unrecognized low-level mtDNA heteroplasmy may explain the variable phenotypic expressivity of apparently homoplasmic mtDNA mutations. *Hum Mutat* 29:248–257.

Bandelt H-J, Kong Q-P, Richards M, Macaulay V. 2006a. Estimation of mutation rates and coalescence times: some caveats. In: Bandelt H-J, Macaulay V, Richards M, editors. *Human mitochondrial DNA and the evolution of Homo sapiens*. Berlin: Springer-Verlag. p. 47–90.

Bandelt H-J, Salas A, Bravi CM. 2006b. What is a “novel” mtDNA mutation—and does “novelty” really matter? *J Hum Genet* 51:1073–1082.

- Bandelt H-J, Yao Y-G, Salas A. 2008. The search of “novel” mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor. *Int J Cardiol* 126:439–442.
- Brandon MC, Lott MT, Nguyen KC, Spolim S, Navathe SB, Baldi P, Wallace DC. 2005. MITOMAP: a human mitochondrial genome database—2004 update. *Nucleic Acids Res* 33:D611–D613.
- Brandstätter A, Sängler T, Lutz-Bonengel S, Parson W, Béraud-Colomb E, Wen B, Kong Q-P, Bravi CM, Bandelt H-J. 2005. Phantom mutation hotspots in human mitochondrial DNA. *Electrophoresis* 26:3414–3429.
- Chae JH, Lee JS, Kim KJ, Hwang YS, Bonilla E, Tanji K, Hirano M. 2007. A novel ND3 mitochondrial DNA mutation in three Korean children with basal ganglia lesions and complex I deficiency. *Pediatr Res* 61:622–624.
- Chrzanowska-Lightowlers ZMA, Temperley RJ, Smith PM, Seneca SH, Lightowlers RN. 2004. Functional polypeptides can be synthesized from human mitochondrial transcripts lacking termination codons. *Biochem J* 377:725–731.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee HK, Vanecek T, Villemis R, Zakharov I. 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet* 81:1025–1041.
- Fraumene C, Belle EMS, Castrì L, Sanna S, Mancosu G, Cosso M, Marras F, Barbujuani G, Pirastu M, Angius A. 2006. High resolution analysis and phylogenetic network construction using complete mtDNA sequences in Sardinian genetic isolates. *Mol Biol Evol* 23:2101–2111.
- Hinttala R, Smeets R, Moilanen JS, Ugalde C, Uusimaa J, Smeitink JA, Majamaa K. 2006. Analysis of mitochondrial DNA sequences in patients with isolated or combined oxidative phosphorylation system deficiency. *J Med Genet* 43:881–886.
- Hinttala R. 2007. Genetic causes of mitochondrial complex I deficiency in children. Academic dissertation, University of Oulu, Oulu, Finland (*Acta Univ Oul D Medica* 904).
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713.
- Jacobs HT, Hutchin TP, Käppi T, Gillies G, Minkkinen K, Walker J, Thompson K, Rovio AT, Carella M, Melchionda S, Zelante L, Gasparini P, Pyykkö I, Shah ZH, Zeviani M, Mueller RF. 2005. Mitochondrial DNA mutations in patients with postlingual, nonsyndromic hearing impairment. *Eur J Hum Genet* 13:26–33.
- Jaksch M, Kleinle S, Scharfe C, Klopstock T, Pongratz D, Müller-Höcker J, Gerbitz KD, Liechti-Gallati S, Lochmuller H, Horvath R. 2001. Frequency of mitochondrial transfer RNA mutations and deletions in 225 patients presenting with respiratory chain deficiencies. *J Med Genet* 38:665–673.
- Janssen GMC, Neu A, 't Hart LM, van de Sande CMT, Antonie Maassen J. 2006. Novel mitochondrial DNA length variants and genetic instability in a family with diabetes and deafness. *Exp Clin Endocrinol Diabetes* 114:168–174.
- Kirby DM, Salemi R, Sugiana C, Ohtake A, Parry L, Bell KM, Kirk EP, Boneh A, Taylor RW, Dahl HH, Ryan MT, Thorburn DR. 2004. NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency. *J Clin Invest* 114:837–845.
- Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.
- Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P. 2003. Phylogeny of east Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73:671–676 [Erratum: 2004; 75:157].
- Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.
- Malyarchuk BA, Perkova MA, Derenko MV, Vanecek T, Lazur J, Gomolcak P. 2008. Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Ann Hum Genet* 72:228–240.
- Manfredi G, Schon EA, Bonilla E, Moraes CT, Shanske S, DiMauro S. 1996. Identification of a mutation in the mitochondrial tRNA(Cys) gene associated with mitochondrial encephalopathy. *Hum Mutat* 7:158–163.
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171–176.
- Olivieri A, Achilli A, Pala M, Battaglia V, Fornarino S, Al-Zahery N, Scozzari R, Cruciani F, Behar DM, Dugoujon JM, Coudray C, Santachiara-Benerecetti AS, Semino O, Bandelt H-J, Torroni A. 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science* 314:1767–1770.
- Ozawa T, Tanaka M, Sugiyama S, Ino H, Ohno K, Hattori K, Ohbayashi T, Ito T, Deguchi H, Kawamura K, Nakane Y, Hashiba K. 1991. 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.
- Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, Sica L, Mougouma-Daouda P, Comas D, Tzur S, Balanovsky O, Kidd KK, Kidd JR, van der Veen L, Hombert JM, Gessain A, Verdu P, Froment A, Bahuchet S, Heyer E, Dausset J, Salas A, Behar DM. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc Natl Acad Sci USA* 105:1596–1601.
- Santorelli FM, Siciliano G, Casali C, Basirico MG, Carrozzo R, Calvosa F, Sartucci F, Bonfiglio L, Murri L, DiMauro S. 1997. Mitochondrial tRNA(Cys) gene mutation (A5814G): a second family with mitochondrial encephalopathy. *Neuromuscul Disord* 7:156–159.
- Sarzi E, Brown MD, Lebon S, Chretien D, Munnich A, Rotig A, Procaccio V. 2007. A novel recurrent mitochondrial DNA mutation in ND3 gene is associated with isolated complex I deficiency causing Leigh syndrome and dystonia. *Am J Med Genet A* 143:33–41.
- Sato A, Endo H, Umetsu K, Sone H, Yanagisawa Y, Saigusa A, Aita S, Kagawa Y. 2003. Polymorphism, heteroplasmy, mitochondrial fusion and diabetes. *Biosci Rep* 23:313–337.
- Sonnenschein A. 2006. Mitochondriale DNA Mutationen und Untersuchungen zum oxidativen Stress beim idiopathischen Parkinsonsyndrom (Mitochondrial DNA mutations and analyses of oxidative stress in Parkinson's disease) [in German]. Doctoral Dissertation, Technische Universität Dresden, Dresden, Germany, 138p.
- Steffann J, Gigarel N, Frydman N, Burlet P, Kerbrat V, Tachdjian G, Bonnefont JP, Frydman R, Munnich A. 2007. Preimplantation diagnosis for mitochondrial DNA disorders: contribution to understanding mitochondrial DNA segregation during early human embryonic development. Presented at the 57th Annual Meeting of the American Society of Human Genetics, October 23–27, 2007, San Diego, CA. Platform Presentation 196 (Abstract). Available at: www.ashg.org/genetics/ashg07s/f/20479.htm. Last accessed: June 11, 2008.
- Sternberg D, Danan C, Lombès A, Laforêt P, Girodon E, Goossens M, Amselem S. 1998. Exhaustive scanning approach to screen all the mitochondrial tRNA genes for mutations and its application to the investigation of 35 independent patients with mitochondrial disorders. *Hum Mol Genet* 7:33–42.
- Sternberg D, Chatzoglou E, Laforêt P, Fayet G, Jardel C, Blondy P, Fardeau M, Amselem S, Eymard B, Lombès A. 2001. Mitochondrial DNA transfer RNA gene sequence variations in patients with mitochondrial disorders. *Brain* 124:984–994.
- Sun C, Kong Q-P, Palanichamy Mg, Agrawal S, Bandelt H-J, Yao Y-G, Khan F, Zhu C-L, Chaudhuri TK, 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.
- Tanaka M, Cabrera VM, González AM, Larruga JM, Takeyasu T, Fuku N, Guo LJ, Hirose R, Fujita Y, Kurata M, Shinoda K, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H. 2004. Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res* 14:1832–1850.

- Tchikviladzé M, Laforêt P, Eymard B, Delbos F, Filaut S, Lombès A, Jardel C. 2007. A novel mutation in the mitochondrial ND3 gene causing Leigh syndrome with late-onset neurological decline [Abstract]. *Neuromuscul Disord* 17:769.
- Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R. 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69:1348–1356.
- Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Lee ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol* 3:e247.
- Vilmi T, Moilanen JS, Finnilä S, Majamaa K. 2005. Sequence variation in the tRNA genes of human mitochondrial DNA. *J Mol Evol* 60: 587–597.
- Wang K, Takahashi Y, Gao Z, Goto J, Wang G, Tsuji S. 2006. Mitochondrial ND3 gene as the novel causative gene for leber hereditary optic neuropathy and dystonia. *Ann Neurol* 60:S55–S56 (Abstract M-87).
- Wong L-JC, Liang M-H, Kwon H, Park J, Bai R-K, Tan D-J. 2002. Comprehensive scanning of the entire mitochondrial genome for mutations. *Clin Chem* 48:1901–1912.
- Wong L-JC. 2007. Pathogenic mitochondrial DNA mutations in protein-coding genes. *Muscle Nerve* 36:279–293.
- Yao Y-G, Salas A, Bravi CM, Bandelt H-J. 2006. A reappraisal of complete mtDNA variation in East Asian families with hearing impairment. *Hum Genet* 119:505–515.
- Young W-Y, Zhao L, Qian Y, Li R, Chen J, Yuan H, Dai P, Zhai S, Han D, Guan M-X. 2006. Variants in mitochondrial tRNA^{Glu}, tRNA^{Arg}, and tRNA^{Thr} may influence the phenotypic manifestation of deafness-associated 12S rRNA A1555G mutation in three Han Chinese families with hearing loss. *Am J Med Genet A* 140:2188–2197.
- Zhao L, Wang Q, Qian Y, Li R, Cao J, Hart LC, Zhai S, Han D, Young W-Y, Guan M-X. 2005. Clinical evaluation and mitochondrial DNA sequence analysis in two Chinese families with aminoglycoside-induced and non-syndromic hearing loss. *Biochem Biophys Res Commun* 336:967–973.