Functional Recurrent Mutations in the Human Mitochondrial Phylogeny: Dual Roles in Evolution and Disease

Liron Levin, Ilia Zhidkov, Yotam Gurman, Hadas Hawlena, and Dan Mishmar*

Department of Life Sciences, Ben-Gurion University of the Negev, Beer Sheva 84105, Israel

*Corresponding author: E-mail: dmishmar@bgu.ac.il.

Accepted: April 2, 2013

Data deposition: This article constitutes an analysis of sequences that are already available in GenBank. Accession numbers are mentioned within the text.

Abstract

Mutations frequently reoccur in the human mitochondrial DNA (mtDNA). However, it is unclear whether recurrent mtDNA nodal mutations (RNMs), that is, recurrent mutations in stems of unrelated phylogenetic nodes, are functional and hence selectively constrained. To answer this question, we performed comprehensive parsimony and maximum likelihood analyses of 9,868 publicly available whole human mtDNAs revealing 1,606 single nodal mutations (SNMs) and 679 RNMs. We then evaluated the potential functionality of synonymous, nonsynonymous and RNA SNMs and RNMs. For synonymous mutations, we have implemented the Codon Adaptation Index. For nonsynonymous mutations, we assessed evolutionary conservation, and employed previously described pathogenicity score assessment tools. For RNA genes' mutations, we designed a bioinformatic tool which compiled evolutionary conservation and potential effect on RNA structure. While comparing the functionality scores of nonsynonymous and RNA SNMs and RNMs with those of disease-causing mtDNA mutations, we found significant difference ($P < 0.001$). However, 24 RNMs and 67 SNMs had comparable values with disease-causing mutations reflecting their potential function thus being the best candidates to participate in adaptive events of unrelated lineages. Strikingly, some functional RNMs occurred in unrelated mtDNA lineages that independently altered susceptibility to the same diseases, thus suggesting common functionality. To our knowledge, this is the most comprehensive analysis of selective signatures in the mtDNA not only within proteins but also within RNA genes. For the first time, we discover virtually all positively selected RNMs in our phylogeny while emphasizing their dual role in past evolutionary events and in disease today.

Key words: homoplasy, mitochondrial DNA, recurrent nodal mutations, selection.

Introduction

Independently recurring mutations (homoplasy) in human mitochondrial DNA (mtDNA) phylogeny is a well-known phenomenon, and is thought to be the consequence of the high mtDNA mutation rate (Pereira et al. 2011). As most mutational events are removed by negative selection, only a subset of the recurrent mutations will be retained in the population, of which the majority would have little or no effect and hence can be considered as neutral. Is it possible that homoplasic mtDNA mutations that were retained and inherited in the mtDNA phylogeny possess adaptive properties and were thus positively selected?

Although population fixation of mtDNA variants has traditionally been attributed to genetic drift, it has been repeatedly shown that certain common human mtDNA genetic variants possess functional attributes and are, therefore, likely to be under selective constraints (Wallace 2005). As natural selection acts on phenotype, one would expect that such mtDNA variation would affect mitochondrial function. Indeed, cell culture experiments demonstrated that human mtDNA polymorphisms affected mitochondrial activities, such as calcium uptake (Kazuno et al. 2006), reactive oxidation species (ROS) production (Moreno-Loshuertos et al. 2006), and mtDNA transcription (Suissa et al. 2009). Genetic association studies revealed the phenotypic impact of human mtDNA genetic backgrounds, especially on age-related phenotypes, such as type 2 diabetes mellitus and Parkinson's disease (reviewed by Mishmar and Zhidkov 2010). Moreover, certain mtDNA variants were shown to possess adaptive attributes in multiple metazoan taxa (Castellana et al. 2011). Finally, co-evolution among mtDNA- and nuclear DNA-encoded factors underlined the functional importance of mtDNA variants for protein–protein interactions (Gershoni et al. 2009, 2010; Bar-Yaacov et al. 2012). Although the functional importance and phenotypic effects of common mtDNA variants is clear, functionality of recurrent mutations found in the stems of unrelated mtDNA phylogenetic nodes, that is, recurrent nodal mutations (RNMs), is less obvious.

Recently, it was suggested that recurring mutations in human mtDNA tend to affect positions of lesser functional importance (Pereira et al. 2011). However, several pieces of evidence indicate that a subset of recurring mtDNA mutations are potentially functional and may even offer adaptive value. First, mtDNA mutations that became recurrently fixed in several independent cancer samples recapitulated ancient mtDNA genetic backgrounds (haplogroups), thus suggesting commonalities in the selective constraints acting on the mitochondrial genome in cancer and during human evolution (Zhidkov et al. 2009). Second, mtDNA haplogroups T and N1b2 share an amino acid replacement (mutation A4917G) that altered a highly evolutionary conserved position, thus affecting susceptibility to reduced sperm motility (Ruiz-Pesini et al. 2000) and to complications in type 2 diabetes patients (Feder et al. 2008). Finally, the mutation at position 1,555 that causes hearing loss in humans exposed to aminoglycosides recurred in orangutan mtDNA and affected mitochondrial function in cell culture (Pacheu-Grau et al. 2011). Although promising, these pieces of evidence do not reveal the extent of this phenomenon, that is, how many of the total set of mtDNA RNMs carry functional attributes.

Here, we tested the hypothesis that a subset of the single nodal mutation (SNMs) and RNMs in the human mtDNA phylogeny carries functional properties and has survived years of evolution due to its adaptive nature. To test our hypothesis, we constructed a detailed phylogenetic tree from 9,868 publicly available, nonredundant, whole human mtDNA sequences (and out-group sequences) that was in complete agreement with the most updated mtDNA phylogenetic tree topology (Behar et al. 2012). Then, we identified all of the mutational events that occurred during human mtDNA phylogeny and associate them with each of the obtained tree branches and lineages. By bioinformatics assessment of the functional potential of nodal synonymous, nonsynonymous, and RNA genes mutations, we identified the most likely candidate SNMs and RNMs to have been retained in the human population due to positive selection.

Materials and Methods

Whole mtDNA Sequences

We retrieved 9,862 publicly available (Genbank) whole human mtDNA sequences, excluding redundancies (supplementary table S1, Supplementary Material online), that were used to construct the mtDNA phylogenetic tree (van Oven and Kayser 2008; Behar et al. 2012). We retrieved the NCBI accession numbers of these sequences and their phylogenetic assignment from the mtDNA Community website (www.mtdnacommunity.org, last accessed April 22, 2013), whereas the sequences were downloaded from the PhyloTree site (www.phylotree.org, last accessed April 22, 2013). The 9,862 sequences, together with six Neanderthal sequences (*Homo sapiens* neanderthalensis accession numbers: NC_011137.1, FM865409.1, FM865407.1, FM865408.1, FM865410.1, and FM865411.1) that were isolated from archaeological specimens (Briggs et al. 2009), were aligned using MAFFT (mafft.cbrc.jp/alignment/server/). MAFFT was used because it is specifically designed to analyze multiple sequence alignments comprising thousands of sequences.

Phylogenetic Analysis

The phylogenetic tree of whole mtDNA sequences was constructed in several stages. First, we used the lineage/cluster nomenclature published in the tree generated by Behar et al (2012) (supplementary table S1, Supplementary Material online). To predict ancestral sequences for each cluster, we used the Phylip software package (www.phylip.com, last accessed April 22, 2013), especially utilizing the maximum likelihood method with the molecular clock option. Briefly, trees were generated for each cluster of sequences and the ancestral sequences were predicted from the best tree out of 500 Jumble repetitions (i.e., randomized input order of sequences, as outlined in the manual of the Phylip software, www.phylip.com, last accessed April 22, 2013). Second, the resulting sequence list, which included an ancestral sequence from each cluster (i.e., the predicted ancestral sequences of each of the subtrees), was used to generate an ancestral phylogenetic tree (supplementary fig. S1, Supplementary Material online). This phylogenetic tree was constructed by MEGA5 (Tamura et al. 2011), using the neighbor-joining method (Saitou and Nei 1987) with the following parameters: gaps/missing data treatment – pairwise deletion, bootstrap – 1000X; substitution model: maximum composite likelihood method. Similar tree topology was received using fastTree (Price et al. 2009) and PhyML (Guindon and Gascuel 2003). For the sake of simplicity, we present only the NJ tree, which is consistent with previously published human mtDNA tree topology (van Oven and Kayser 2008; Behar et al. 2012). We used DNA parsimony (Phylip software package) to analyze the ancestral phylogenetic tree (supplementary fig. S2, Supplementary Material online) containing 563 ancestral sequences. We developed a bioinformatics tool to integrate data from the DNA Parsimony Algorithm program with the phylogenetic analysis (i.e., the 563 sub-trees, corresponding to the ancestral sequences created by the ML method by the molecular clock option). This enabled a global view of all mutational events that occurred in human mtDNA phylogeny (excluding ambiguous mutations, that is, mutations in which there is no certainty as to the identity of the nucleotide changes).

All sequences and their accession numbers, sequence alignment files, scripts, phylogenetic analysis, and sequences groups are available as supplementary data, Supplementary Material online.

Identifying Nodal Mutational Events

For the purposes of the current study and for the sake of simplicity, a mutational event was considered to be "nodal" only when fulfilling two criteria: 1) the mutation in a given mtDNA lineage should be shared by at least five sequences within this lineage/sub lineage (i.e., phylogenetic node). 2) These sequences should comprise at least 85% of all clustered sequences within the studied tree node. This proportion was calculated as follows: the average proportion of sequences within branches throughout the tree that were in agreement with criterion "1" was 95 ± 12%. We chose to use a proportion within one SD of this mean. Therefore, taken together, we defined a "nodal mutation" as a mutation shared by at least five sequences comprising at least 85% of the sequences within a given branch. Accordingly, RNMs were defined as mutations that lie in the stems of at least two unrelated tree nodes, following the principles of parsimony. Notably, although we are aware of the fact that mutations occur in the branch leading to a node rather than in the base of a given node, for the sake of simplicity we used the term "nodal mutations" as outlined earlier. Additionally, a mutation could occur in a branch leading to a phylogenetic node yet only became "nodal" in a particular subnode.

Assessing the Functional Potential of Nonsynonymous Mutations

Two basic measurements were performed to assess the functional potential of all the mutations in mtDNA-encoded protein-coding genes. We calculated the Conservation Index (Glaser et al. 2003) and the SIFT pathogenicity score (http://sift.bii.a-star.edu.sg/, last accessed April 22, 2013) (Kumar et al. 2009) of amino acid substitutions. In both tests, we generated sequence alignment files (T-Coffee, with minor manual corrections) using orthologous sequences from 296 mammalian species for each of the 13 mtDNA-encoded protein genes (supplementary table S2, Supplementary Material online, NCBI-Organelle Genome Resources). As previously suggested (Kumar et al. 2009), SIFT pathogenicity scores ≤ 0.05 were considered as having the highest deleterious potential. The Conservation Index was calculated by ConSurf (consurf.tau.ac.il, last accessed April 22, 2013) (Glaser et al. 2003) with the Evolutionary Substitution Model, that is, mtREV for mitochondrial proteins using default settings. Conservation Index scores $X \geq 7$ were considered as having the highest functional potential, being within 1SD from the mean Conservation Index scores of mtDNA disease-causing mutation (8.4 ± 1.8, see also results). The pathogenicity scores of amino acid substitutions were calculated using SIFT, based on the aforementioned multiple sequence alignment files. Two additional functionality measurements were applied to all nonsynonymous nodal mutations: MutPred general score (Li et al. 2009) and Panther P-deleterious score (Thomas et al. 2003) using the default cutoff values (MutPred > 0.5, Panther > 0.5). The resulting values were compared with those obtained from mtDNA disease-causing mutations (discussed later).

Assessing the Functional Potential of Synonymous Mutations

To assess the functional potential of "nodal" synonymous mutations, we estimated codon bias. The effective number of codons (NC) (Wright 1990), the GC content within the third codon position (i.e., GC3s) and the Codon Adaptation Index (CAI) (Sharp and Li 1987) were calculated using CodonW (codonw.sourceforge.net, last accessed April 22, 2013). Calculations were applied to all the observed codons in all protein-coding genes in our entire data set of 9,862 whole human mtDNA sequences. The CAI values were used to calculate the difference before and after the mutation occurrence as follows: ΔCAI of a given synonymous mutation equals CAI value after the synonymous mutational event minus the CAI value before the mutation occurred.

Assessing the Functional Potential of Mutations in RNA Genes

1. Assessing Conservation Indexes of Mutations in RNA Genes

Evolutionary conservation was previously used to evaluate the functionality of human mtDNA variants within RNA genes (Ruiz-Pesini and Wallace 2006). To calculate the Conservation Index of mutations identified in each of the 22 mtDNA-encoded tRNA genes, multiple sequence alignment files were retrieved from the Mamit-tRNA website (mamit-trna.u-strasbg.fr, last accessed April 22, 2013) using sequence orthologs from at least 114 mammalian species and analyzed by ConSurf (using default settings for nucleotide sequences). To calculate the Conservation Index for mutations in the two mtDNA-encoded rRNA genes (*12S* and *16S*), we used orthologous sequences from 296 mammalian species (supplementary table S2, Supplementary Material online, NCBI-Organelle Genome Resources) aligned by ClustalW (default settings). Conservation Index values were calculated using the Rate4Site application, a stand-alone program within the ConSurf server, using default settings for nucleotide sequences.

2. Assessing the Potential Structural Impact of Mutations in RNA Genes

The potential effect of a given variant on structure stability within rRNA and tRNA genes (free energy, ΔG) was assessed using the RNAeval application of the Vienna RNA Package (version 1.8.5, www.tbi.univie.ac.at/RNA/, last accessed April 22, 2013). This application evaluates the free energy of an RNA molecule in a fixed secondary structure, excluding "pseudoknots," which are RNA structures that are composed of two helical segments at least, connected by single-stranded regions or loops (Staple and Butcher 2005). For analysis of tRNA genes, we retrieved the RNA structures from the Mamit-tRNA website (mamit-trna.u-strasbg.fr/, last accessed April 22, 2013), while rRNA gene structures were retrieved from the CRW website (www.rna.icmb.utexas.edu/, last accessed April 22, 2013). For the latter website, the *12S rRNA* structure was retrieved from the "Current *12S rRNA* Structures" section and the *16S rRNA* was retrieved from the "Appendix to a large rRNA folding manuscript" section. We calculated ΔG by subtracting the assessed free energy of an RNA variant before the mutational event from the assessed free energy in the sequence harboring the mutation. Absolute values were used, because we were merely interested in the magnitude of the effect on structural stability.

3. Novel RNA-Mutation Scoring Method to Assess Functionality

As mentioned earlier, two measurements were performed to assess the functional potential of mutations within RNA-coding genes, namely assessment of the Conservation Index and calculation of the potential free energy change before and after the mutation (ΔG). In general, mutations in stems within stem-and-loop elements of RNA molecules can alter structural stability (ΔG). However, such mutations are not necessarily evolutionarily conserved as a result of compensatory mutations. Moreover, evolutionarily conserved mutations could be found in loops, having little or no predicted effect on structures. Therefore, a simple compilation of the two tests could easily lead to contrasting results. To overcome this obstacle, we formulated a scoring method lending equal weight to the results obtained from each test. To this end, the obtained values were transformed to scales of arbitrary units ranging from 1 to 9: 1) as already formulated in ConSurf, for the Conservation Index, the value "1" represented the most variable site and "9" the most conserved site. 2) For the ΔG test, the value "1" represented a lack of effect and "9" represented the largest effect on structural stability. Specifically, the score "9" is assigned to the greatest change in ΔG observed in the analysis within a studied gene or in disease-causing mutations. The combined RNA score is calculated as follows: RNA score = Conservation Index + ΔG index, yielding values ranging from 2 to 18. The resulting values were compared with those obtained with mtDNA disease-causing mutations (discussed later).

mtDNA Disease-Causing Mutations

To assess the functional potential of the phylogenetic variants considered, we sought a set of mutations with experimentally verified functionality. To this end, we applied the earlier-described tests of functionality to a set of mtDNA disease-causing mutations, of which 38 were nonsynonymous and 27 were in RNA genes. These mutations comprise all confirmed disease-causing mutations listed by MITOMAP, that is, were found in patients but not in controls in at least two independent studies (26 confirmed nonsynonymous and 27 RNA genes mutations, www.mitomap.org, last accessed April 22, 2013) in addition to 12 nonsynonymous mutations that were identified in patients and were experimentally verified for their functionality (listed by Ruiz-Pesini et al. 2004) (supplementary tables S3 and S4, Supplementary Material online).

Statistics

To compare the results obtained from the earlier described tests of functionality between nodal mutations and disease-causing mutations, we performed a Mann–Whitney *U* test and a resampling simulation with sequential Bonferroni correction (discussed later) for multiple testing. We preformed resampling analyses, which control for differences in sample size of the different subsets of mutations (Simon 1993; Good 2006) to determine the portion of nodal mutations in RNA genes that are most similar to the disease-causing mutations. The detailed procedure is describes in supplementary figure S3, Supplementary Material online. Briefly, each set of mutations was sorted according to functionality score. The different percentile of each set of mutations was tested against that of documented disease-causing mutations. Significant levels were adjusted for multiple tests, using a sequential Bonferroni correction (Rice et al. 2008). We used $R \times C$ (rows × columns) test of independence to test for differences in the frequency of mutations that passed the functionality test cutoff values in SNMs and RNMs as compared with disease causing mutations.

Results

Identifying Single and Recurrently Nodal Mutations in the Human mtDNA Phylogeny

We first sought to identify all of the mutational events that occurred during human mtDNA phylogeny. To this end, we constructed a phylogenetic tree from 9,868 publicly available whole mtDNA sequences (supplementary fig. S2, Supplementary Material online), including a nonredundant set of 9,862 whole human mtDNA sequences and six Neanderthal mtDNA sequences (supplementary table S1, Supplementary Material online). Indeed, the resulting tree topology closely resembled that of recently published human mtDNA trees (Ruiz-Pesini et al. 2007; van Oven and Kayser 2008; Behar et al. 2012).

As the topology of our phylogenetic tree was consistent with previous studies, we extensively analyzed the tree to identify all variants encompassing mutations that lie at the base of certain branches (nodal mutations), as well as

RNMs. It is noteworthy that recording of all variants was not conducted by comparison with a common reference sequence, such as the rCRS (revised Cambridge Reference Sequence), but rather by comparing the premutation node with the branch in which that mutation occurred. Our screen revealed 31,714 mutational events, of which 4,176 defined phylogenetic branches (nodal mutations). Of these, 1,606 lie in the stem of single phylogenetic branches (SNMs), and 2,570 mutational events occurred in the stem of two or more parsimoniously unrelated tree branches; the latter inhabit 679 nucleotide positions (i.e., 679 RNMs) (supplementary table S5, Supplementary Material online). Out of the total number of RNMs, 188 were either nonsynonymous ($N = 121$ RNMs, encompassing 357 mutational events) or RNA gene mutations ($N = 67$ RNMs, encompassing 219 mutational events). It is expected that if human mtDNA mutates at random, then the mutational proportion in noncoding sequences should be comparable with the proportion of such sequences in the human mitochondrial genome, that is, approximately 7%. In contrast, we found that noncoding mtDNA sequences harbored 35.8% of the total observed nodal mutations (including both RNMs and SNMs) and 49% of the RNMs, that is, 5-fold (χ^2 test, $P < 0.001$) and 7-fold (χ^2 test, $P < 0.001$) more than expected at random, respectively. This indicates the existence of strong purifying selection against the transmission of nodal mutations in coding mtDNA sequences and even stronger selective constraint acting upon RNMs.

Assessing the Functional Potential of Nonsynonymous RNMs and SNMs

Previously, two tests were used to assess the functionality of nonsynonymous mtDNA variants, that is, evolutionary conservation (Miller and Kumar 2001; Ruiz-Pesini et al. 2004) and SIFT pathogenicity score (Pereira et al. 2011). In our analysis, we have calculated the Conservation Index and SIFT scores for all the nonsynonymous mutation. These tests are not redundant but rather complementary, because a given mutation may alter a highly conserved position yet did not radically change the physical–chemical properties of the amino acid, and vice versa. In consistence with previously published results (Templeton 1996; Pereira et al. 2011), a comparison of the Conservation Index distribution of nonsynonymous mutation in general as well as those that passed the SIFT pathogenicity score cutoff threshold ($X \leq 0.05$) revealed that nodal mutations had significantly lower Conservation Index than either disease-causing mutations (Mann–Whitney *U* test, $P < 0.001$) or mutations occurring at the tips of the tree (Mann–Whitney *U* test, $P < 0.05$, fig. 1*A* and *B*). RNMs had the lowest Conservation Index, thus suggesting stronger negative selection acting against RNMs in particular.

We next asked whether a subset of the nodal mutations (RNMs and SNMs) have a high functional potential, and hence are candidates to be positively selected. As confirmed mtDNA disease-causing mutations are clearly functional (38 nonsynonymous mutations, supplementary table S3, Supplementary Material online), we calculated their scores using the earlier-mentioned tests (i.e., Conservation Index and SIFT scores) and used them as a reference of functionality to compared with our identified RNMs and SNMs. We further augmented our analysis by two additional functionality assessment tools, MutPred, and Panther, which were recently reported to perform well in comparison with other available assessment methods (Thusberg et al. 2011). First, our analysis validated the use of the disease-causing mutations as a reference for functionality as approximately 92% of these mutations not only passed the functionality threshold of SIFT and had a high Conservation Index (8.4 ± 1.8) but also passed at least one additional test—either MutPred or Panther (fig. 1*C*). Second, our results indicate that most of the 121 nonsynonymous RNMs or 283 SNMs (91% and 85%, respectively), did not pass the thresholds as explained for disease-causing mutations and in that are significantly different from the latter ($R \times C$ [rows × columns] test of independence, $G = 108.85$, df $= 2$, $P < 0.001$). Thus, along with the observed overrepresentation of nodal mutations in noncoding sequences, and the differences from the tip mutations our results suggest stronger selective constraints acting on nodal mutations. However, 11 out of the 121 nonsynonymous RNMs and 42 out of the 283 nonsynonymous SNMs passed the functionality score threshold of SIFT, had comparable Conservation Index with disease-causing mutations (i.e., within 1SD from the mean ConSurf values of disease-causing mutations), and passed at least one additional test (MutPred or Panther), similarly to the disease-causing mutations. We thus interpret these particular nonsynonymous RNMs and SNMs as the best candidates to bare adaptive properties.

Assessing the Functional Potential of Synonymous Nodal Mutations

Synonymous mutations may have functional consequences because of selective constraints acting on certain codons (Chamary et al. 2006). Codon bias (i.e., nonrandom codons usage) in highly expressed genes may reflect their adaptation toward translational efficient codons due to the abundance of their associated tRNAs (Ikemura 1985; Salinas et al. 2012). We used our 9,862 whole human mtDNA sequence data set to calculate the effective number of used codons (NC) for each amino acid in each studied protein-coding gene (Wright 1990). NC values range from 20 to 61, for example, from the usage of a single effective codon per amino acid to equal abundance of all codons. We found that the mtDNA NC values ranged from 31 (ND3) to 41 (COX2), thus reflecting moderate codon bias, which differs among mtDNA genes. NC-plot analysis (Wright 1990) excluded the possibility that certain transitions or transversions at the third codon position

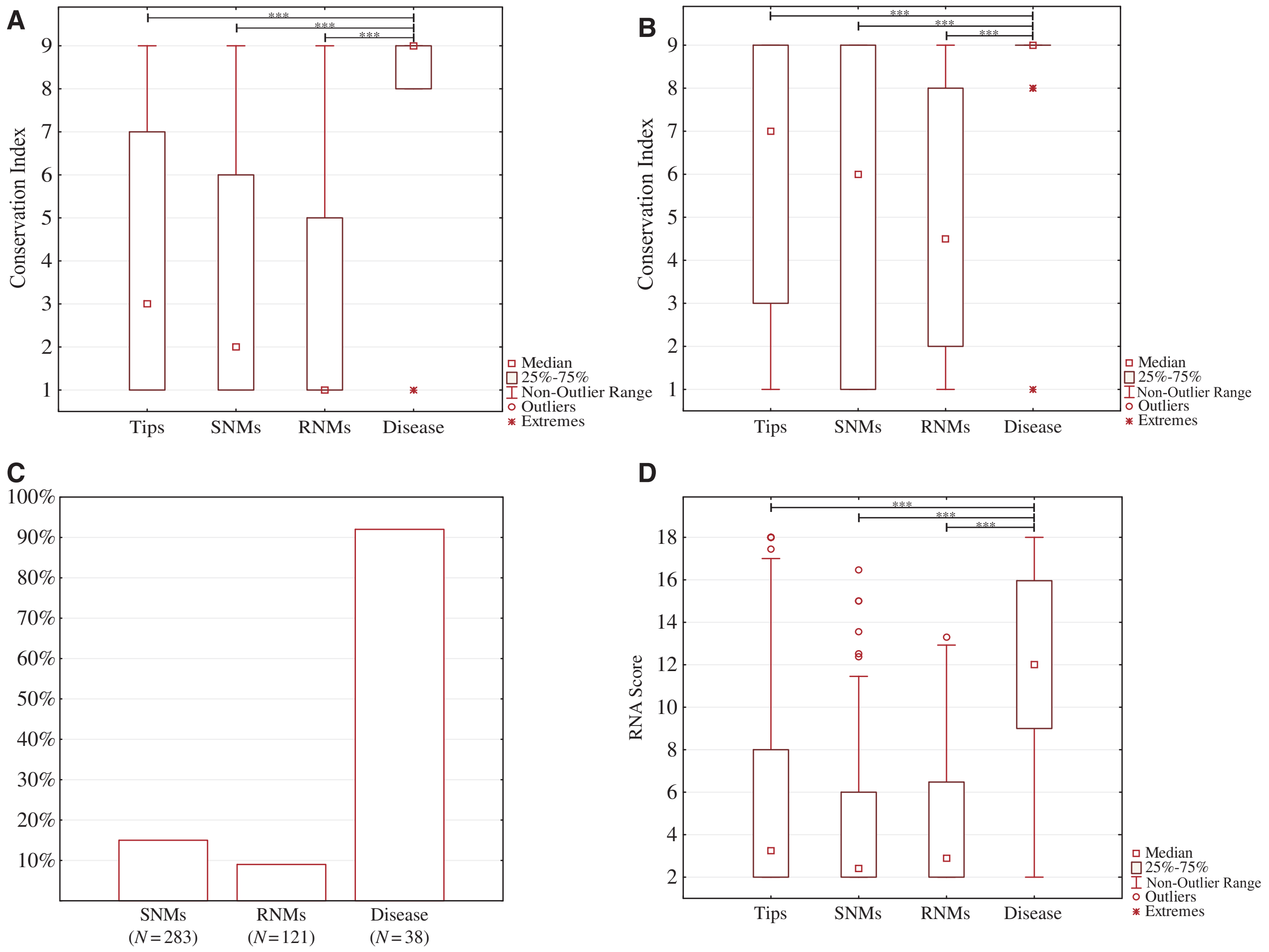

Fig. 1.—Comparison of functionality assessments of nonsynonymous and RNA genes mutations to disease causing and tips mutations. (*A*) The distribution of Conservation Index scores of nonsynonymous mutations in the tree Tips, SNMs, RNMs, and of disease-causing mutations. (*B*) The distribution of Conservation Index scores of nonsynonymous mutations, which also passed the SIFT score cutoff ($X \leq 0.05$) in the tree Tips, SNMs, RNMs, and of disease-causing mutations. (*C*) The percentage of nonsynonymous mutations (SNMs, RNMs, and disease-causing mutations) that passed the threshold of SIFT, the Conservation Index as well as the cutoff value of either MutPred or Panther. Total number of mutations is indicated for each category of tested mutations. (*D*) The score distribution of RNA genes mutations in the tree Tips, SNMs, RNMs, and of disease-causing mutations. ***Significant difference between the disease-causing mutations to the Tips, SNMs, or RNMs (Mann–Whitney *U* test, $P < 0.001$).

are responsible for the observed codon bias (GC3s) (supplementary fig. S4*A*, Supplementary Material online, χ^2 test, $P < 0.001$). Next, we have calculated the CAI (Sharp and Li 1987) in all mtDNA genes. In brief, CAI measures how the codon usage of a given gene resembles the codon usage of the most highly expressed genes, because the latter are the most likely to adapt toward using translational efficient codons. While inspecting recently published human mtDNA transcriptome data of 16 human tissues (Mercer et al. 2011), we identified ND3, the concatenate of ND4 and ND4L, COX3 and CYTB as the most highly expressed mtDNA transcripts. Our analysis revealed a correlation between the relative expression of mtDNA genes and their CAI values, after correction for peptide length (CAI/amino acid length). Using each of the four highly expressed genes as references for the CAI calculation, produced similar correlation whereas the highest correlation was obtained using either CYTB or the concatenated gene ND4\ND4L as references (supplementary fig. S4*B*, Supplementary Material online, $r = 0.67$, $P = 0.024$ and $r = 0.68$, $P = 0.021$, respectively). Excluding the ND4\ND4L concatenate from the correlation analysis revealed even higher correlation of the remaining mtDNA genes (supplementary fig. S4*C*, Supplementary Material online, $r = 0.84$, $P = 0.0024$). Therefore, for the sake of simplicity, we used CYTB as a single reference for further CAI calculations. As adaptation toward translational efficiency is correlated with tRNA abundance, we examined which are the most frequently used codons in the highly expressed CYTB gene as compared with the expected random codons usage (equal distribution). We identified 28 codons that are used significantly more than expected by chance ($P < 0.01$, two way χ^2 test), of which 19 correspond to the mtDNA-encoded tRNAs, which is significantly more than expected by chance (χ^2 test, $P < 0.001$). Accordingly, a recent study showed higher abundance of mtDNA-encoded tRNAs in the mitochondria as compared with imported tRNAs (Mercer et al. 2011). These findings further support the notion that CYTB adapted toward using more translational efficient codons. We next estimated the functional potential of mtDNA synonymous mutations by calculating the difference in CAI values before and after the occurrence of each identified nodal synonymous mutation (i.e., ΔCAI) using CYTB as a reference (supplementary table S5, Supplementary Material online). While comparing the distribution of ΔCAI values between either SNMs or RNMs to tree tips mutations, we found significant differences (fig. 2, Mann–Whitney *U* test, $P < 0.001$). Nevertheless, whereas the SNMs showed general reduction in values, the RNMs showed the opposite trend (fig. 2). Hence, our analysis did not detect consistent selective signature among nodal synonymous mutations. Moreover, in the absence of mtDNA disease-causing synonymous mutations, there is no set of clearly functional synonymous mutations to be used as a "functionality reference" of CAI values.

Identifying Nodal Mutations with Functional Potential in RNA Genes

Similar to the analysis of nonsynonymous mutations, we compared the sequence attributes of RNMs and SNMs in mtDNA-encoded RNA genes (tRNAs and rRNA genes) with those of disease-causing mutations in the same genes. As RNA genes are not translated, we generated a novel combined test which takes into account evolutionary conservation and the potential effect of the tested mutation on the predicted structural stability of the molecule (Δ*G*) (see Materials and Methods). Our novel RNA mutations functionality scoring method was applied to 27 disease-causing mutations in RNA genes (supplementary table S4, Supplementary Material online) and used as a reference upon assessment of the functional potential of RNMs and SNMs. As observed in the distribution of nonsynonymous functionality scores, the RNA scores of RNA nodal mutations (i.e., either RNMs or SNMs, fig. 1*D*) had significantly lower values than that of disease-causing mutations (Mann–Whitney *U* test, $P < 0.001$). However, only SNMs also had significantly lower values than mutations occurring at the tips of the tree (Mann–Whitney *U* test, $P < 0.01$). To identify the group of nodal mutations that possesses comparable functionality values with those of disease-causing mutations, we used resampling simulations with sequential Bonferroni correction for multiple testing (see Materials and Methods). This approach revealed that for RNMs in RNA genes ($N = 67$), the upper 20th percentile (13 mutations), lost its statistical significance and therefore had functionality values (RNA mutations functionality scores) comparable with those of disease-causing mutations (supplementary table S6*A*, Supplementary Material online). Additionally, using the resampling simulations analysis, we detected that the upper 11th percentile subset of SNMs in RNA genes (25 mutations), lost its statistical significance making them the RNA genes' SNMs with functionality values most similar to those of disease-causing mutations (supplementary table S6*B*, Supplementary Material online).

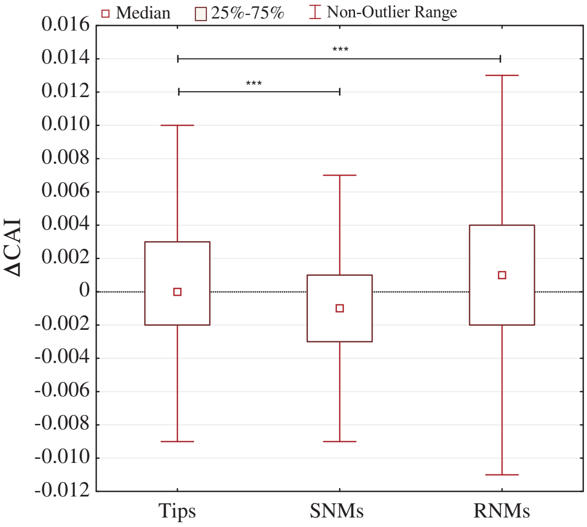

Fig. 2.—Comparison of functionality assessments of synonymous mutations. The distribution of ΔCAI values of synonymous mutations in the tree Tips, SNMs, and RNMs. ***Significant difference between the Tips to the SNMs or RNMs (Mann–Whitney *U* test, $P < 0.001$).

The "functional" RNMs Associate with Phenotypes

Taken together, 24 RNMs and 67 SNMs (nonsynonymous and RNA genes mutations) had comparable functionality values with those of disease-causing mutations (tables 1 and 2 and supplementary table S7, Supplementary Material online). For the sake of brevity, we refer to these as "functional" RNMs and SNMs. Our analysis indicated that RNMs were subjected to the strongest negative selection during human evolution. As mentioned earlier, we hypothesized that the RNMs with the highest functional potential are the best candidates to possess adaptive properties. An outcome of this possibility is that the very same RNMs could associate with diseases if the environment changes. Indeed, the environment and climate dramatically changed during the course of human evolution. Thus, we asked whether our discovered "functional" RNMs associate with known diseases from the one hand or were reported to be adaptive to certain environments from the other. Firstly, an inspection of the phylogenetic distribution of our identified 24 "functional" RNMs revealed their occurrence throughout human mtDNA phylogeny (fig. 3). These

Table 1

Summary of Nonsynonymous RNMs Events Having Functional Potential[a]

Mutation Number	mtDNA Mutation	Haplo-Group	Number of Sequences	Haplo-Group	Number of Sequences	Gene	Amino Acid Change	Conservation Index	SIFT Score	PANTHER *P* Deleterious	MutPred
1	T3394C	J1c1[b]	45	M9a[b]	101	*ND1*	Y30H	8	0.02	0.39	0.73
2	A3547G	B2[b]	57	within HV1b[b]	9	*ND1*	I81V	8	0.01	0.29	0.57
3	T3644C	Within M13[b]	11	D4h[b]	56	*ND1*	V113A	9	0.01	0.49	0.80
4[c]	A4917G	R1a[b] N1b2[b]	15 11	T[b]	446	*ND2*	N150D	7	0.06	0.79	0.58
5	G7697A	Within C1d[b] Within M9a[b]	21 26	Within M36[b]	6	*COX2*	V38I	9	0.02	0.60	0.60
6	T8843C	H45[b] Within H2a5[b]	7 8	Within A11[b]	8	*ATP6*	I106T	7	0	0.80	0.69
7	A10086G	L3b[b]	53	Within W1[b]	8	*ND3*	N10D	8	0.01	0.53	0.16
8	A11084G	Within M2a1[b]	5	Within M7a[b]	16	*ND4*	T109A	9	0	0.59	0.49
9	G11969A	C4[b]	107	M11[b]	12	*ND4*	A404T	7	0.04	0.91	0.70
10	C13129T	Within L0d[b]	8	N1b2[b]	11	*ND5*	P265S	7	0.03	0.90	0.76
11	G15119A	Within N5[b]	5	Within C1b[b]	8	*CYTB*	A125T	9	0.04	0.44	0.53
12	G15257A	J2[b]	106	Within K1b1[b]	20	*CYTB*	D171N	7	0.02	0.52	0.72

[a] Summary of all RNM events detected in the phylogenetic analysis bearing functional potential (see Materials and Methods).

[b] Group of sequences with the same lineage nomenclature; "within"—the mutational event occurred or became "nodal" inside a phylogenetic branch.

[c] Mutation number 4 did not pass the SIFT score cutoff (see main text).

Table 2

Summary of All RNA RNMs Events Having Functional Potential[a]

Mutation Number	mtDNA Mutation	Haplo-Group	Number of Sequences	Haplo-Group	Number of Sequences	Haplo-Group	Number of Sequences	Gene	ΔG Index	Conservation Index	RNA Score
13	G750A	H2a2[b]	91	L3e3/4/5[b]	34			*12S*	1	8	9
14	G951A	H2a1[b]	60	D3[b]	9			*12S*	7.2	4	11.2
15	T1243C	W[b]	141	L0K[b]	7			*12S*	7.2	1	8.2
16	G1719A	H7a[b] Within L3h[b] D6[b]	21 21 6	X2[b] Within D4h[b] N1[b]	114 9 189	M28[b] Within P[b] D4m[b]	8 5 5	*16S*	1	8	9
17	T2083C	Within M40[b]	8	Within U4b[b]	6			*16S*	3.4	6	9.4
18	T3027C	U5a1d[b]	13	E[b]	62			*16S*	3.2	6	9.2
19	G5821A	C7a[b]	15	M53[b]	11			*tRNA-Cys*	7.1	1	8.1
20	T7581C	Within D4j[b] Within H4a[b]	23 5	Within U1[b]	22	N2a[b]	5	*tRNA-Asp*	7.3	6	13.3
21	G12236A	L2b/c[b]	43	Within L5[b]	9			*tRNA-Ser(AGY)*	9	3	12
22	A14693G	Y[b]	19	Within M2a1[b]	12			*tRNA-Glu*	3.5	6	9.5
23	G15927A	B5b[b] Within HV1a[b]	14 8	Within G3[b] Within U6a[b]	7 25	X2b[b]	51	*tRNA-Thr*	6.9	6	12.9
24	G15928A	T[b]	442	Within Z[b]	9	Within M35[b]	18	*tRNA-Thr*	8.3	1	9.3
25	A15951G	D4b1[b] + D3[b]	33	Within L0d[b]	8			*tRNA-Thr*	1.9	7	8.9

[a] Summary of all RNM events detected in the phylogenetic analysis bearing functional potential (see Materials and Methods).

[b] Group of sequences with the same lineage nomenclature; "within"—the mutational event occurred or became "nodal" inside a phylogenetic branch.

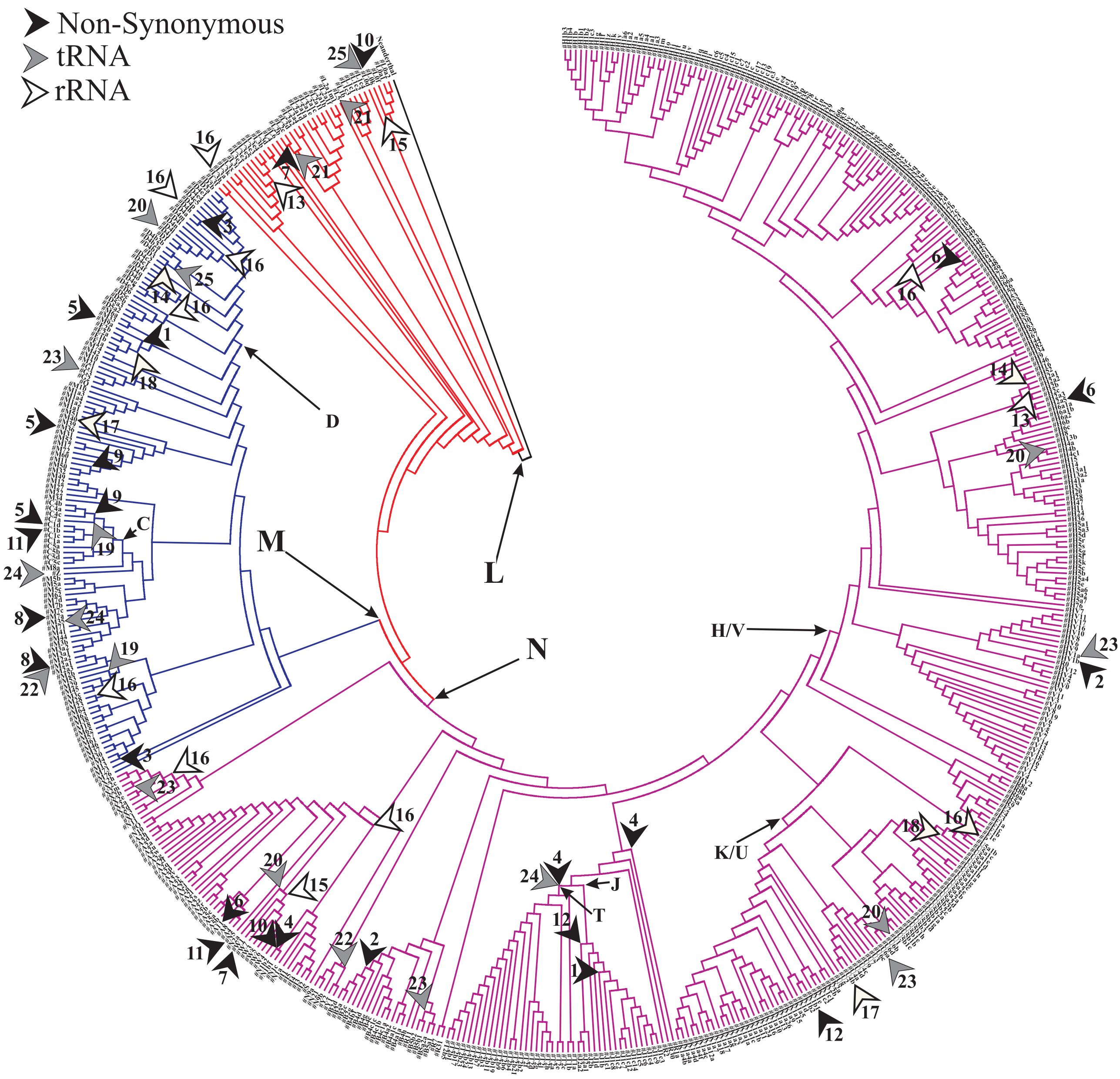

Fig. 3.—Distribution of recurrent nodal mutational events with functional potential across human mtDNA phylogeny. A neighbor-joining phylogenetic tree of whole human mtDNA sequences was created using MEGA5 (see Materials and Methods). The tree was generated from 563 ancestral sequences representing 9,868 sequences clustered according to previously published nomenclature (Behar, et al. 2012; van Oven and Kayser 2008). The mutation numbers are correlated with those listed in tables 1–3 and figure 4. Arrowheads indicate the branch at which the mutational events occurred. Arrowheads pointing to the terminal nodes refer to mutational events that lie in the stem of a subnode of the indicated ancestral sequence.

RNMs are also found throughout the human mitochondrial genome, albeit with no clear preference for particular genes (fig. 4). Second, we screened the literature and found reports showing that lineages harboring our identified RNMs were associated with altered tendencies to develop genetic disorders (table 3). Some lineages that share such RNMs were also associated with the same phenotypes. Of particular interest is the T3394C variant (table 1, mutation number 1), identified as nodal in lineages J1c1 and M9a. These lineages were independently identified as modifying the phenotypic expression of Leber Hereditary Optic Neuropathy (LHON) (Carelli et al. 2006; Liang et al. 2009; Zhang et al. 2010). Moreover, recent studies have shown that the T3394C variant is enriched in haplogroup M9 within the high altitude Tibetan population (Gu et al. 2012; Ji et al. 2012) and affects mitochondrial function in cybrids (Ji et al. 2012).

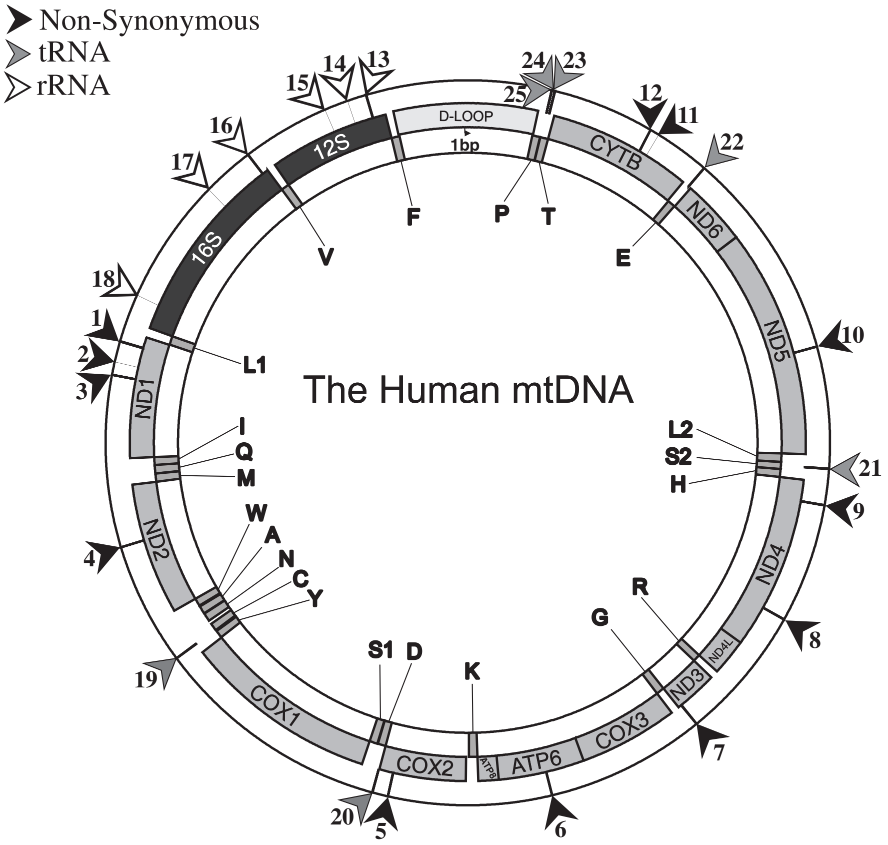

Fig. 4.—Distribution of RNM events with functional potential across human mtDNA. The mutation numbers are correlated with those listed in tables 1–3 and figure 3. The bars in the inner circle represent the locations of tRNA genes, whereas the bars in the middle circle represent the locations of the rRNA and protein-coding genes. The stripes in the outer circle and the arrowhead mark the location of the mutational events.

Discussion

In this study, we conducted a comprehensive analysis of the entire human mtDNA phylogeny and identified 188 RNMs and 518 SNMs that were either nonsynonymous ($N = 121$ and $N = 283$, respectively) or RNA gene mutations ($N = 67$ and $N = 235$, respectively). As intuitively expected both RNMs and SNMs, exhibited significantly different functional attributes from disease-causing mutations. Strikingly, 24 RNMs (11 nonsynonymous and 13 RNA mutations) and 67 SNMs (42 nonsynonymous, 25 RNA mutations) displayed comparable functionality values with those of disease-causing mutations. This implies the strong functional potential of these particular nodal mutations.

While analyzing nodal synonymous mutations in the human phylogeny, we identified preference of mtDNA protein-coding genes for codons recognized by mtDNA-encoded tRNAs rather than imported tRNAs. This most likely reflects adaptation toward using more translational efficient codons in the mitochondria. However, no apparent trend was observed in the distribution of the functional potential values of synonymous mutations (ΔCAI) as compared between all types of nodal mutations (SNMs and RNMs) and tree tips mutations. Therefore, assessment of the functional potential of nodal synonymous mutations still awaits the identification of disease causing synonymous mtDNA mutations for comparison.

Previously, several studies including those performed by us showed that nearly a quarter of the nodal mutations in the human mitochondrial phylogeny that occurred within the coding mtDNA region had similar characteristics to disease-causing mutations, and thus reflecting positive selection (Ruiz-Pesini et al. 2004). Since nodal mutations, in general, and functional RNMs or SNMs, in particular, have been retained in the human population over a prolonged period, it is logical to assume that they survived the effects of natural selection. Although mtDNA disease-causing mutations recur multiple times independently in unrelated families, such mutations mostly associate with the terminal tips of the phylogenetic tree, most likely due to strong negative selection. This is in sharp contrast to the "functional" RNMs and SNMs. Although such mutations share many characteristics with disease-causing mutations, they are detected within the stems of certain

Table 3
Summary of All RNM Association to Phenotype Events Having Functional Potential[a]

Mutation Number	mtDNA Mutation	Haplo-Groups	Phenotype	References
1	T3394C	J1c1[b] M9a[b]	1. LHON in haplogroups M9 and J1c1 2. Adaptation to high altitude in M9	1. Carelli et al. (2006); Liang et al. (2009); Zhang et al. (2010) 2. Gu et al. (2012); Ji et al. (2012)
3	T3644C	Within M13[b] D4h[b]	Bipolar disorder in patients vs. controls	Munakata et al. (2004)
4[c]	A4917G	R1a[b] T[b] N1b2[b]	1. Age-related macular degeneration in patients vs. controls/reduced sperm motility in haplogroup T 2. Coronary-artery disease and diabetic retinopathy association haplogroup T 3. Altered susceptibility to the common complications of T2DM in haplogroup N1b1 (N1b2)	1. Canter et al. (2008) 2. Ruiz-Pesini et al. (2000) 3. Feder et al. (2008); Kofler et al. (2009)
19	G5821A	C7a[b] M53[b]	Enhanced penetrance of hearing loss in Chinese families	Lu et al. (2010)
5	G7697A	Within C1d[b] Within M36[b] Within M9a[b]	Hypertrophic cardiomyopathy (HCM) in patients vs. controls in Chinese families	Wei et al. (2009)
7	A10086G	L3b[b] Within W1[b]	African Americans with hypertension-associated end-stage renal disease in patients vs. controls	Watson et al. (2001)
8	A11084G	Within M2a1[b] Within M7a[b]	1. Mitochondrial Encephalopathy Lactic Acidosis and Stroke-like Episodes (MELAS) 2. Parkinson's disease (PD) in Japanese individuals	1. Lertrit et al. (1992) 2. Takasaki (2009)
21	G12236A	L2b/c[b] Within L5[b]	Maternally inherited nonsyndromic hearing impairment in haplogroup H (The G12236A mutation)	Leveque et al. (2007)
22	A14693G	Y[b] Within M2a1[b]	1. Modulating the phenotypic manifestation of deafness-associated A1555G mutation in Chinese families. 2. Modulating the manifestation of LHON G3460A mutation in a Chinese family. 3. Associates with MELAS.	1. Ding et al. (2009); Lu et al. (2010) 2. Tong et al. (2007) 3. Tzen et al. (2003)

(continued)

Table 3 Continued

Mutation Number	mtDNA Mutation	Haplo-Groups	Phenotype	References
12	G15257A	J2[b] Within K1b1[b]	1. Association of mitochondrial haplogroup J2 with longevity and reduced mtDNA oxidative damage in high altitude Pyrenees Mountains population. 2. Association of G15257A with LHON.	Brown et al. (1992); Dominguez-Garrido et al. (2009); Niemi et al. (2003)
23	G15927A	B5b[b] Within G3[b] X2b[b] Within U6a[b] Within HV1a[b]	1. Enhanced penetrance of hearing loss in Chinese families with haplogroup B5b and the pathological mutation A1555G. 2. Haplogroup B5b over representation in Japanese centenarians, Parkinson's and type 2 diabetic patients. 3. Haplogroup X2 is associated with centenarians in the Amish.	1. Chen et al. (2008); Wang et al. (2008) 2. Takasaki (2009) 3. Courtenay et al. (2012)
24	G15928A	T[b] Within Z[b] Within M35[b]	1. Association with idiopathic repeated pregnancy loss. 2. Protection against Alzheimer Disease in French Canadians.	1. Seyedhassani et al. (2010) 2. Chagnon et al. (1999)
25	A15951G	D4b1[b] + D3[b] within L0d[b]	Influences phenotypic expression of LHON-associated G11778A mutation in a Chinese family	Li et al. (2006)

[a] Summary of all RNM events detected in the phylogenetic analysis bearing functional potential (see Materials and Methods).

[b] Group of sequences with the same lineage nomenclature; "within"—the mutational event occurred or became "nodal" inside a phylogenetic branch.

[c] Mutation number 4 did not pass the SIFT score cutoff (see main text).

phylogenetic branches, survived natural selection and thus have likely been positively selected (Ruiz-Pesini et al. 2004; Ruiz-Pesini and Wallace 2006). Nevertheless, we cannot exclude the possibility that some RNMs or SNMs survived natural selection because they were only mildly deleterious and had only little effect on fitness.

The functional RNMs caught our attention, because they could, in principle, be involved in events of convergent evolution. As the functional RNMs are positively selected and hence adaptive, they likely played a role in human survival in the face of past environmental conditions. As the environment and life style of modern humans have dramatically changed with time and as modern humans reach much older ages than did our ancestors, it is expected that some functional RNMs would alter susceptibility to disease or age-related phenotypes. Indeed, the identities of some functional RNMs support this hypothesis, as these RNMs reside in the stem of unrelated mtDNA haplogroups that independently associate with altered susceptibility to complex and mitochondrial genetic disorders. Interestingly, our literature search revealed that in some cases the mtDNA haplogroups that share RNMs also alter the susceptibility to the same diseases (table 3). As the RNMs are the only "functional" mutations shared among these haplogroups, it is tempting to suggest that these RNMs played important roles in the molecular basis of these phenotypes.

If certain RNMs associate with altered susceptibility to the same phenotypes in unrelated lineages in modern times, then one can ask whether it is possible that these very mutations reflect similar adaptive properties of these unrelated lineages, namely reflecting convergent adaptive evolution? Naturally, disease association underlines the functional properties of a given mutation but does not readily lend clues to the exact adaptive role of that mutation in our phylogenetic history. Moreover, association with age-related diseases is typically invisible to selection due to the onset of such conditions after reproductive age. Previously, we showed that certain nodal mutations in our phylogenetic history associated with the ability of our ancestors to survive in different climatic conditions, mainly due to an altered balance between heat and ATP production, that is, coupling efficiency (Mishmar et al. 2003). As "functional" RNMs occur in unrelated lineages from different ethnicities that frequently inhabit different parts of the globe, it is less obvious how to assess shared adaptive properties. Nevertheless, recent studies have shown that the mtDNA variant, T3394C (a "functional" RNM, which is nodal in lineages J1c1 and M9a; table 1, mutation number 1), is enriched in haplogroup M9 within the high altitude Tibetan population (Gu et al. 2012; Ji et al. 2012) and affects mitochondrial function in cybrids (Ji et al. 2012). As mentioned earlier, the same RNM (T3394C) associated with increased risk for LHON. Hence, RNMs that played adaptive role during the phylogenetic history of our kind alter susceptibility to disease in modern times, given that environmental conditions have since changed dramatically.

Similar to the T3394C RNM, the RNM G15927A (table 2, mutation number 23) that was identified as nodal in lineages B5b, a sub-lineage of G3, a sub-lineage of U6a, a sub-lineage of HV1a, and X2b associated with longevity in the frame of two of these lineages, that is, B5b (Takasaki 2009) and X2 (Courtenay et al. 2012). However, low resolution of the X2 haplogroup assignment in the latter study prevented us from drawing more conclusive phenotypic association of the G15927A variant. Thus, to better understand the functional potential of nodal mutations (both RNMs and SNMs) future mtDNA disease-association studies will benefit from higher resolution of haplogroup assignment.

Clearly, mtDNA variants are not sufficient to cause a phenotype, mainly because mitochondrial functions operate via mitochondrial–mitochondrial and nuclear–mitochondrial epistatic interactions, thus calling for modifying compensatory mutations (Hudson et al. 2005). As such, it is predicted that mutations with functional properties will associate with different phenotypes or with varying phenotypic severity/expression in different genetic backgrounds. Functional RNMs are no exception in this matter. The mtDNA RNM A4917G (table 1, mutation 4) that was identified as nodal in the mtDNA haplogroups T and N1b2 associated with two different phenotypes. Specifically, haplogroup T associated with reduced sperm motility (Ruiz-Pesini et al. 2000), and haplogroup N1b2 (sometimes termed N1B1), associated with altered susceptibility to common complications of type 2 diabetes mellitus (Feder et al. 2008). Hence, although exhibiting clear characteristics of functional potential, some "functional" mtDNA RNMs are instead expected to exhibit their phenotypic effect only in combination with other factors, suggesting variability in the degree of functionality among RNMs.

Our detailed analysis of RNMs over the entire human mtDNA phylogeny, including those with functional potential, became possible due to a novel in-house-developed bioinformatics tool entitled FuRNED (Functional Recurrent Nodal Events Detector). This tool associated genetic variants with specific lineages in the human mtDNA phylogeny and assessed the functional potential of coding region mutations. The scripts of this tool are available as supplementary data, Supplementary Material online.

Conclusions

In summary, we have analyzed and identified the largest set of rare and common variants in the human mtDNA phylogeny to date, including single and recurrent nodal variants (SNMs and RNMs, respectively). A novel in-house developed bioinformatics tool enabled identification of SNMs and RNMs that exhibit similar characteristics as do disease-causing mutations, yet unlike the latter, survived natural selection over time and hence, are likely adaptive. As some of the functional RNMs became independently nodal in unrelated mtDNA lineages that associate with the same traits and diseases, it is possible that such mutations are major players in assigning the traits in question, thus providing a novel explanation for the molecular basis of such phenotypes. From the evolutionary perspective, these RNMs potentially share common functionality in unrelated lineages and thus constitute the best candidates to play an adaptive role in a convergent manner during human phylogentic history. To our knowledge, this is the most comprehensive analysis of selective signatures within human mtDNA-encoded RNA and protein genes. For the first time, we discover virtually all positively selected SNMs and RNMs in our phylogeny while emphasizing their dual role in evolution and disease etiology today.

Supplementary Material

Supplementary data and figures S1–S4 and tables S1–S7 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Israeli Science Foundation grant 387/08 to D.M and a partial Kreitman scholarship for excellent graduate students (BGU) awarded to L.L.

Literature Cited

Bar-Yaacov D, Blumberg A, Mishmar D. 2012. Mitochondrial-nuclear coevolution and its effects on OXPHOS activity and regulation. Biochim Biophys Acta. 1819:1107–1111.

Behar DM, et al. 2012. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet. 90:675–684.

Briggs AW, et al. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325:318–321.

Brown MD, et al. 1992. Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. Genetics 130: 163–173.

Canter JA, et al. 2008. Mitochondrial DNA polymorphism A4917G is independently associated with age-related macular degeneration. PLoS One 3:e2091.

Carelli V, et al. 2006. Haplogroup effects and recombination of mitochondrial DNA: novel clues from the analysis of Leber hereditary optic neuropathy pedigrees. Am J Hum Genet. 78:564–574.

Castellana S, Vicario S, Saccone C. 2011. Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes. Genome Biol Evol. 3:1067–1079.

Chagnon P, et al. 1999. Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population. Am J Med Genet. 85:20–30.

Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet. 7:98–108.

Chen B, et al. 2008. Mitochondrial ND5 T12338C, tRNA(Cys) T5802C, and tRNA(Thr) G15927A variants may have a modifying role in the phenotypic manifestation of deafness-associated 12S rRNA A1555G mutation in three Han Chinese pedigrees. Am J Med Genet A. 146A: 1248–1258.

Courtenay MD, et al. 2012. Mitochondrial haplogroup X is associated with successful aging in the Amish. Hum Genet. 131:201–208.

Ding Y, et al. 2009. Mitochondrial tRNA(Glu) A14693G variant may modulate the phenotypic manifestation of deafness-associated 12S rRNA A1555G mutation in a Han Chinese family. J Genet Genomics. 36: 241–250.

Dominguez-Garrido E, et al. 2009. Association of mitochondrial haplogroup J and mtDNA oxidative damage in two different North Spain elderly populations. Biogerontology 10:435–442.

Feder J, et al. 2008. Differences in mtDNA haplogroup distribution among 3 Jewish populations alter susceptibility to T2DM complications. BMC Genomics 9:198.

Gershoni M, et al. 2010. Coevolution predicts direct interactions between mtDNA-encoded and *n*DNA-encoded subunits of oxidative phosphorylation complex I. J Mol Biol. 404:158–171.

Gershoni M, Templeton AR, Mishmar D. 2009. Mitochondrial bioenergetics as a major motive force of speciation. Bioessays 31:642–650.

Glaser F, et al. 2003. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19:163–164.

Good PI. 2006. Resampling methods: a practical guide to data analysis. Boston: Birkhauser.

Gu M, et al. 2012. Differences in mtDNA whole sequence between Tibetan and Han populations suggesting adaptive selection to high altitude. Gene 496:37–44.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52: 696–704.

Hudson G, et al. 2005. Identification of an X-chromosomal locus and haplotype modulating the phenotype of a mitochondrial DNA disorder. Am J Hum Genet. 77:1086–1091.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol. 2:13–34.

Ji F, et al. 2012. Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and high-altitude Tibetans. Proc Natl Acad Sci U S A. 109:7391–7396.

Kazuno AA, et al. 2006. Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics. PLoS Genet. 2:e128.

Kofler B, et al. 2009. Mitochondrial DNA haplogroup T is associated with coronary artery disease and diabetic retinopathy: a case control study. BMC Med Genet. 10:35.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 4:1073–1081.

Lertrit P, et al. 1992. A new disease-related mutation for mitochondrial encephalopathy lactic acidosis and strokelike episodes (MELAS) syndrome affects the ND4 subunit of the respiratory complex I. Am J Hum Genet. 51:457–468.

Leveque M, et al. 2007. Whole mitochondrial genome screening in maternally inherited non-syndromic hearing impairment using a microarray resequencing mitochondrial DNA chip. Eur J Hum Genet. 15: 1145–1155.

Li B, et al. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25:2744–2750.

Li R, et al. 2006. The mitochondrial tRNA(Thr) A15951G mutation may influence the phenotypic expression of the LHON-associated ND4 G11778A mutation in a Chinese family. Gene 376:79–86.

Liang M, et al. 2009. Leber's hereditary optic neuropathy is associated with mitochondrial ND1 T3394C mutation. Biochem Biophys Res Commun. 383:286–292.

Lu J, et al. 2010. Mitochondrial haplotypes may modulate the phenotypic manifestation of the deafness-associated 12S rRNA 1555A>G mutation. Mitochondrion 10:69–81.

Mercer TR, et al. 2011. The human mitochondrial transcriptome. Cell 146: 645–658.

Miller MP, Kumar S. 2001. Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet. 10:2319–2328.

Mishmar D, et al. 2003. Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A. 100:171–176.

Mishmar D, Zhidkov I. 2010. Evolution and disease converge in the mitochondrion. Biochim Biophys Acta. 1797:1099–1104.

Moreno-Loshuertos R, et al. 2006. Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants. Nat Genet. 38: 1261–1268.

Munakata K, et al. 2004. Mitochondrial DNA 3644T–>C mutation associated with bipolar disorder. Genomics 84:1041–1050.

Niemi AK, et al. 2003. Mitochondrial DNA polymorphisms associated with longevity in a Finnish population. Hum Genet. 112:29–33.

Pacheu-Grau D, et al. 2011. 'Progress' renders detrimental an ancient mitochondrial DNA genetic variant. Hum Mol Genet. 20: 4224–4231.

Pereira L, Soares P, Radivojac PLB, Samuels DC. 2011. Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity. Am J Hum Genet. 88:433–439.

Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 26:1641–1650.

Rice TK, Schork NJ, Rao DC. 2008. Methods for handling multiple testing. In: Rao DC, Gu C, editors. Advances in genetics. Academic Press. p. 293–308.

Ruiz-Pesini E, et al. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Res. 35:D823–D828.

Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. Science 303:223–226.

Ruiz-Pesini E, Wallace DC. 2006. Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA. Hum Mutat. 27:1072–1081.

Ruiz-Pesini E, et al. 2000. Human mtDNA Haplogroups associated with high or reduced spermatozoa motility. Am J Hum Genet. 67:682–696.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406–425.

Salinas T, et al. 2012. Co-evolution of mitochondrial tRNA import and codon usage determines translational efficiency in the green alga *Chlamydomonas*. PLoS Genet. 8:e1002946.

Seyedhassani SM, et al. 2010. The point mutations of m itochondrial tRNA. Iran J Reprod Med. 8:45–50.

Sharp PM, Li WH. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281–1295.

Simon JL. 1993. Resampling: the new statistics. Boston: Wadsworth.

Staple DW, Butcher SE. 2005. Pseudoknots: RNA structures with diverse functions. PLoS Biol. 3:e213.

Suissa S, et al. 2009. Ancient mtDNA genetic variants modulate mtDNA transcription and replication. PLoS Genet. 5:e1000474.

Takasaki S. 2009. Mitochondrial haplogroups associated with Japanese centenarians, Alzheimer's patients, Parkinson's patients, type 2 diabetic patients and healthy non-obese young males. J Genet Genomics. 36:425–434.

Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.

Templeton AR. 1996. Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. Genetics 144:1263–1270.

Thomas PD, et al. 2003. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13:2129–2141.

Thusberg J, Olatubosun A, Vihinen M. 2011. Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat. 32:358–368.

Tong Y, et al. 2007. The mitochondrial tRNA(Glu) A14693G mutation may influence the phenotypic manifestation of ND1 G3460A mutation in a Chinese family with Leber's hereditary optic neuropathy. Biochem Biophys Res Commun. 357:524–530.

Tzen CY, Thajeb P, Wu TY, Chen SC. 2003. Melas with point mutations involving tRNALeu (A3243G) and tRNAGlu(A14693g). Muscle Nerve 28:575–581.

van Oven M, Kayser M. 2008. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat. 30: E386–E394.

Wallace DC. 2005. The mitochondrial genome in human adaptive radiation and disease: on the road to therapeutics and performance enhancement. Gene 354:169–180.

Wang X, et al. 2008. Mitochondrial tRNAThr G15927A mutation may modulate the phenotypic manifestation of ototoxic 12S rRNA A1555G mutation in four Chinese families. Pharmacogenet Genomics. 18:1059–1070.

Watson B Jr, Khan MA, Desmond RA, Bergman S. 2001. Mitochondrial DNA mutations in black Americans with hypertension-associated end-stage renal disease. Am J Kidney Dis. 38:529–536.

Wei YL, et al. 2009. Novel mitochondrial DNA mutations associated with Chinese familial hypertrophic cardiomyopathy. Clin Exp Pharmacol Physiol. 36:933–939.

Wright F. 1990. The "effective number of codons" used in a gene. Gene 87:23–29.

Zhang M, et al. 2010. Mitochondrial haplogroup M9a specific variant ND1 T3394C may have a modifying role in the phenotypic expression of the LHON-associated ND4 G11778A mutation. Mol Genet Metab. 101: 192–199.

Zhidkov I, Livneh EA, Rubin E, Mishmar D. 2009. mtDNA mutation pattern in tumors and human evolution are shaped by similar selective constraints. Genome Res. 19:576–580.

Associate editor: Judith Mank