

MitoLSDB: The Human Mitochondrial Locus Specific Database

Shamna Mole¹, Saakshi Jalali², Vinod Scaria², Anshu Bhardwaj*³

¹Stella Maris College, University of Madras, Chennai, India

²Genomics and Molecular Medicine, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110007, India

³Council of Scientific and Industrial Research (CSIR), 2 Rafi Marg, Delhi 110001, India

Email: Shamna Mole - shem8916@gmail.com; Saakshi Jalali - saakshi.jalali@igib.res.in; Vinod Scaria - vinods@igib.in; Anshu Bhardwaj* - anshu@csir.res.in, anshub@osdd.net;

*Corresponding author

Abstract

Background: Human mitochondrial DNA (mtDNA) encodes a set of 37 genes which are essential structural and functional components of the electron transport chain. Variations in these genes have been implicated in a broad spectrum of diseases and are extensively reported in literature and various databases. In this study, we describe MitoLSDB, an integrated platform to catalogue disease association studies on mtDNA (<http://mitolsdb.igib.res.in>). The main goal of MitoLSDB is to provide a central platform for direct submissions of novel variants that can be curated by the Mitochondrial Research Community. Description: MitoLSDB provides access to standardized and annotated data from literature and databases encompassing information from 5231 individuals, 675 populations and 27 phenotypes. This platform is developed using the Leiden Open (source) Variation Database (LOVD) software. MitoLSDB houses information on all 37 genes in each population amounting to 132397 variants, 5147 unique variants. For each variant its genomic location as per the Revised Cambridge Reference Sequence, codon and amino acid change for variations in protein-coding regions, frequency, disease/phenotype, population, reference and remarks are also listed. MitoLSDB curators have also reported errors documented in literature which includes 94 phantom mutations, 10 NUMTs, six documentation errors and one artefactual recombination. Conclusion: MitoLSDB is the largest repository of mtDNA variants systematically standardized and presented using the LOVD platform. We believe that this is a good starting resource to curate mtDNA variants and will facilitate direct submissions enhancing data coverage, annotation in context of pathogenesis and quality control by ensuring

non-redundancy in reporting novel disease associated variants.

Keywords

MitoLSDB, Human mitochondrial genome variation, Errors detected in GenBank

Background

Mitochondria are the essential energy-generating organelles in eukaryotes possessing the oxidative phosphorylation system (OXPHOS). Mitochondrial disorders are caused by mutations in mitochondrial genes encoded by nuclear or mitochondrial DNA (mtDNA) [1]. The OXPHOS comprises of five protein complexes and majority of their protein subunits are nuclear encoded with only a subset of 37 genes encoded by mtDNA [2]. Of these 37 genes, 13 are protein subunits, 22 tRNAs, and 2 rRNAs. These genes are essential components of electron transport chain complexes I, III and IV and complex V (ATP synthase) [3]. Mutations in these genes have been associated with a broad spectrum of diseases [4]. At least 1 in every 200 births is thought to have a potentially pathogenic mitochondrial DNA mutation [5]. The disease phenotypes attributed to mutations in mtDNA have diverse and overlapping symptoms and also multi-organ involvement [6]. Many deleterious point mutations have been identified to date, the most frequent ones being the m.3243A>G MELAS mutation [7], the LHON primary mutations [8], and the m.8344A>G MERRF mutation [9]. Others are found less often, while still others have been described only as case studies or in families. The investigation of pathogenic mtDNA mutations has revealed a complex relation between patient genotype and phenotype [10]. The phenotypic variability is due to the peculiarities of mitochondrial properties, such as heteroplasmy, different mutation rates in different tissues and highly polymorphic nature [11-13]. Therefore, the patho-mechanisms of mtDNA point mutations are still not very well understood. Furthermore, there appears to be a class of slightly deleterious mutations that modify the risks of developing certain complex diseases or traits [14]. Besides, heteroplasmic and homoplasmic mtDNA have also been observed along with large number of basal polymorphisms in the mitochondrial genome across databases like OMIM [<http://www.ncbi.nlm.nih.gov/omim>] [15], MitoMap [16], Mitovariome [17] and mtDB [18]. These facts highlight the challenges in assessing the role of mtDNA variants in diseases or phenotypes.

Recent reports indicate role of mitochondrial dysfunction in the pathogenesis of or influence the risk of diseases such as Alzheimer's, Parkinson's, cardiovascular disease including cardiomyopathy, etc. [19-21]. But the genotype-phenotype relationship is unclear and debatable [22, 23]. More than 5000 complete or

coding-region sequences of publicly available mtDNA were analyzed to study the diversity of the global human population [24]. This study has generated useful data in the form of all possible transitions and transversions and their analysis lead to interesting observations that may help in understanding the role of mtDNA variants in disease. Besides, there has been an increase in DNA variant data resulting from new automatic sequencing technologies [25]. Thus, it is imperative to catalogue this information on a standard web-based platform for sharing and evaluating the potential pathological effects of mtDNA variants. To this end, we have used the Leiden Open (source) Variation Database (LOVD) Software [26, 27] for creating a catalog of human mtDNA variants, through manual curation of data from literature and from public databases. LOVD is a commonly used tool for organizing locus-centric variation data. As of now, MitoLSDB has patient and variant information from 5231 individuals from 675 different populations [24, 28] from 27 different groups including patients with Alzheimer’s disease, Asthanozoospermic, Atypical psychosis, Breast cancer, Diabetes, Angiopathy, Deafness, Glioma, Parkinson’s disease, Teratozoospermic, Thyroid cancer, etc and can be accessed at <http://mitolsdb.igib.res.in>. MitoLSDB is a Locus-Specific DataBase (LSDB) for human mtDNA genes and provides access to standardized and annotated data compiled from different resources which are otherwise difficult to search and comprehend. This is in line with the objectives of LSDBs which are expected to contain comprehensive information from disparate resources and are open for direct submissions. It has also been observed that a large amount of variant data from case studies or reports never get published and LSDBs have served as a viable platform for the scientific community to benefit from and actively contribute to [29]. For each variant curated in MitoLSDB, its genomic location as per the Revised Cambridge Reference Sequence, codon and amino acid change for variations in protein-coding regions, frequency, disease/phenotype, population, reference and remarks are also listed. MitoLSDB curators have also reported errors documented in literature which includes 94 phantom mutations, 10 NUMTs (Nuclear mitochondrial DNA sequences), six documentation errors and one artefactual recombination. We believe that this is a good starting resource to curate mtDNA variants and will facilitate direct submissions enhancing data coverage, annotation in context of pathogenesis and quality control by ensuring non-redundancy in reporting novel disease associated variants.

Data Collection and Integration

Data Collection

The variant data and other patient information of 5139 individuals from different populations were obtained from the study by Pereira et al [24] and the public databases www.phylotree.org [30] and Ian Logan’s website

<http://www.ianlogan.co.uk/checker/genbank.htm> [31] which belongs to 26 different groups and a set of 92 complete genomes from sporadic ataxia patients [28]. A set of PERL scripts were developed to extract variant data from the sources. The dataset obtained from Pereira et al's study [24] gives the details on sample ID, variant, reference, haplotype reported, origin/ethnicity and phenotype. However, this dataset only provides variant positions in each sample. This information was complemented with variant detail with help of other resources. These variants are reported as per the coordinates of the revised rCRS (Revised Cambridge Reference Sequence) [GI:251831106] . We have also reported errors documented in literature which includes 94 phantom mutations [31], 10 NUMTs [32], 6 documentation errors and one artefactual recombination [31] in the remarks section of the database. This data is converted to match the 'import file' specifications of LOVD.

The Database

The database is customized on the LOVD platform which is supported on the backend by a MySQL relational database management system. Links are provided to genes to assist the user in searching detailed information related to the gene. In addition, plug-ins have been created to export the data to a standard meta-tagged format for interoperability with other resources. This would aid the user to have a genome centered and holistic view of the variants and this would be helpful in interpreting the biological impact of variations. In human mtDNA there are five instances of overlapping bases among genes and thus these have been mapped to both genes. Codon assignments for mtDNA are different from the universal genetic code and thus the alternate codon table is utilized for reporting codon changes [33]. The database provides for each variant information on its genomic location, gene name, frequency, phenotype, tissue, sample information, methodology, codon and amino acid change for protein variation, variant submission link, advance search options and registration guidelines for a new submitter.

Results and Discussion

MitoLSDB comprise of variants data from 5231 individuals from 675 populations which belong to 27 different categories [Figure 1 and Supplementary Table 1]. The base changes reported as per the genomic coordinates of rCRS so that this data can be compared easily with other datasets. Overall, 132397 variants are catalogued in MitoLSDB. Of these 5147 are unique genomic variants, wherein 4226 belong to protein-coding genes, 538 to rRNA genes and 383 are tRNA variants. Of the 4226 protein-coding variants, 158 are nucleotide ambiguities. In the remaining 4068 protein coding variants, 1349 and 2719 are non-synonymous

and synonymous changes, respectively with 1066, 528 and 2474 at first, second and third codon positions, in that order. Presence of variation at specific codon position may also be related to the strength of association with the phenotype. In disease association studies, variations occurring with high frequency in patients as compared to normal individuals are considered to be disease associated. For most of the disease phenotypes included in MitoLSDB, there is no information on the normal population variants and hence it is not possible to report disease association based on frequency differences. We have instead reported the frequency of each variant within each population that may assist in evaluating the pathogenic status of these variants during subsequent data analyses.

A closer look at the data highlights that MT-ATP8 shows the maximum number of non-synonymous variations after normalization for gene length. Similarly, MT-ND6 shows the maximum number of synonymous changes [Figure 2]. However, MT-ND2 and MT-ND4L harbor least number of synonymous changes and non-synonymous changes, respectively.

MT-CYB shows the maximum number of polymorphisms with frequency one, which is 2048 in number. For example, the variant m.15326A>G is seen in all the 77 samples from Finland CADASIL population and there are many more variants captured with frequency one. m.8860A>G, m.750A>G, m.15326A>G are some of the variants seen repeatedly in different samples from various populations. This highlights the systemic involvement of these mitochondrial variants in diseases or phenotypes. The statistics of base changes shows a clear skew towards transitions (4293) being more common as compared to transversions (575) [Figure 3]. The A>G transitions are most common, while G>T transversions are least frequent. As stated earlier, variations are more frequent at the third position in codons as compared to second position.

Errors Detected and Reported

We have reported a number of errors documented in literature for the datasets integrated in MitoLSDB. These errors include 94 phantom mutations, 10 NUMT (Nuclear mitochondrial DNA sequences) contaminations, 6 documentation errors, and one sequence with artefactual recombination. Data reported by Pereira et al [24] are directly retrieved from GenBank. It has been observed that many mtDNA sequences available in GenBank are reported errors and unintended mistakes [31, 34]. Many of these errors have already been reported in literature or sometimes even corrected by the authors. Unfortunately in several instances the new corrected versions of sequences have not been updated in GenBank [31]. These documentation errors include missing variants, phantom mutations and artefactual recombinants that may lead to wrong conclusions. Missing variants are those that are expected in a particular mtDNA haplotype according to its haplogroup

status. For example, the sequence [GenBank:DQ826448] lacks an additional nine expected variants to group that sequence into haplogroup M7b1. Phantom mutations are defined by the exclusive presence of a rare transversion [31]. These are systematic artefacts generated in the course of the sequencing process. The amount of artefacts depends not only on the automated sequencer and sequencing chemistry employed, but also on other lab-specific factors [35] and it is also observed that the pattern of phantom mutations differs significantly from that of natural mutations [36]. In particular, phantom mutation hotspots could lead to spurious mapping of somatic mutations and to misinterpretations in clinical mtDNA studies [1]. Another type of error reported in GenBank mtDNA sequences is the NUMTs. These are the mitochondrial DNA sequences in the nuclear genome [37] (nuclear mitochondrial pseudogenes) which on accidental amplification can pose a serious problem for mitochondrial disease studies [32]. Primers designed for amplification of mtDNA can potentially anneal with sequences in nuclear genome that present at high homology to mtDNA. In fact NUMTs have already been mistaken as heteroplasmic positions in the case of reported association of mutations in MT-CO1 and MT-CO2 with development of Alzheimer's disease, which were later shown to be an artefact resulting from the accidental amplification of nuclear mitochondrial pseudo genes [38]. Studies on artefactual recombinations [39, 40] and various missing mutations [41, 42] have also been used to report the status of variants in the data sets used in MitoLSDB. For example, the mtDNA sequence [GenBank:DQ834258] may be a recombinant since it bears m.8701A>G (MT-ATP6) and m.9540T>C (MT-CO3), characteristic of non-N status; but this sequence was misclassified as haplogroup HV, due to artefactual recombination [31]. This sequence is reported as 'recombinant sequence' in the database remarks section.

Conclusion

MitoLSDB is a systematic compilation of variant information and is expected to facilitate the submission of novel variants by the users. This is proposed as a good starting resource to curate mitochondrial DNA variants, which would facilitate researchers in genotype-phenotype studies and also streamline the task of reporting novel mutations. It would also allow cross-comparison of different mtDNA association studies and help understand the molecular correlates of mitochondrial disease phenotypes, which otherwise is a very daunting and challenging task given the complexity of mitochondrial genetics. Variants are integrated in MitoLSDB in a standard updatable format, with a very user friendly interface [Figure 4]. We believe that MitoLSDB may work as a central repository for reporting novel pathogenic variants and provide a solution to documented issues in context of spurious reports and faulty conclusions on disease association status of mtDNA variants [43].

MitoLSDB is a freely accessible website that allows researchers to retrieve mitochondrial genome variation data on 5231 individuals from various populations. Unlike other available sources, users can browse and obtain the variation data gene wise. It also allows the user to list the variants based on patient origin. Contrasting to other existing resources MitoLSDB provides data on variants caused by insertions and deletions. The MitoLSDB curators do not report the missing mutation or haplogroup information in the first version of the database because of the ambiguities reported in the haplogroup status, which may lead the researcher to wrong conclusions. We have reported the available corrected errors in the database remarks column. It would be the best to get these ambiguities confirmed by the original authors.

Future Perspectives

To the best of our knowledge, MitoLSDB is the largest repository of mtDNA variants systematically standardized and presented using the LOVD platform. The curators have integrated data from 675 populations comprising of 5231 individuals and 5147 unique variants. We are attempting to make the data interoperable with various genomic databases and computational workflows, which would facilitate easy and automated analysis of the variants. This would facilitate researchers in genotype-phenotype studies. MitoLSDB would also allow cross-comparison or meta-analysis of different mtDNA association studies and help understand the molecular correlates of mitochondrial disease phenotypes, which otherwise is a very daunting and challenging task given the complexity of mitochondrial genetics. It has been demonstrated earlier that publications contain a significant number of reporting errors that have been corrected or reported by curators and submitters of LSDBs [29]. We expect a similar trend for mtDNA variations and believe that community participation will further enhance data coverage, improved annotations in context of pathogenic status of variants and quality checks for spurious reports and correctness of the submitted data.

Acknowledgements

The authors also thank Dr. Arijit Mukhopadhyay and Dr. Mohammed Faruq for helpful comments on the manuscript. The research leading to results of MitoLSDB has received funding from the European Community's Seventh Framework Programme under grant agreement 200754 (the GEN2PHEN project).

Figures

Figure 1 - Frequency distribution of variants across various disease phenotypes

The abbreviations in the pie are : AD - Alzheimer's disease; AT - Ataxia; AP - Atypical Psychosis; OB - Obese; Non-OB - Non obese; PCT - primary cancerous tissue; BC - breast cancer; ThC- thyroid cancer; TZ - Teratozoospermic; semi-SC - semi-supercentenarian; T2D - Type 2 diabetes; T2DA - Type 2 Diabetes with Angiopathy; DD - Diabetes and deafness; NFFT1 - Neurofibromatosis type I; AZ -Asthanozoospermic; PD - Parkinson's disease; POLG1-T251I - genotype: POLG1 variant T251I; POLG1-G268A - genotype: POLG1 G268A; MELAS - Mitochondrial Encephalopathy, Lactic acidosis and Stroke like episodes; LHON - Leber's Hereditary Optic Neuropathy; MERRF - Myoclonic Epilepsy with Raged Red Fibers; CADASIL - Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and- Leukoencephalopathy; Stroke and vision loss - stroke-like episodes, lactic acidosis, exercise intolerance similar to that seen in MELAS; two episodes of transient central vision loss similar to that seen in LHON; CPEO - chronic progressive external ophthalmoplegia; OXPHOS deficiency; Glioma; Centenarian

Figure 2 - Normalized genewise distribution of synonymous and non-synonymous changes

The number of polymorphic sites in each protein coding gene is divided by the gene length to obtain normalized values of polymorphism across these genes in mtDNA. The graph shows the distribution of synonymous and non-synonymous changes in each gene.

Figure 3 - Frequency of transitions and transversions

This graph shows the frequency of base changes in the entire dataset. The frequency of transitions (4293) is higher than transversions (575).

Figure 4 - Screenshot of the MitoLSDB database

Screenshot shows the MitoLSDB home page for a gene and the variant listing. It also shows that mitolsdb is listed as a public LOVD installation.

References

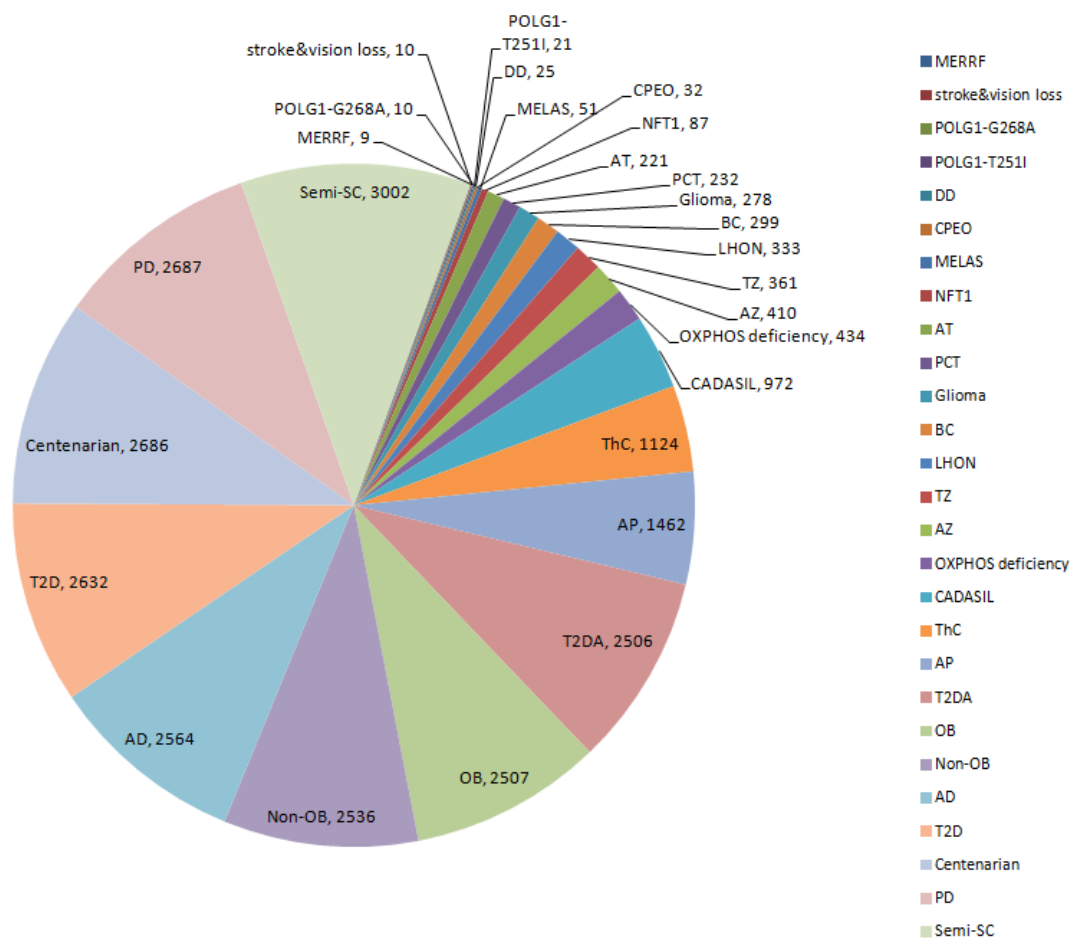
1. Taylor RW, Turnbull DM: **Mitochondrial DNA mutations in human disease.** *Nat Rev Genet* 2005, **6**(5):389-402.
2. Wallace DC: **Mitochondrial diseases in man and mouse.** *Science* 1999, **283**(5407):1482-1488.
3. Wallace DC: **A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine.** *Annu Rev Genet* 2005, **39**:359-407.
4. Gropman A, Chen TJ, Perng CL, Krasnewich D, Chernoff E, Tifft C, Wong LJ: **Variable clinical manifestation of homoplasmic G14459A mitochondrial DNA mutation.** *Am J Med Genet A* 2004, **124A**(4):377-382.
5. Elliott HR, Samuels DC, Eden JA, Relton CL, Chinnery PF: **Pathogenic mitochondrial DNA mutations are common in the general population.** *Am J Hum Genet* 2008, **83**(2):254-260.
6. Zeviani M, Bertagnolio B, Uziel G: **Neurological presentations of mitochondrial diseases.** *J Inherit Metab Dis* 1996, **19**(4):504-520.
7. Lam CW, Lau CH, Williams JC, Chan YW, Wong LJ: **Mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS) triggered by valproate therapy.** *Eur J Pediatr* 1997, **156**(7):562-564.
8. Yang J, Zhu Y, Chen L, Zhang H, Tong Y, Huang D, Zhang Z, Chen S, Han X, Ma X: **Novel A14841G mutation is associated with high penetrance of LHON/C4171A family.** *Biochem Biophys Res Commun* 2009, **386**(4):693-696.
9. Enriquez JA, Chomyn A, Attardi G: **MtDNA mutation in MERRF syndrome causes defective aminoacylation of tRNA(Lys) and premature translation termination.** *Nat Genet* 1995, **10**(1):47-55.
10. Schon EA, Bonilla E, DiMauro S: **Mitochondrial DNA mutations and pathogenesis.** *J Bioenerg Biomembr* 1997, **29**(2):131-149.
11. Morgan-Hughes JA, Sweeney MG, Cooper JM, Hammans SR, Brockington M, Schapira AH, Harding AE, Clark JB: **Mitochondrial DNA (mtDNA) diseases: correlation of genotype to phenotype.** *Biochim Biophys Acta* 1995, **1271**(1):135-140.
12. Wong LJ: **Diagnostic challenges of mitochondrial DNA disorders.** *Mitochondrion* 2007, **7**(1-2):45-52.
13. Kierdaszuk B, Jamrozik Z, Tonska K, Bartnik E, Kaliszewska M, Kaminska A, Kwiecinski H: **Mitochondrial cytopathies: clinical, morphological and genetic characteristics.** *Neurol Neurochir Pol* 2009, **43**(3):216-227.
14. Wallace DC: **Mitochondrial DNA sequence variation in human evolution and disease.** *Proc Natl Acad Sci U S A* 1994, **91**(19):8739-8746.
15. McKusick-Nathans Institute of Genetic Medicine (JHUB, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD): **Online Mendelian Inheritance in Man, OMIM (TM) [<http://www.ncbi.nlm.nih.gov/omim>]** 2009.
16. Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, Kreuziger J, Baldi P, Wallace DC: **An enhanced MITOMAP with a global mtDNA mutational phylogeny.** *Nucleic Acids Res* 2007, **35**(Database issue):D823-828.
17. Lee YS, Kim WY, Ji M, Kim JH, Bhak J: **MitoVariome: a variome database of human mitochondrial DNA.** *BMC genomics* 2009, **10** Suppl 3:S12.
18. Ingman M, Gyllensten U: **mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences.** *Nucleic Acids Res* 2006, **34**(Database issue):D749-751.
19. Chandrasekaran K, Giordano T, Brady DR, Stoll J, Martin LJ, Rapoport SI: **Impairment in mitochondrial cytochrome oxidase gene expression in Alzheimer disease.** *Brain Res Mol Brain Res* 1994, **24**(1-4):336-340.

20. Kazuno AA, Munakata K, Nagai T, Shimoazono S, Tanaka M, Yoneda M, Kato N, Miyawaki A, Kato T: **Identification of mitochondrial DNA polymorphisms that alter mitochondrial matrix pH and intracellular calcium dynamics.** *PLoS Genet* 2006, **2**(8):e128.
21. Cho YM, Park KS, Lee HK: **Genetic factors related to mitochondrial function and risk of diabetes mellitus.** *Diabetes Res Clin Pract* 2007, **77 Suppl 1**:S172-177.
22. Hanagasi HA, Ayribas D, Baysal K, Emre M: **Mitochondrial complex I, II/III, and IV activities in familial and sporadic Parkinson's disease.** *Int J Neurosci* 2005, **115**(4):479-493.
23. Saxena R, de Bakker PI, Singer K, Mootha V, Burt N, Hirschhorn JN, Gaudet D, Isomaa B, Daly MJ, Groop L *et al*: **Comprehensive association testing of common mitochondrial DNA variation in metabolic disease.** *Am J Hum Genet* 2006, **79**(1):54-61.
24. Pereira L, Freitas F, Fernandes V, Pereira JB, Costa MD, Costa S, Maximo V, Macaulay V, Rocha R, Samuels DC: **The diversity present in 5140 human mitochondrial genomes.** *Am J Hum Genet* 2009, **84**(5):628-640.
25. van Eijnsden RG, Gerards M, Eijssen LM, Hendrickx AT, Jongbloed RJ, Wokke JH, Hintzen RQ, Rubio-Gozalbo ME, De Coo IF, Briem E *et al*: **Chip-based mtDNA mutation screening enables fast and reliable genetic diagnosis of OXPHOS patients.** *Genet Med* 2006, **8**(10):620-627.
26. Fokkema IF, den Dunnen JT, Taschner PE: **LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach.** *Hum Mutat* 2005, **26**(2):63-68.
27. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT: **LOVD v.2.0: the next generation in gene variant databases.** *Hum Mutat* 2011, **32**(5):557-563.
28. Bhardwaj A, Mukerji M, Sharma S, Paul J, Gokhale CS, Srivastava AK, Tiwari S: **MtSNPScore: a combined evidence approach for assessing cumulative impact of mitochondrial variations in disease.** *BMC Bioinformatics* 2009, **10 Suppl 8**:S7.
29. Celli J, Dalgleish R, Vihinen M, Taschner PE, den Dunnen JT: **Curating gene variant databases (LSDBs): Toward a universal standard.** *Hum Mutat* 2011.
30. van Oven M, Kayser M: **Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation.** *Hum Mutat* 2009, **30**(2):E386-394.
31. Yao YG, Salas A, Logan I, Bandelt HJ: **mtDNA data mining in GenBank needs surveying.** *Am J Hum Genet* 2009, **85**(6):929-933; author reply 933.
32. Yao YG, Kong QP, Salas A, Bandelt HJ: **Pseudomitochondrial genome haunts disease studies.** *J Med Genet* 2008, **45**(12):769-772.
33. Knight RD, Landweber LF, Yarus M: **How mitochondria redefine the code.** *J Mol Evol* 2001, **53**(4-5):299-313.
34. Bandelt HJ, Achilli A, Kong QP, Salas A, Lutz-Bonengel S, Sun C, Zhang YP, Torroni A, Yao YG: **Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies.** *Biochem Biophys Res Commun* 2005, **333**(1):122-130.
35. Brandstatter A, Sanger T, Lutz-Bonengel S, Parson W, Beraud-Colomb E, Wen B, Kong QP, Bravi CM, Bandelt HJ: **Phantom mutation hotspots in human mitochondrial DNA.** *Electrophoresis* 2005, **26**(18):3414-3429.
36. Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V: **The fingerprint of phantom mutations in mitochondrial DNA data.** *Am J Hum Genet* 2002, **71**(5):1150-1160.
37. Bensasson D, Zhang D, Hartl DL, Hewitt GM: **Mitochondrial pseudogenes: evolution's misplaced witnesses.** *Trends Ecol Evol* 2001, **16**(6):314-321.
38. Goios A, Prieto L, Amorim A, Pereira L: **Specificity of mtDNA-directed PCR-influence of NUClear MTDNA insertion (NUMT) contamination in routine samples and techniques.** *Int J Legal Med* 2008, **122**(4):341-345.
39. Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D *et al*: **The dawn of human matrilineal diversity.** *Am J Hum Genet* 2008, **82**(5):1130-1140.

40. Kong QP, Salas A, Sun C, Fuku N, Tanaka M, Zhong L, Wang CY, Yao YG, Bandelt HJ: **Distilling artificial recombinants from large sets of complete mtDNA genomes.** *PLoS One* 2008, **3**(8):e3016.
41. Sun C, Kong QP, Palanichamy MG, Agrawal S, Bandelt HJ, Yao YG, Khan F, Zhu CL, Chaudhuri TK, Zhang YP: **The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes.** *Mol Biol Evol* 2006, **23**(3):683-690.
42. Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP: **Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia.** *Am J Hum Genet* 2004, **75**(6):966-978.
43. Bandelt HJ, Salas A, Taylor RW, Yao YG: **Exaggerated status of "novel" and "pathogenic" mtDNA sequence variants due to inadequate database searches.** *Hum Mutat* 2009, **30**(2):191-196.

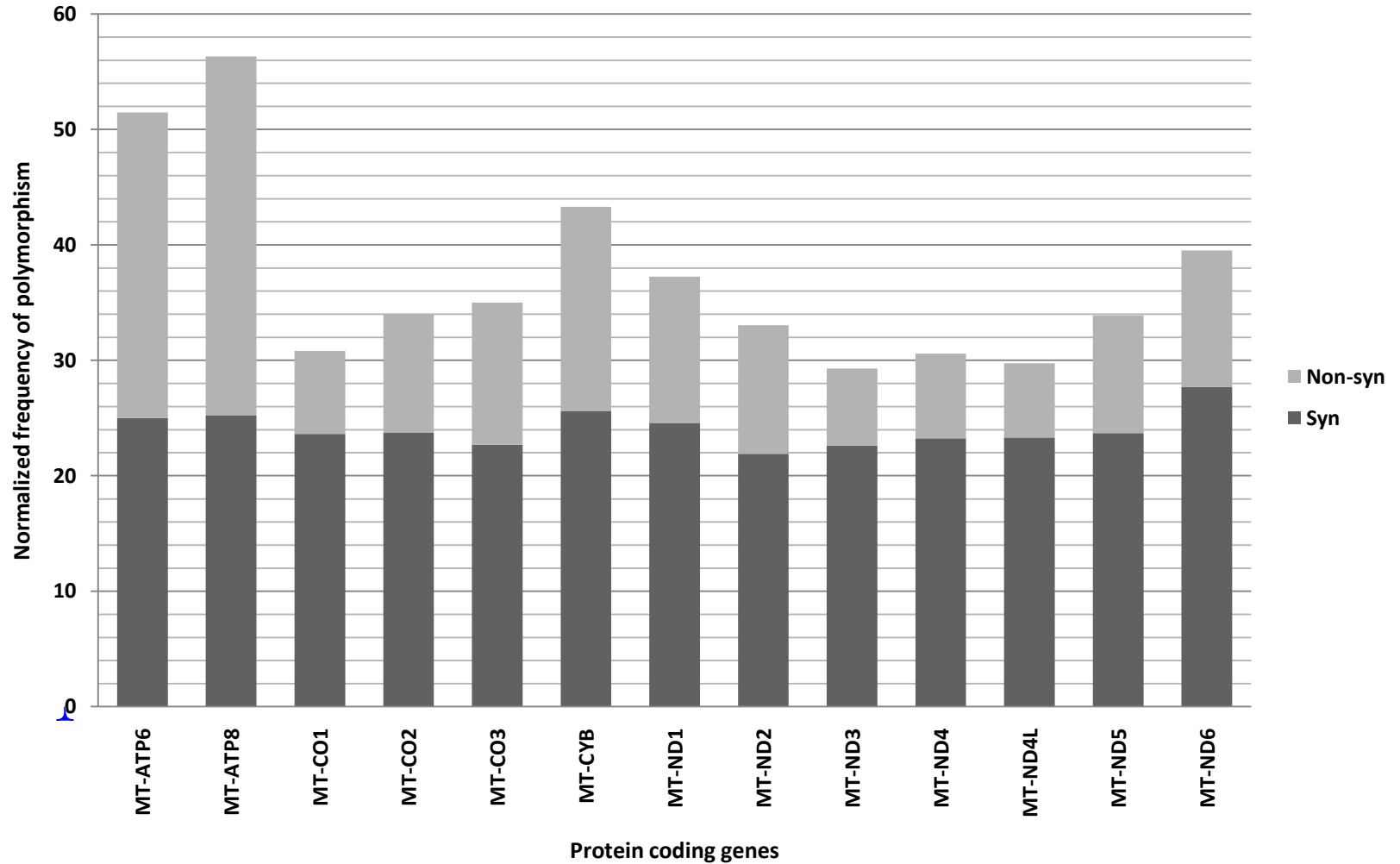
Figure 1

Frequency of polymorphism across various disease phenotypes

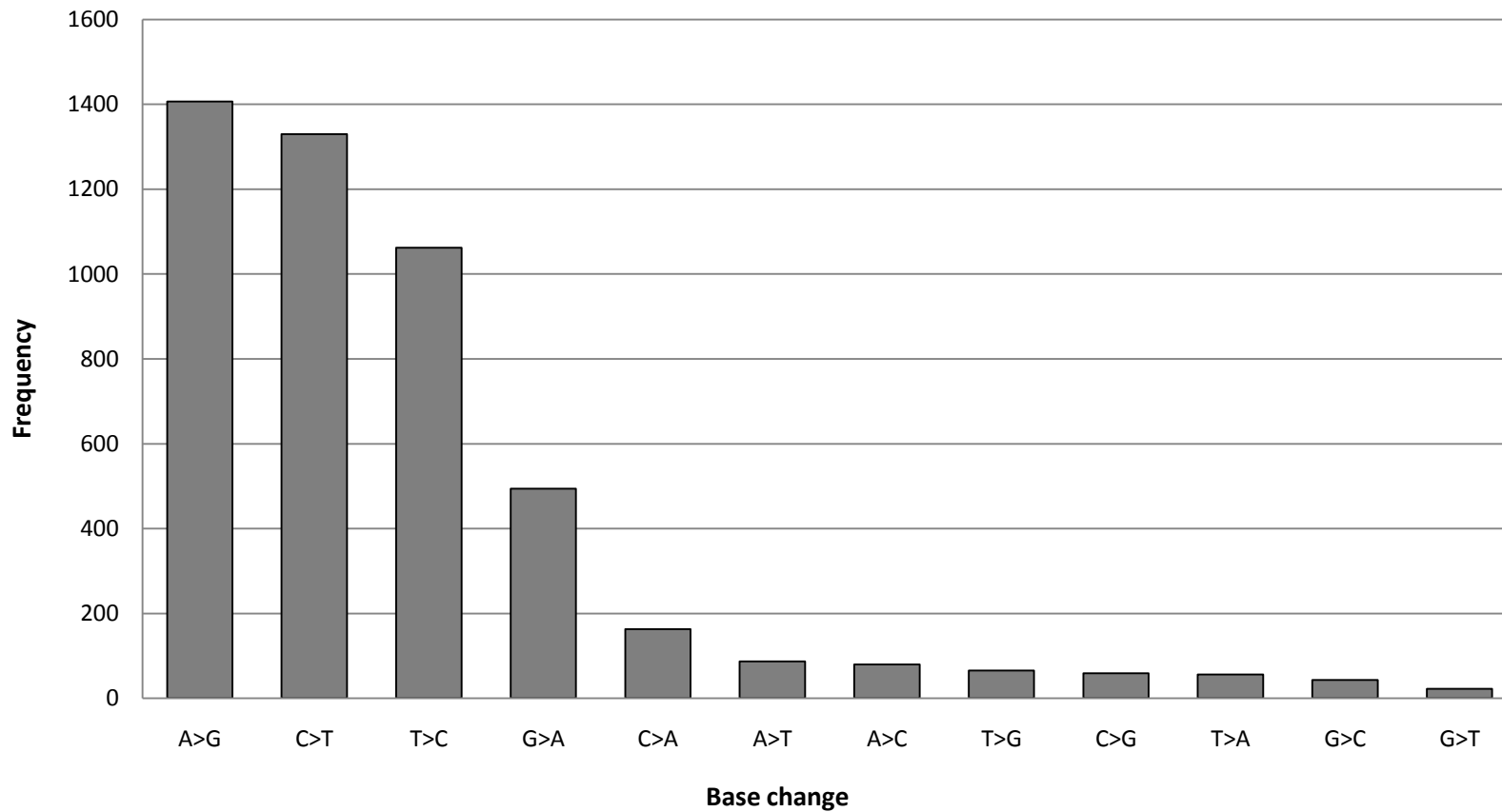


Phenotype	Frequency of variants	Abbreviation of phenotypes used in Pie chart
MERRF	9	MERRF
stroke& central vision loss genotype: POLG1 G268A	10	stroke&vision loss
stroke genotype: POLG1 variant T251I	10	POLG1-G268A
Diabetes with deafness	25	POLG1-T251I
CPEO	32	DD
MELAS	51	CPEO
neurofibromatosis type 1	87	MELAS
AT	221	NFT1
primary cancerous tissue	232	AT
Glioma	278	PCT
Breast cancer	299	Glioma
LHON	333	BC
Teratozoospermic	361	LHON
Asthenozoospermic	410	TZ
OXPHOS deficiency	434	AZ
CADASIL	972	OXPHOS deficiency
Thyroid cancer	1124	CADASIL
Atypical psychosis	1462	ThC
T2DA	2506	AP
OB	2507	T2DA
Non-obese	2536	OB
AD	2564	Non-OB
T2D	2632	AD
Centenarian	2686	T2D
PD	2687	Centenarian
Semi-supercentenarian	3002	PD
		Semi-SC

Normalized gene wise distribution of synonymous and non-synonymous changes



Frequency of transitions and transversions



LOVD *Leiden Open Variation Database*
mitosldb *mitochondrially encoded cytochrome b (MTCYB)*
 Curator: Anshu Bhardwaj

Home Variants Submitters Submit Documentation

MTCYB homepage Switch gene

LOVD Gene homepage

General information

Gene name	mitochondrially encoded cytochrome b
Gene symbol	MTCYB
Chromosome Location	M
Database location	192.168.6.66
Curator	Anshu Bhardwaj
PubMed references	View all (unique) PubMed references in the MTCYB database
Date of creation	September 16, 2011
Last update	October 20, 2011
Version	MTCYB111020
Add sequence variant	Submit a sequence variant
First time submitters	Register here
Reference sequence	Genomic reference sequence for describing sequence variants
GenBank reference	NC_012920
Total number of unique DNA variants reported	510
Total number of individuals with variant(s)	5163
Total number of variants reported	29833

Graphical displays and utilities

Summary tables	Summary of all sequence variants in the MTCYB database, sorted by type of variant (with graphical displays and statistics)
UCSC Genome Browser	Show variants in the UCSC Genome Browser (compact view)
Ensembl Genome Browser	Show variants in the Ensembl Genome Browser
NCBI Sequence Viewer	Show distribution histogram of variants in the NCBI Sequence Viewer

Sequence variant tables

Unique sequence variants	Listing of all unique sequence variants in the MTCYB database, without patient data
Complete sequence variant listing	Listing of all sequence variants in the MTCYB database
Variants with no known pathogenicity	Listing of all MTCYB variants reported to have no noticeable phenotypic effect (note: excluding variants of unknown effect)
Download table	Download the full sequence variant table of the MTCYB database in tab-delimited text format.

Search the database

By type of variant	View all sequence variants of a certain type
Simple search	Query the database by selecting the most important variables (exon number, type of variant, disease phenotype)
Advanced search	Query the database by selecting a combination of variables
Based on patient origin	View all variants based on your patient origin search terms
Search through hidden entries	Find the number of variant entries in the database (including hidden entries) matching your search terms.

Links to other resources

HGNC	7427
------	------

LOVD *Leiden Open Variation Database*
LOVD v.2.0 - Leiden Open Variation Database
 Online gene-centered collection and display of DNA variations

Home News FAQ Documentation Screenshots Download Contact

2.0 LOVD 3.0 Public list of LOVD installations LOVD 1.1.0 homepage LOVD 1.0.2 homepage

For all details about LOVD, see our [LOVD flyer!](#) (last updated February 24th, 2010)

List of public LOVD installations

If you are looking for a specific gene database, please also check the list of gene variant databases [at the HGVS site](#).

Please note that the latest available build is always installed on our Leiden server. We offer free use of this LOVD installation for those interested, e.g. to maintain/start a LSDB. [Contact us](#) for more information.

Show only LOVDs with gene symbol:

You can also [download this list](#) to see the complete list of genes in the databases.

In total: 388,294 variants (64,389 unique) in 102,443 patients in 24443 genes in 80 LOVD installations			
http://mitosldb.igib.res.in/	LOVD 2.0-31	37 genes	142495 variants
mitosldb	MTATP6,MTATP8,MTCO1,MTCO2,MTCO3,MTCYB...		5170 unique
http://www.dmd.nl/nmdb2/	LOVD 2.0-33	92 genes	66263 variants
Leiden Muscular Dystrophy pages	ACTA1,ACTC1,AGRN,ANO5,BAG3,BIN1,CAPN3...		9143 unique
http://chromium.liacs.nl/LOVD2/colon_cancer/	LOVD 2.0-33	10 genes	27135 variants
Colon cancer gene variant databases	APC,CDH1,EPCAM,MLH1,MLH3,MSH2,MSH6,MU...		4410 unique
http://grenada.lumc.nl/LOVD2/MR/	LOVD 2.0-33	542 genes	27093 variants
X-chromosome gene database	ABCB7,ABCD1,ACE2,ACOT9,ACSL4,ACTRT1,A...		3270 unique

LOVD - Variant listings

[Unhide all columns](#) | [Hide Specific Columns](#) | [Hide all columns](#)

[About this overview](#) [[Show](#)]

29833 public entries
 entries per page

Path.	DB-ID	DNA change	Protein	Codon Change	Template	Technique	Patient ID	Disease	Geographic origin	Population	Detection/Tissue	Referenc
??	MTCYB_00002	m.14766C>T + 38 others	p.T71	ACT>ATT	DNA	SEQ	EU828637	Tuvan LHON	Russia	North-East Siberia	Blood	15638829
??	MTCYB_00018	m.14783T>C + 38 others	p.L13L	TTA>CTA	DNA	SEQ	EU828637	Tuvan LHON	Russia	North-East Siberia	Blood	15638829
??	MTCYB_00010	m.15043G>A + 38 others	p.G99G	GGG>GGA	DNA	SEQ	EU828637	Tuvan LHON	Russia	North-East Siberia	Blood	15638829
??	MTCYB_00203	m.15080A>G + 38 others	p.T112A	ACC>GCC	DNA	SEQ	EU828637	Tuvan LHON	Russia	North-East Siberia	Blood	15638829

Supplementary Table1: Distribution of phenotypes across individuals from different			
Phenotype	Ethnicity/Population	Number of individuals	MeSH Heading
Alzheimer disease	Chiba Japan	96	Alzheimer disease
Asthenozoospermic	Portugal	20	Sperm Motility
Atypical psychosis	Japan	57	Psychotic Disorders
Breast cancer	Italy	19	Breast Neoplasm
	Romania	1	
CADASIL	Finland	77	CADASIL
Centenarian	Gifu Japan	11	
	Gifu Japan	11	Aged, 80 and over
	Tokyo Japan	85	
chronic progressive external ophthalmoplegia (CPEO)	Bonn, Germany	1	Ophthalmoplegia, Chronic Progressive External
diabetes and deafness	The Netherlands	1	Diabetes and deafness, maternally inherited
Diabetes Type II	Aichi Japan	96	Diabetes Mellitus, Type 2
	Ashkenazi Jew; Belarus	1	
Diabetic with angiopathy	Tokyo Japan	96	Diabetic Angiopathies
genotype: POLG1 G268A	Russia: Belgorod	1	DNA-Directed DNA Polymerase
genotype: POLG1 variant T251I	Russia: Belgorod	1	DNA-Directed DNA Polymerase
Glioma	Italy	16	Glioma
LHON	Central China	2	Optic Atrophy, Hereditary, Leber
	China	1	
	Italy	7	
	Brazil	1	
	Tuvan, Russia, North-East Siberia	1	
	slavic, Siberia, Russia	2	
MELAS	Turkey	1	MELAS Syndrome
	Vietnam	1	
	Germany	1	

MERRF	Germany	1	MERRF Syndrome
neurofibromatosis type 1		4	Neurofibromatosis 1
Non-obese young male	Aichi Japan	96	No suitable term
Obese young male	Aichi Japan	96	No suitable term
OXPPOS deficiency	The Netherlands: Nijmegen	28	Mitochondrial Diseases
Parkinson disease	Tokyo; Japan	96	Parkinson Disease
primary cancerous tissue	China	10	No suitable term
Semi-supercentenarian	Japan	112	Aged, 80 and over
stroke-like episodes, lactic acidosis, exercise intolerance similar to that seen in MELAS; two episodes of transient central vision loss similar to that seen in LHON	Ashkenazi, Jew, Poland	1	
Teratozoospermic	Portugal	23	No suitable term
Thyroid cancer	Italy	64	Thyroid Neoplasms
	Albania	1	
	Chile	1	