



Published in final edited form as:

Curr Protoc Bioinformatics. ; 1(123): 1.23.1–1.23.26. doi:10.1002/0471250953.bi0123s44.

mtDNA Variation and Analysis Using MITOMAP and MITOMASTER

Marie T. Lott,

Center for Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; CMEM admin phone: 1-267-425-3078; CMEM admin fax: 1-267-426-0978

Jeremy N. Leipzig,

Center for Biomedical Informatics, The Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; phone: 1-267-426-1375; fax: 1-215-590-5245

Olga Derbeneva,

Center for Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; phone: 1-267-425-3064; fax: 1-267-426-0978; work cell: 1-215-866-8121

H. Michael Xie,

Center for Biomedical Informatics, The Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; phone: 1-267-426-0675; fax: 1-215-590-5245

Dimitra Chalkia,

Center for Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA

Mahdi Sarmady,

Center for Biomedical Informatics, The Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; phone: 1-267-426-1373; fax: 1-215-590-5245

Vincent Procaccio, and

National Center for Neurodegenerative and Mitochondrial Diseases CHU Angers, Biochemistry and Genetics Department, Angers, France

Douglas C. Wallace

Center for Mitochondrial and Epigenomic Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; Department of Pathology and Laboratory Medicine; University of Pennsylvania, Philadelphia, PA; phone: 1-267-425-3078; fax: 1-267-426-0978

Marie T. Lott: LottM1@email.chop.edu; Jeremy N. Leipzig: LeipzigJ@email.chop.edu; Olga Derbeneva: DerbenevaO@email.chop.edu; H. Michael Xie: XieM1@email.chop.edu; Mahdi Sarmady: SarmadyM@email.chop.edu; Vincent Procaccio: ViProcaccio@chu-angers.fr; Douglas C. Wallace: WallaceD1@email.chop.edu

Abstract

The MITOMAP database of human mitochondrial DNA (mtDNA) information has been an important compilation of mtDNA variation for researchers, clinicians and genetic counselors for the past twenty-five years. The MITOMAP protocol shows how users may look up human mitochondrial gene loci, search for public mitochondrial sequences, and browse or search for

reported general population nucleotide variants as well as those reported in clinical disease. Within MITOMAP is the powerful sequence analysis tool for human mitochondrial DNA, MITOMASTER. The MITOMASTER protocol gives step-by-step instructions showing how to submit sequences to identify nucleotide variants relative to the rCRS, to determine the haplogroup, and to view species conservation. User-supplied sequences, GenBank identifiers and single nucleotide variants may be analyzed.

Keywords

biological database; information retrieval; human mitochondrial DNA; haplogroups; species conservation; GenBank sequences; single nucleotide variants

Introduction

The MITOMAP database of human mitochondrial DNA (mtDNA) information has been an important resource for information about the human mitochondrial DNA (mtDNA) for researchers, clinicians and genetic counselors for the past twenty-five years. Essential information about the mitochondrial reference sequence is provided along with an extensive compilation of mtDNA variants. The MITOMAP curators search research literature for published reports of mitochondrial DNA variants and index those variants in the database. Those variants which are reported as having possible association with disease are noted. A new addition to MITOMAP is the inclusion of data from full-length human mtDNA sequences in GenBank.

The MITOMAP protocol section shows how users may look up human mitochondrial gene loci, search for public mitochondrial sequences, and browse or search for reported general population nucleotide variants as well as those reported in clinical disease. Within MITOMAP is the powerful sequences analysis tool for human mitochondrial DNA, MITOMASTER.

The MITOMASTER protocol section gives step-by-step instructions showing how to submit sequences to identify nucleotide variants relative to the rCRS, to determine the haplogroup, and to view species conservation. User-supplied sequences, GenBank sequences and single nucleotide variants may be analyzed.

Basic Protocol 1: Exploring mtDNA Variants with MITOMAP

The MITOMAP database can be accessed at <http://www.mitomap.org> (Figure 1).

MITOMAP consists of three main sections: 1) Background information about the human mitochondrial DNA; 2) An annotated listing of mtDNA variants, both general population and patient; and 3) The MITOMASTER analysis tool (Basic Protocol 2).

Necessary Resources

Hardware—Internet connection

Software—An up-to-date web browser such as Firefox, Chrome, Safari, or Internet Explorer (version 9 or higher)

MITOMAP's Background Information about Human Mitochondrial DNA

1. Access the MITOMAP database at <http://www.mitomap.org>. Several pages of important background information are available, shown in Figure 2.
2. Click on Link 1, [The Annotated Human Mitochondrial DNA Sequence](#) (Figure 3) to find the information about the revised Cambridge reference sequence (Andrews et al., 1999) and the sequence itself.

This is critical information because for many years researchers mistakenly used either the original but outdated version of the reference sequence (Anderson et al., 1981) or an African Yoruban sequence which was listed as the primary reference sequence in GenBank for several years (AF347015 / formerly RefSeq NC_001807.4). Use of the Yoruban sequence on DNA sequencing chips has often resulted in confusion and, occasionally, misinterpretation of data.

3. Click on Link 2, [The rCRS & other mtDNAs](#) (Figure 4) to go to a companion page where you will find the version history of the rCRS, links to other representative mtDNA sequences from different continental populations as well as search tools to retrieve full length mtDNA sequences from GenBank.
4. Click on the “search for complete human mtDNAs in GenBank” link (Figure 4, red arrow) to retrieve sequences with a minimum length of 15400 bp and a maximum of 16600 bp.
5. Click on the link “complete eukaryote (non-human) genomes in GenBank: (Figure 4, blue arrow) to retrieve mtDNA sequences of other species. In February 2013, these searches returned 17851 full-length human mitochondrial DNA sequences and 9608 complete non-human mitochondrial DNA sequences.
6. Click on Link 3, [Amino Acid Translation Table](#) (Figure 5) to take you to the human mitochondrial genetic code, with important notes as to the differences between it and the nuclear genetic code.
7. Click on Link 4, [Mitochondrial References](#) to browse a library of the ~5000 publications indexed by the Mitomap curators, with links to PubMed.
8. Select Link 5, [Haplogroup Markers and Frequencies](#) to reach background haplotyping information (Figure 6-7). Helpful maps of world migrations and haplogroup relationships are located nearby (Figures 8-9).
9. Click on [Mitochondrial DNA Function Locations](#) (Link 6) to pull up delineated positions of gene loci (Figure 10). Mitomap's classic map of loci with selected pathological DNA variants is available in the adjacent section of illustrations (Figure 11).

Human Mitochondrial DNA Variants

- 10 An extensive database of published mitochondrial DNA variants is available. These can be searched (Figure 12a, <http://www.mitomap.org/bin/view/Main/SearchAllele>) or browsed (Figure 13).
- 11 To search for variants,
 - a. Locate the bulleted “to search for point mutations, click **here**” link in the “Mitomap Quick Reference” section near the top of the MITOMAP home page (Figure 12a). Click “here” to open the Allele Search box (Figure 12b).
 - b. Enter
 - i. a single nucleotide position in the “Start” box or
 - ii. a range of up to 100 nucleotides by also entering a position number in the “End” box.
- 12 Click the “Search” button. Results will return as a listing of reported variants and their references as seen in Figure 12c.

Population Variants

- 13 There are two links for Control Region Variants and Coding/RNA Variants. Information is listed for nucleotide position, rCRS nucleotide, variant nucleotide, and, for coding region variants, amino acid change and codon position.
 - a. Click on the link *Control Region Variants (16024-576)* to find variants located between tRNA Proline and tRNA Phenylalanine.
 - b. Click on the link *Coding & RNA Variants (577-16023, MTF-MTTP)* to find variants located in the region including the beginning of tRNA Phenylalanine through the end of tRNA Proline.

New for 2013 is the frequency of each variant in a large set of over 18,000 human mitochondrial DNA sequences from GenBank. These sequences have a minimum length of 15.4 kb and are extracted from GenBank on a quarterly basis.

- 14 Click on the GB set frequency for a given variant to retrieve a listing of sequences which contain the variant of interest (Figure 16) and the relevant Pub Med reference. In addition, MITOMAP's companion analysis tool MITOMASTER (detailed in Protocol 2) predicts haplogroup for each sequence, calculates the number of different haplogroups seen carrying this variant, and displays the total number of each haplogroup found in the GenBank set of sequence. Clicking on each ID and haplogroup yields more information (Figures 16-20). Haplotyping is based on Phylotree (van Oven and Kayser, 2009) and is generated by MITOMASTER using the Haplogrep engine (Kloss-Brandstatter et al., 2011).

Patient Variants

- 15 To locate information about reported mtDNA variation in patients click on one of the links listed in the section “**mtDNA Mutations with Reports of Disease-Associations**”, shown in Figure 13.

Variants reported as having possible association with disease are grouped into two categories: those found in the Coding or Control Region (Figure 21) and those found in the mitochondrial rRNA or tRNA loci (Figure 22). In addition to the same information presented in the population variant tables, reports of heteroplasmy are indicated as well as a general category of pathogenicity. A pathogenic status of “Reported” indicates that a publication has considered the mutation as possibly pathologic. A status of “Cfrm” (confirmed) indicates that several independent laboratories have published strong evidence of the pathogenicity of a specific mutation. These mutations are generally accepted by the mitochondrial research community as being pathogenic.

Basic Protocol 2: Analyzing mtDNA Variants with MITOMASTER

Mitochondrial sequence analysis typically begins with specialized mitochondrial SNP genotyping microarrays, using Sanger-based capillary sequencing, or various forms high-throughput sequencing. MITOMASTER is designed to accommodate output from each of three strategies. Small batches of sequence can be “copy-and-pasted” into a text field or uploaded from FASTA-formatted files. Query by GenBank ID is available for individual or multiple record numbers. Single nucleotide variants, identified by microarray, can be submitted using the SNV Query tab.

Necessary Resources

Hardware—Internet connection

Software—An up-to-date web browser - MITOMASTER is optimized for Firefox, Chrome, or Safari. If Internet Explorer is used, it must be version 9 or higher.

Files—Your data of interest, either as a FASTA file (such as `sequence.fasta`, downloaded in this protocol) or as a text file from which you can copy and paste

Sequence Query Walk-Through

1. Navigate to <http://www.ncbi.nlm.nih.gov/nuccore/EU915478>. This sequence, a full length mitochondrial sequence, can demonstrate some of the features (Figure 23):
2. Click on the “Send” popup dialog link in the upper right of the main section of the sequence record. Select “File” as a destination and FASTA as a format (Figure 24).

The file will be saved to your local drive as “`sequence.fasta`” in your default download folder. Make a note of this folder location. For details of the FASTA format, see http://www.bioinformatics.nl/tools/crab_fasta.html.

3. Navigate to <http://mitomaster.mitomap.org/>. This is also accessible as a frame within the MITOMAP website at <http://mitomap.org/MITOMASTER>. You may also use the MITOMASTER link on the left menu of the MITOMAP home page.

Please note that currently the only browsers supported are Chrome, Mozilla, Firefox, and Internet Explorer 9 or higher.

By default the “**Sequence**” tab in the top menu is selected (Figure 25).

4. Skip Step 1 for now. We will discuss this feature later in the walk-through. By default, 45 species are selected for comparison.
5. Click “**Select File**” in Step 2, Option 1 and navigate to the ‘sequence.fasta’ file which you downloaded earlier.
6. Click “**Submit**”. The **Alignment Summary** page (Figure 26) is then shown after the sequence is processed:

There are two portions to the **Alignment Summary** screen:

- **rCRS track view**
 - The rCRS track view shows the coverage of the query sequence with respect to the rCRS reference sequence. The example sequence EU915478, indicated by a bar at the top of the screen, displays complete coverage. This is expected as it is a full-length sequence. Tracks below the query sequence correspond to the locations of protein coding genes, ribosomal RNAs, and transfer RNAs.
 - **Sequence alignment**
 - For each query sequence analyzed you will find (1) the predicted haplogroup as calculated by HaploGrep, (2) the total number of variants relative to the rCRS, and (3) a condensed list of the variants observed. In the example given (Figure 28, circled in blue), the predicted haplogroup is J1b, the total number of variants detected is 41, with a summary listed in the right-most column.
7. Next, click on the sequence name to bring up the Alignment Details page (Figure 29). In the example above, you will click on “gi|194441041|gb|EU915478.1” (Figure 28, blue arrow). The sequence name is taken from the first line of the ‘sequence.fasta’ file. After clicking on the sequence name, the Alignment Details page will open (Figure 29).

The **Alignment Detail** page (Figure 29) shows the rCRS reference position, the query sequence position, the reference base and query base with respect to the L-strand of the rCRS (NC_012920). The mutation type (substitution type or indel type), the locus or loci intersected, and the predicted transcript effect, if any.

8. The amount of information displayed can be adjusted with the drop-down menu at the top, ‘**Display 10 records**’. The default value is 10 records per page, but more records (25, 50, 100 or “All”) can be selected.

- 9 Each column can be sorted by clicking the column headers. To explore the data columns, first click on the header “**Patient Report**” twice to sort descending (Figure 30).

In Figure 30, the **Patient Report** column shows variants in this sequence which have been published in the literature as possibly having some disease association in patients with Leber Hereditary Optic Neuropathy (LHON), cyclic vomiting and AD/PD. Do not assume that a listing of a published patient report is a confirmation of a variant's disease association. In this screenshot, 14484C is a well known LHON mutation; however, 3010A is a relatively common variant, found in 17% of all sequences in the mined GenBank set (currently numbering over 18,000) and in 98% of haplogroup J1b sequences. Please read more in the **Commentary** section about interpreting disease reports.

- 10 For coding mutations, a **Conservation** value is calculated based on the user's initial selection of species. By default, 45 species are used.
 - a. To explore this, select the first Conservation value (51.11%, circled in red dashes in the above figure).
 - b. Open this blue ‘**51.11%**’ link in a new tab or window by right-clicking if using a PC or control-clicking if using a Mac. This will open the **Conservation Grid** (Figure 31A). Keep this window or tab open.
- 11 Also open the **Species** link found in the top menu bar (Figure, 30, circled in red dots) in a new tab. Keep this window or tab open as well.
- 12 First, let's examine the **Conservation Grid** (Figure 31A). Amino acid conservation is shown for genic mutations. Note the query sequence has an ‘A’ residue while rCRS has a ‘T’ residue. Finally, 23 of the 45 species (51.11%) have the ‘T’ residue at that position – note this is the fraction which matches the rCRS value, not the query residue.
- 13 Now examine the “**Species Selection**” tab or window (Figure 31B). Species are subcategorized into checkbox groups. Users select which species are used in the conservation index calculations.
- 14 As an example, click on both “**Primates**” and “**Artiodactyla**” to open the drop-down view of the contents of these two categories. You may move your mouse to hover over any genus/species name to view the common name. Click each category name again to close the drop downs.
- 15 Another popular selection option is to use the “Select All”, “Clear All” or “Primates Only” at the top of the page.
- 16 Click on “**Primates Only**” – the change in selected species is instant. Subsequent queries will use this subset of 12 primate species only.

GenBank queries

- 17** To aid basic research, Mitomaster offers one-step analysis of any human mitochondrial sequences stored in GenBank — 104,705 sequences at the time of this publication (17,869 sequences 15.40kb-16.66kb in length plus another 86,836 sequences less than 15.4 kb). One or more GenBank sequence identifiers can be used as query parameters; Mitomaster fetches these sequences using the NCBI web service and returns reports. Acceptable inputs are single GenBank accession number or GI sequence identifier (e.g., EF060316 or 93116889), a comma-separated list of identifiers with or without intervening spaces (e.g., EF060316, 302376313, DQ112752 or EF060316,302376313,DQ112752), or a range of sequence numbers (AF346963-AF346968). Data reports are produced as described earlier in walk-through of the FASTA sequence analysis. (Figure 32)

SNV submission

- 18** Single Nucleotide Variants can be analyzed alone, without the need of a complete sequence. Data reports are produced as described earlier in walk-through of the FASTA sequence analysis.

Single Nucleotide Variants can be formatted the following ways and may either be pasted into the “copy/paste” window or uploaded as a text file (Figure 33).

- a. A four-column tab-delimited format consisting of sample name, rCRS mitochondrial position, reference base, and query base (“sample pos ref var”).

With header row				Without header row			
Sample	pos	ref	var	Sample1	73	A	G
Sample1	73	A	G	Sample1	146	T	C
Sample1	146	T	C	Sample2	263	A	:
Sample2	263	A	:	Sample2	709	:	G
Sample2	709	:	G				

- b. A more compact non-delimited reference-position-query format is also permitted, with or without sample names.

With sample names	Without sample names		Without rCRS value
Sample1	A73G	A73G	73G
Sample2	T146C	T146C	146C

- c. Indels can be represented using the adjacent-base notation (C573CC) or with the use of decimal positions (573.1C), explicitly (573insC), or using dash or colon symbols (:573C).

With sample names	Without sample names	Explicit/decimal format	-/: format
Sample123	A263d	A263d	263delA
			A263-

With sample names	Without sample names	Explicit/decimal format	-/: format
Sample456 C573CCC	573CC	573.1insC	:573C

Guidelines for Understanding Results

With the recent explosion of sequence data, MITOMAP has found it necessary to augment the hand-curated portion of the database with mined sequence data from GenBank. The new GenBank frequency data is derived from sequences with size equal to or larger than the complete coding region. These sequences have been pre-loaded into MITOMASTER and represent almost all haplogroups known to date. As of August 2013 this data set contained over 18000 sequences. The size of the sequence set is expected to increase with quarterly scans of GenBank and possibly other public sequence repositories. Please keep in mind that pre-loaded sequences from GenBank have not been individually reviewed by MITOMAP.

GenBank sequences may not be of equal quality (Yao et al., 2009). Published results may have sequencing errors, partial data, and analysis mistakes that, even if corrected, still might not be downloadable as a corrected sequence directly from GenBank and might only be found as published erratum to the corresponding publications.

When considering the frequency of any particular variant in GenBank, understand that the numbers are not the true worldwide frequency but are only a reflection of the actual sequences currently collected. Many populations in the world are under-represented in sequence repositories due to remote location and economics. Some sequences in public databases are occasionally duplicated with different record numbers and therefore do not represent unique individuals. Also, multiple sequences might be from closely related family members with identical mitochondrial DNA.

More and more sequences from patients with known or suspected mitochondrial disease are now being banked in public databases. It is important to keep in mind that any human mitochondrial sequence has the potential to contain variants relevant to past, current or future disease. This is especially true in the case of diseases which develop later in life or after particular environment stressors and thus might not be evident in an individual at the time of sequencing.

COMMENTARY

Background Information

MITOMAP had its beginnings as a 1994 report to the Human Genome Committee, appearing in print form in 1995 (Wallace et al., 1995). This was the first attempt to record and document all published human mitochondrial DNA variation in the general population as well as summarizing continent-specific and disease-specific variants. It also served as a unified reference for human mitochondrial gene loci. MITOMAP was first launched as an online mtDNA database in 1996 and has been an essential tool for genetic researchers, counselors, and clinicians ever since. With its well-documented and hand-curated dataset,

MITOMAP allows users to browse or search for reported mtDNA variants from both the general and clinical disease populations.

Critical Parameters

Alignment and Positional Caveats—Mitomaster uses a standard pair-wise BLAST local alignment to determine substitutions and mismatches. A local alignment will only include subsequences to achieve optimal alignment given a set of parameters, and so areas of extreme dissimilarity will be unaligned and no results will be given. This can result in truncated alignments or alignments composed of discontinuous high-scoring pairs. The genome viewer on the summary page shows the portion of the rCRS reference covered by the query sequence.

Calculation of a mutation's effect is done on an individual basis, isolated from any potential interactions of other mutation events within the query sequence. For instance, an insertion followed closely by a deletion within a protein coding gene would result in both being reported as independent frameshift mutations.

Also, be aware that insertions and deletions in regions of nucleotide polytracts (for example, “ACCCCT”) are often reported using different conventions. Some publications list the insertion/deletion point at the beginning of nucleotide run, some at the end. Further complications arise in repetitive polytract regions (for example “CCCCCTCTACCCCTCTAG”). This will often result in inaccurate sequence frequency numbers when MITOMASTER scans published indel reports for matches with GenBank sequences.

Large Data Sets—Large data sets may take a long time to process in MITOMASTER depending upon the number of FASTA files submitted and the server load at the time of submittal. To allow high-throughput sequence analyses using programmatic scripts, a web service utility is provided. This service is for use by programmers and database support personnel. Data may be submitted in large file batches (*e.g.*, 100 or more sequences) to the MITOMASTER web service via script clients to programmatically submit queries using a simple POST mechanism. Sample client scripts in Perl and Python are provided on the MITOMASTER site. Input is by FASTA file, and output is in a tab-delimited text format with similar content as delivered on the interactive web application.

Pathogenicity Cautions—Mitochondrial DNA is highly variable in and between populations. New variants will continue to be discovered as more and more sequencing studies are performed. It is possible that some sporadic mutations as well as known haplogroup-defining or polymorphic variants might be involved in a disease, but to make any conclusions concerning pathogenicity, more evidence and data analyses are required. Caution is advised when using the listing of mtDNA variants found in patient groups. A status of “Confirmed” (“Cfrm”) in MITOMAP is not an assignment of pathogenicity but is a general consensus of what is reported in published literature. Researchers and clinicians are advised that additional data, searches of other databases, and/or analyses are usually required to confirm the pathological significance of some of these mutations. Such due

diligence will also reduce the number of false reports of “novel” and “pathogenic” mtDNA variants in the literature.

Haplogroups—Haplogroup definition, while in some cases possible on a truncated sequencing data set, is best done using a complete mtDNA sequence. Caution is advised against popular “allelic” analysis approach where a researcher is only looking at a single nucleotide variant in relation to a disease phenotype. This approach is not directly applicable to mtDNA analysis. There are sets of different variants that are linked together and define mtDNA highly hierarchical phylogeny and haplogroup assignments. Some variants are polymorphic and could be found in many different mitochondrial lineages on different haplogroup backgrounds, some could be haplogroup-defining and rarely found anywhere else, while there are variants that are haplogroup-defining and polymorphic at the same time. The latter variants could be pathogenic on a different (non-defining) haplogroup background or in a different environment, but be non-pathogenic in their defining haplogroup. Thus they might occasionally be found in MITOMAP as being listed in both the polymorphism table and the disease table. Ethnic, geographical, and historic factors can also come into play when attempting to correlate a haplogroup with a medical condition.

Acknowledgments

This work was supported by NIH Grants NS21325, NS070298, AG24373, and DK73691 plus Simons Foundation Grant 205844 awarded to DCW.

Literature Cited

- Anderson S, et al. Sequence and organization of the human mitochondrial genome. *Nature*. 1981; 290:457–465. [PubMed: 7219534]
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999; 23:147. [PubMed: 10508508]
- Kloss-Brandstatter A, Pacher D, Schonherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat*. 2011; 32:25–32. [PubMed: 20960467]
- van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat (Online)*. 2009; 30:E386–E394.
- Wallace, DC.; Lott, MT.; Brown, MD.; Huoponen, K.; Torroni, A. Report of the committee on human mitochondrial DNA. In: Cuticchia, AJ., editor. In *Human Gene Mapping 1994, a Compendium*. The Johns Hopkins University Press; Baltimore, MD: 1995. p. 910-954.
- Yao YG, Salas A, Logan I, Bandelt HJ. mtDNA data mining in GenBank needs surveying. *Am J Hum Genet*. 2009; 85:929–33. author reply 933. [PubMed: 20004768]

Key References

- Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res*. 2002; 12:1611–1618. Supporting publication for Bioperl. [PubMed: 12368254]
- Steinbiss S, Gremme G, Scharfer C, Mader M, Kurtz S. AnnotationSketch: a genome annotation drawing library. *Bioinformatics*. 2009; 25:533–534. Supporting publication for AnnotationSketch. [PubMed: 19106120]

- Yao YG, Salas A, Logan I, Bandelt HJ. mtDNA data mining in GenBank needs surveying. *Am J Hum Genet.* 2009; 85:929–33. author reply 933. Important caveats about use of data mined from public sequences. [PubMed: 20004768]
- Zaragoza MV, Brandon MC, Diegoli M, Arbustini E, Wallace DC. Mitochondrial cardiomyopathies: how to identify candidate pathogenic mutations by mitochondrial DNA sequencing, MITOMASTER and phylogeny. *Eur J Hum Genet.* 2011; 19:200–207. Using MITOMASTER to investigate pathogenicity. [PubMed: 20978534]

Internet Resources

<http://www.mitomap.org>

MITOMAP

<http://mitomaster.mitomap.org>

MITOMASTER

<http://www.phylotree.org>

Phylotree

<http://haplogrep.uibk.ac.at/>

Haplogrep

<http://www.ncbi.nlm.nih.gov>

NCBI

<http://blast.ncbi.nlm.nih.gov/>

BLAST

http://www.ncbi.nlm.nih.gov/nuccore/NC_012920.1

The rCRS: Homo sapiens mitochondrion, complete genome

http://www.bioinformatics.nl/tools/crab_fasta.html

FASTA format

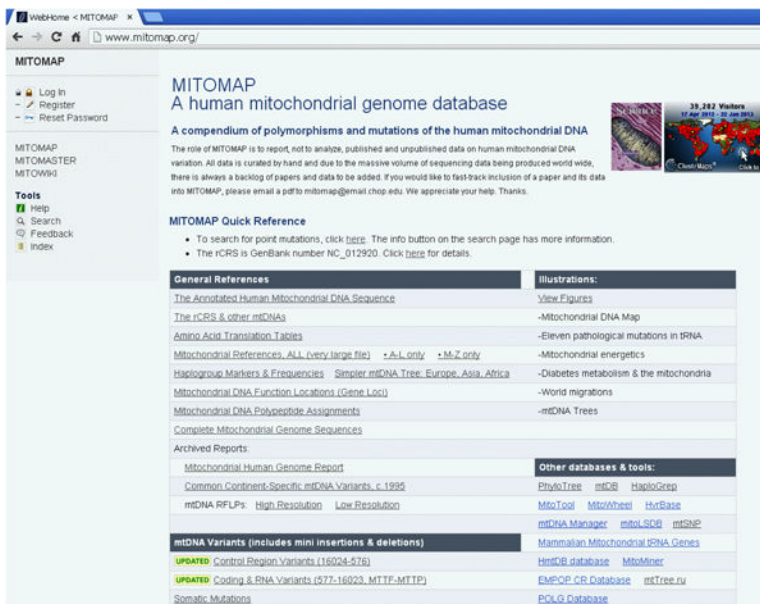


Figure 1. The home page of www.mitomap.org

General References

- 1: [The Annotated Human Mitochondrial DNA Sequence](#)
- 2: [The rCRS & other mtDNAs](#)
- 3: [Amino Acid Translation Tables](#)
- 4: [Mitochondrial References, ALL \(very large file\)](#) • [A-L only](#) • [M-Z only](#)
- 5: [Haplogroup Markers & Frequencies](#) [Simpler mtDNA Tree: Europe, Asia, Africa](#)
- 6: [Mitochondrial DNA Function Locations \(Gene Loci\)](#)

Figure 2. Essential mtDNA Background Information

Revised Cambridge Reference Sequence (rCRS) of the Human Mitochondrial DNA

The rCRS sequence is a fully corrected version of the original Cambridge Reference Sequence.

The rCRS is GenBank sequence [NC_012920](#) gi:251831106

Get the more information about the rCRS and download the rCRS plus other complete mtDNA reference sequences at GenBank [here](#).

IMPORTANT: Do not use RefSeq NC_001807 as "the rCRS" as it is an African (Yoruban) sequence with over 40 variant nucleotides from the rCRS. On July 8, 2009 the sequence was removed from GenBank as a reference sequence but may be found, if needed, as [AF347015](#), one of 53 African sequences deposited in GenBank by Ingman et al in 2001. Unfortunately, mistaken use of this Yoruban sequence as the rCRS is still seen occasionally in new publications today.

The rCRS sequence below is the Cambridge Reference Sequence [Anderson et al 1981](#) as revised by [Andrews et al 1999](#). It differs from the original CRS and other complete mtDNA GenBank sequences in that it has eighteen corrected or confirmed nucleotides as annotated below. See the [summary table](#) of the reanalysis by Andrews et al.

- **Seven nucleotides are considered to be rare polymorphisms** and were determined to be correct as originally sequenced ([J01415](#) gi:337188). Nucleotides [263A](#), [311C-315C](#), [750A](#), [1438A](#), [4769A](#), [8860A](#), and [14326A](#) are considered to be rare polymorphisms and are maintained as part of the true reference sequence. The seven rare polymorphisms are shown below in bold green capital letters.
- **Eleven nucleotide errors in the original CRS have been corrected** by re-sequencing the original placental material. Nucleotides [3107del](#), [3423T](#), [4985A](#), [9559C](#), [11335C](#), [13702C](#), [14199T](#), [14272C](#), [14365C](#), [14368C](#), [14766C](#) are corrections of the original Cambridge sequence ([J01415](#) gi:337188). The errors in the original Cambridge sequence have been attributed to sequencing errors (8 instances) and to the inclusion of bovine (2 instances) or HeLa (1 instance) DNA. See [summary table](#). Corrected sequencing errors are shown below in bold red capital letters. "[3107del](#)" is maintained in this revised sequence with the gap represented by an "N". **THIS ALLOWS HISTORICAL NUCLEOTIDE NUMBERING TO BE MAINTAINED.** [Note: We would have preferred to have used a "-", "d", or "x" at 3107, but these were not allowed in a Reference Sequence by GenBank. When submitting mtDNA sequences to GenBank, please do NOT include 3107 as an "N". Please submit your actual sequence or use "-" to indicate a gap.]
- The L-strand of the rCRS NC_012920 is shown. [View double-stranded version](#).
- For strand composition asymmetry and an explanation of L-strand/H-strand terminology, click [here](#).

```

1 gatcacaggt ctatcaacct attaacact cacgggagct cccatgcat ttggtatatt
61 cgtctggggg gtatgcaagc gatagcatt cgagacgctg gagccggagc accctatgtc
121 gcagtatctg tctttgattc ctgctcctc ctattatatta tgcgaactac gttcaatatt
181 acaggogaac atactacta aagtgtgta ataatatgct gctttagga cataataa
241 acaattgaat gtctgcaag ccctttcca cacagacatc ataacaaaa atttccacca
301 aacccccctt cccccgcttc tggccacagc acttaaacac atcttgcca aacccccaaa
361 acaaaagaac ctaaacaccg cctaaccaga ttccaattt tatcttttg cggtagcacc

```

Figure 3. The rCRS

Complete Mitochondrial DNA Sequences

The revised Cambridge Reference Sequence (rCRS) is GenBank number NC_012920.

Please use this new number when citing the rCRS in publications. The rCRS is a reference sequence, not a "consensus" sequence. It is a single reference individual from haplogroup H2a2 and has been used as a standard for reporting variants for over 30 years.

The Cambridge Reference Sequence, revised & original:

Version	GenBank #	Fasta format	Article links
Revised Cambridge Reference Sequence ("rCRS") Two identical versions of the rCRS are available on GenBank. NC_012920, formerly AC_000021.2, is in GenBank's RefSeq database. It is the most commonly used rCRS and is the standard comparison sequence for human mtDNA research. For new publications, please cite NC_012920 as the rCRS. J01415.2 is a fully corrected update of the original Cambridge sequence and is identical to NC_012920.	NC_012920 gi:251831106	rCRS-fasta	Andrews et al. 1999 (PubMed) <ul style="list-style-type: none"> • Read the paper (PDF) • Summary table of corrections.
Original Cambridge Reference Sequence ("CRS")	J01415 gi:337188	CRS1981-fasta	<ul style="list-style-type: none"> • Anderson et al. 1981 (PubMed) • Read the paper (PDF)

Other comparison hmtDNAs in GenBank & elsewhere:

African (Yoruba) Sequence [AF347015](#), formerly NC_001807.4. This sequence has over 40 variant nucleotides from the rCRS.
 African (Uganda) Sequence [D38112](#). This sequence has over 90 variant nucleotides from the rCRS.
 Swedish Sequence [X93334](#). This sequence has over 30 variant nucleotides from the rCRS.
 Japanese Sequence [AB055387](#). This sequence has over 50 variant nucleotides from the rCRS.
 Root Sequence of [Behar et al. 2012](#). This is an artificial sequence constructed for rooting phylogenetic trees, the "RSBS".

To find >17,000 complete* human mtDNAs in GenBank: [execute search](#) 

*includes sequences that are complete coding region but minus the control region (15400 nucleotides minimum).

To find >9,000 other complete eukaryote (non-human) mtDNA genomes in GenBank: [execute search](#) 

Partial and full sequences are also available for *Homo sapiens neanderthalensis* and *Homo sp. Altai* mtDNA.

Figure 4. Complete Mitochondrial DNA Sequences

The Human Mitochondrial Genetic Code

Translations: [ATP6](#), [ATP8](#), [ND1](#), [ND2](#), [ND3](#), [ND4](#), [ND4L](#), [ND5](#), [ND6](#), [CO1](#), [CO2](#), [CO3](#), [CytB](#).

Phe F	UUU UUC	Thr T	ACU ACC ACA ACG	Asp D	GAU GAC
Leu (1) L (UUA/G)	UUA UUG	Ala A	GCU GCC GCA GCG	Glu E	GAA GAG
Leu (2) L (CUN)	CUU CUC CUA CUG	Tyr Y	UAU UAC	Cys C	UGU UGC
Ile I	AUU* AUC	Ter	UAA UAG	Trp W	UGA UGG
Met M	AUA AUG	His H	CAU CAC	Arg R	CGU CGC CGA CGG
Val V	GUU GUC GUA GUG	Gln Q	CAA CAG	Ser (2) S (AGU/C)	AGU AGC
Ser (1) S (UCN)	UCU UCC UCA UCG	Asn N	AAU AAC	See note below+	AGA AGG
Pro P	CCU CCC CCA CCG	Lys K	AAA AAG	Gly G	GGU GGC GGA GGG

Figure 5. The mitochondrial genetic code

Estimated Worldwide Haplotype Frequencies (%)

Compiled for MitoMap by O. Derbeneva 2009
 Please note: These numbers are for illustrative purposes, not for solid calculations.
 They are simple means from published frequencies and not always have all haplogroups been typed.
 Frequencies ≥ 20% are bolded, frequencies ≥ 10% are highlighted.

REGION	A	B	C	D	E	F	G	H	I	J	K	L	M	N	R	S	T	U	V	W	X	Y	Z	n.d.
Africa								17	0	5	2	46	3	1	0		4	15	3	0	1			3
Middle East	1	<0.5	1	1		0	0	22	2	13	5	6	2	4	1	0	9	15	1	2	3	0	0	13
South East Asia			13																					87
Australia and Oceania			23									1	7	48	7									14
West Europe				1	0			41	2	9	5	1	1	1	0		8	18	7	2	2		0	3
East Europe	1	<0.5	2	2		1	1	35	2	8	4		2	1	1		11	22	3	1	1	0	1	-
Caucasus	1	<0.5	4	4		1		23	2	7	6	0	1	3	3		10	22	1	2	4		0	7
Central Asia	7	5	12	15		5	5	15	1	3	1	0	6	2	1		6	10	0	2	0	1	2	3
East Asia	7	16	5	26	0	11	4	1		1			15	3	2		0	<0.5		0		1	2	7
Siberia	8	3	30	19	0	4	5	4	0	2	0		3	3	1	1	2	6	0	0	0	4	1	5
North America		43	23	18	10	1																	3	2
South and Central Americas		28	27	21	19	1						2												2

Figure 6. Haplogroup Frequency Estimates

Haplogroup Markers

www.mitomap.org

Nucleotide positions for RFLPs represent the first base of the recognition sequence in the top strand, not the actual cut site.

"," = enzyme site absent, "+" = enzyme site present

All diagnostic markers are indexed to the rCRS, [NC_012920](#).

Haplogroup	Diagnostic RFLP site*	Diagnostic SNP <small>The rCRS nucleotide is on the left of the SNP position, the diagnostic value is on the right.</small>	HVS1 motif <small>All sites are transitions except where indicated.</small>
L0, L1, L2	+3592h	C3594T	no specific sites
L3	-3592h	C3594C	no specific sites
M	+10397a (+10394c)	C10400T + A10398G	no specific sites
N	-10397a and +10871z	C10400C + A10398A and T10873T	no specific sites
M subgroups:			
C	+13262a (-13259o)	A13263G	16223 16327 16298
D	-5176a	C5178A	16223 16362
E	-13619x	C13626T	16362 16390
G	+4830n (+4831f)	A4833G	16223 16278 16362
Q	no RFLP	A5843G	16129 16241 16311
Z	no RFLP	T9090C	16185 16223 16260 16298
N/R subgroups:			
R	+12704j	C12705C	16223C (=CRS)
B	8281-8289 nps 9bp del	8280:8290 =A[delCCCCCTCTA]G	16183C 16189 16217

Figure 7. Diagnostic RFLP and SNPs for Major Haplogroups

Human mtDNA Migrations

from <http://www.mitomap.org>

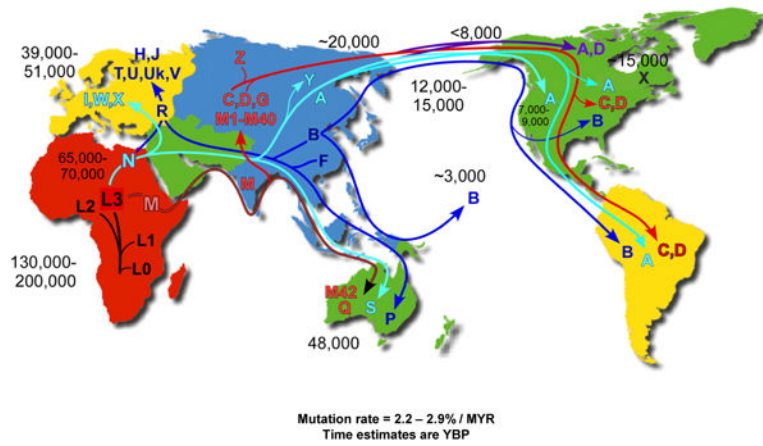


Figure 8. mtDNA Haplogroup Migration Map

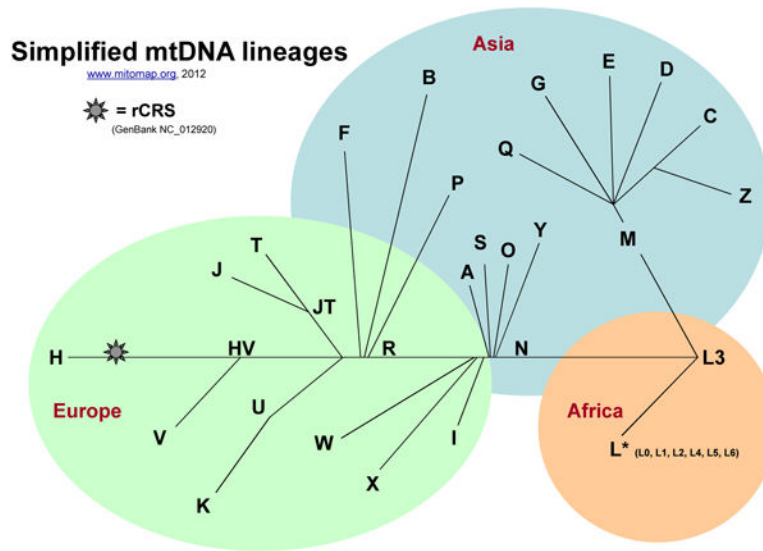


Figure 9. Simplified Mitochondrial Haplogroup Relationships

MITOMAP: Mitochondrial DNA Function Locations

Last Edited: Aug 18, 2009

Map Locus	Starting	Ending	Shorthand	Description	Reference
MT-HV2	57	372	CR.HVS2/HV2	Hypervariable segment 2 [classic:73-340]	references
MT-OHR	110	441	CR.OH	H-strand origin	references
MT-CSB1	213	235	CR.CSB1	Conserved sequence block 1	references
MT-TFX	233	260	CR.TFX	mtTF1 binding site	references
MT-TFY	276	303	CR.TFY	mtTF1 binding site	references
MT-CSB2	299	315	CR.CSB2	Conserved sequence block 2	references
MT-HPR	317	321	CR.HPR	replication primer	references
MT-CSB3	346	363	CR.CSB3	Conserved sequence block 3	references
MT-4H	371	379	CR.mt4H	mt4 H-strand control element	references
MT-3H	384	391	CR.mt3H	mt3 H-strand control element	references
MT-LSP	392	445	CR.PL	L-strand promoter	references
MT-TFL	418	445	CR.mtTF1	mtTF1 binding site	references
MT-HV3	438	574	CR.HVS3/HV3	Hypervariable segment 3	references
MT-TFH	523	550	CR.TFH	mtTF1 binding site	references
MT-HSP1	545	567	CR.PH1	Major H-strand promoter	references
MT-TF	577	647	F	tRNA phenylalanine	references
MT-HSP2	645	645	PH2	Minor H-strand promoter	references
MT-RNR1	648	1601	12S	12S ribosomal RNA	references
MT-TV	1602	1670	V	tRNA valine	references
MT-RNR2	1671	3229	16S	16S ribosomal RNA	references
MT-RNR3	3206	3229	-	5S-like sequence	references
MT-TER	3229	3256	-	Transcription terminator	references
MT-TL1	3230	3304	L(UUAG)	tRNA leucine 1	references
MT-NC1	3305	3306	NC1	non-coding nucleotides	--
MT-ND1	3307	4262	ND1	NADH Dehydrogenase subunit 1	references
MT-TI	4263	4331	I	tRNA isoleucine	references

Figure 10. Mitochondrial Functional Locations

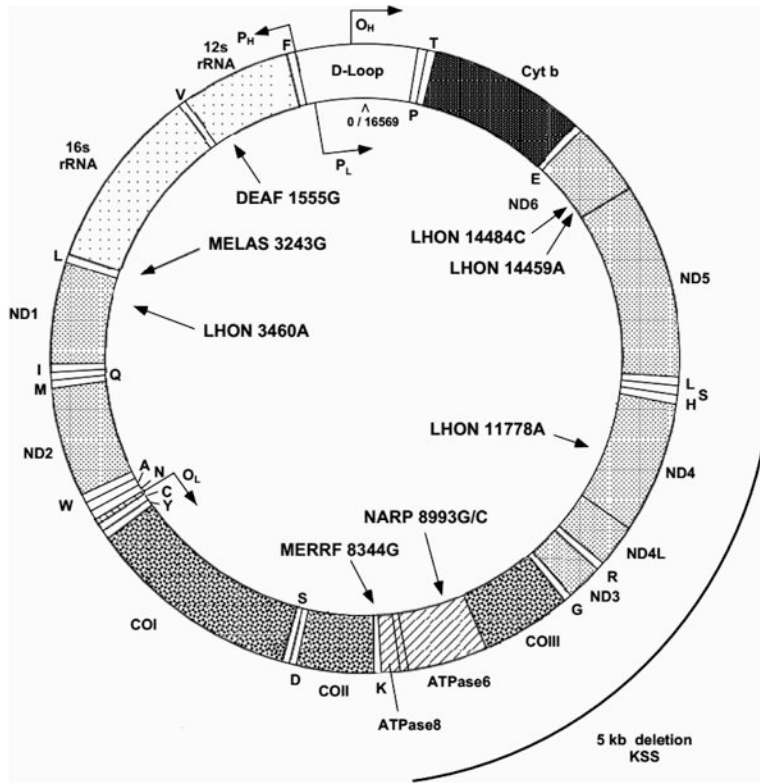


Figure 11. Gene map of the human mitochondrial DNA with representative disease variants shown

MITOMAP Quick Reference

- To search for point mutations, click [here](#). The info button on the search page has more information.
- The rCRS is GenBank number [NC_012920](#). Click [here](#) for details.


Figure 12a.
Search box for specific variant(s).

Allele Search

Start: End: [Reference Search](#) | [Site Search](#)

Enter a single nucleotide position or a range (up to 100 bps). Values must be positive integers from 1 to 16569.
You may also [search unpublished variants](#) submitted to MITOMAP.

Figure 12b.
Search box for specific variant(s). In this example, a single position is queried.

Allele Search 

Searched nucleotide position(s): 3240 - 3260

MITOMAP: mtDNA Coding Region Sequence Polymorphisms

Nucleotide Position	Locus	Nucleotide Change	Amino Acid Change	References
3250	MT-TER.MT-TL1	T-C	non-coding	references
3254	MT-TER.MT-TL1	C-A	non-coding	references
3254	MT-TER.MT-TL1	C-T	non-coding	references

MITOMAP: Reported Mitochondrial DNA Base Substitution Diseases: rRNA/tRNA Mutations

Locus	Disease	Allele	RNA	Homoplasmy	Heteroplasmy	Status	Reference
MT-TL1	MM / HCM+renal tubular dysfunction	G3242A	tRNA Leu (UUR)	+	+	Reported	references
MT-TL1	MELAS / LS	A3243G	tRNA Leu (UUR)	-	+	Cfm	references
MT-TL1	DMDF / MIDD / SNHL / FSGS / Cardiac+multi-organ dysfunction	A3243G	tRNA Leu (UUR)	-	+	Cfm	references
MT-TL1	MM / MELAS / SNHL / CPEO	A3243T	tRNA Leu (UUR)	-	+	Reported	references
MT-TL1	CPEO / MM	A3243G	tRNA Leu (UUR)	-	+	Cfm	references

Figure 12c. Results for variant search. In this example, a range was queried. This screenshot shows only a portion of the results returned.

mtDNA Variants (includes mini insertions & deletions)
UPDATED Control Region Variants (16024-576)
UPDATED Coding & RNA Variants (577-16023, MTTF-MTTP)
Somatic Mutations
mtDNA Mutations with Reports of Disease-Associations
Organized by mtDNA location:
rRNA/tRNA Mutations
Coding & Control Region Mutations

Figure 13.

Links to MITOMAP Variants. Variants are organized into two categories – general variants (top section) and those with reports of possible disease-associations (lower section).

MITOMAP: mtDNA Control Region Sequence Variants

Last edited: Aug 17, 2013

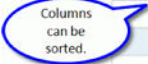
Columns can be sorted.

Nucleotide Position	Nucleotide Change	GB Frequency	Curated References
3	T-C	0	references
7	A-G	0	references
8	G-T	0	references
9	G-A	5	references
9	G-T	0	references
10	T-C	9	references
10	T-G	1	references
11	C-T	1	references
16	A-T	17	references
17	C-T	0	references
23	T-C	1	references
26	C-T	3	references

Figure 14a.
Control Region Variants.

MITOMAP: mtDNA Control Region Sequence Variants

Last Edited: Feb 13, 2013



Nucleotide Position	Nucleotide Change	GB Frequency	Curated References
3	T-C	0	references
7	A-G	0	references
8	G-T	0	references
9	G-A	5	references
9	G-T	0	references
10	T-C	9	references
10	T-G	1	references
11	C-T	1	references
16	A-T	16	references
17	C-T	0	references
23	T-C	1	references
26	C-T	3	references

Figure 14b.
Outside of the Control Region: Coding and RNA Variants.

MITOMAP: mtDNA Coding Region & RNA Sequence Variants

Last Edited: Feb 19, 2013

Locus	Nucleotide Position	Nucleotide Change	Codon number	Codon Position	Amino Acid Change	GB frequency	Curated References
MT-ND3	10373	G-A	105	3	syn	153	references
MT-ND3	10376	A-G	106	3	syn	9	references
MT-ND3	10379	A-G	107	3	syn	2	references
MT-ND3	10385	A-C	109	3	K-N	0	references
MT-ND3	10387	G-C	110	2	G-A	0	references
MT-ND3	10388	T-C	110	3	syn	0	references
MT-ND3	10389	T-C	111	1	syn	15	references
MT-ND3	10391	A-G	111	3	syn	1	references
MT-ND3	10394	C-T	112	3	syn	32	references
MT-ND3	10397	A-G	113	3	syn	119	references
MT-ND3	10398	A-G	114	1	T-A	7081	references

Figure 15.
Variant information can be sorted.

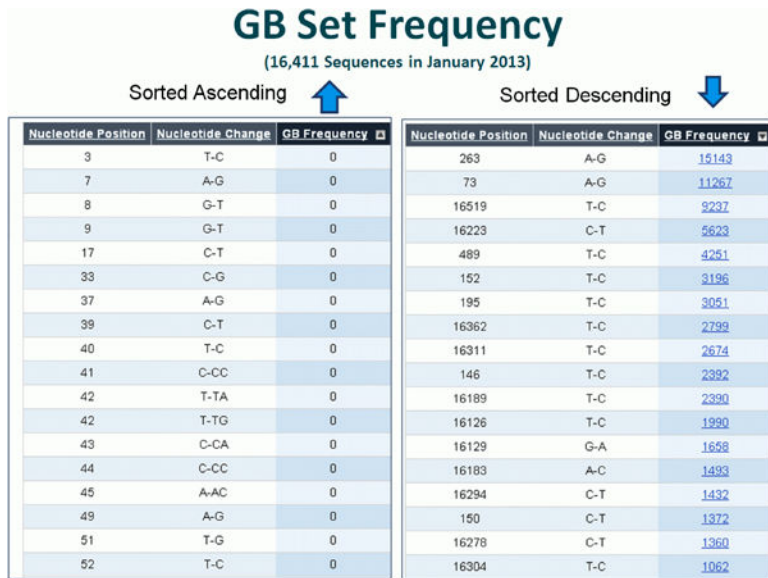


Figure 16. Details of GB set frequency

GenBank Record for Control Polymorphism A-T at rCRS position 16

click hyperlink to view GenBank Record, PubMed Reference, and MitoMaster running Results

Display 10 records

Copy CSV Excel PDF Print

Genbank ID	Pubmed Reference	Predicted Haplogroup	#Haplogroup with given polymorphism	Total # Haplogroup in GenBank	MitoMaster Results
DQ272108.1	16714301	G3a	3	9	view
EU523125.1		K1a	9	496	view
EU603402.1		K1a	9	496	view
EU623472.1		K1a	9	496	view
GU014566.1	19955425	G3a	3	9	view
GU014569.1	19955425	G3a	3	9	view
GU332639.1		K1a	9	496	view
JN657206.1		K1	3	59	view
JQ044858.1	22130972	L2c	2	48	view
JQ044989.1	22130972	L2c	2	48	view

Showing 1 to 10 of 17 entries

← Previous 1 2 Next →

Results out of 18363 sequences in Mitomap's GenBank Set

Figure 17. GenBank ID links to sequence

Genbank ID	Pubmed Reference	Predicted Haplogroup	#Haplogroup with given polymorphism	Total # Haplogroup in GenBank	MitoMaster Results
JQ044989.1	22130972	L2c	2	48	view

NCBI Resources | How To | Search | JQ044989.1

Nucleotide | Nucleotide | Links | Advanced | Search | Help

Display Settings | GenBank | Send | Change region shown | Customise view

Homo sapiens isolate BF047 mitochondrion, complete genome

GenBank: JQ044989.1

FASTA | Statistics

Go to: |

LOCUS JQ044989 16544 bp DNA circular PR2 29-MAR-2012

DEFINITION Homo sapiens isolate BF047 mitochondrion, complete genome.

ACCESSION JQ044989

VERSION JQ044989.1 GI:159469905

KEYWORDS

SOURCE

ORGANISM Homo_sapiens

Excerpted from: (bacteria) Chlamydia; (eukaryota) Eukaryota; (metazoa) Metazoa; (chordata) Chordata; (mammalia) Mammalia; (primates) Primates; (hominidae) Hominidae; (homo) Homo; (homo_sapiens) Homo_sapiens; (chromosomes) Chromosomes; (mitochondrion) Mitochondrion; (mitochondrial DNA) Mitochondrial DNA.

REFERENCE 1 (bases 1 to 16544)

AUTHOR Baheti,C., Whitten,R., Beyer,K., Subeibar,S., Li,S. and Pabendorf,R.

TITLE Contrasting maternal and paternal histories in the linguistic context of human Oa3

JOURNAL Mol. Biol. Evol. 29 (4), 1213-1223 (2012)

PMID 22130972

REFERENCE 2 (bases 1 to 16544)

AUTHOR Baheti,C., Whitten,R., Li,S. and Pabendorf,R.

TITLE Pleistocene Subequatorial Forest Disappearance: A Population Genetic Perspective

JOURNAL Molecular Biology and Evolution 29(10):2011-2021 (2012)

PMID 22130972

FEATURES

Source

1..16544

/organism="Homo sapiens"

/organism_locus="Mitochondrion"

/mol_type="genomic DNA"

/label="BF047"

/db_xref="taxon:9606"

/db_xref="taxon:33203"

/taxonomy="L2c"

Related information

Related Sequences

Gene

GeneView in dSNAP

Protein

PubMed

PubMed (MeSH)

Taxonomy

Recent activity

Contrasting maternal and paternal histories in the linguistic context of human Oa3

The initial peopling of the Americas: a growing number of founding mitochondrial lineages

Homo sapiens isolate BF047 mitochondrion, complete genome

Figure 18. Pub Med ID links to publication

Genbank ID	Pubmed Reference	Predicted Haplogroup	#Haplogroup with given polymorphism	Total # Haplogroup in GenBank	MitoMaster Results
JQ044989.1	22130972	L2c	2	48	view

↓

GenBank Record for haplogroup L2c

click hyperlinks to view GenBank Record, PubMed Reference, and MitoMaster running Results

Display 10 records Copy CSV Excel PDF Print

Genbank ID	Pubmed Reference	MitoMaster Results
AF346995.1	11130070	view
AF381981.1	11553319	view
AV195785.2	12509511	view
DQ112706.2	16172908	view
DQ112707.2	16172908	view
DQ112735.2	16172908	view
DQ112883.3	16172908	view
DQ304966.1	19083815	view
DQ304967.1	19083815	view
DQ304968.1	19083815	view

Showing 1 to 10 of 48 items -- Previous 1 2 3 4 5 Next --

Check other Haplogroup

Haplogroup Name

Figure 19. Predicted Haplogroup links to total listing of sequences with that haplogroup in the GenBank sequence set

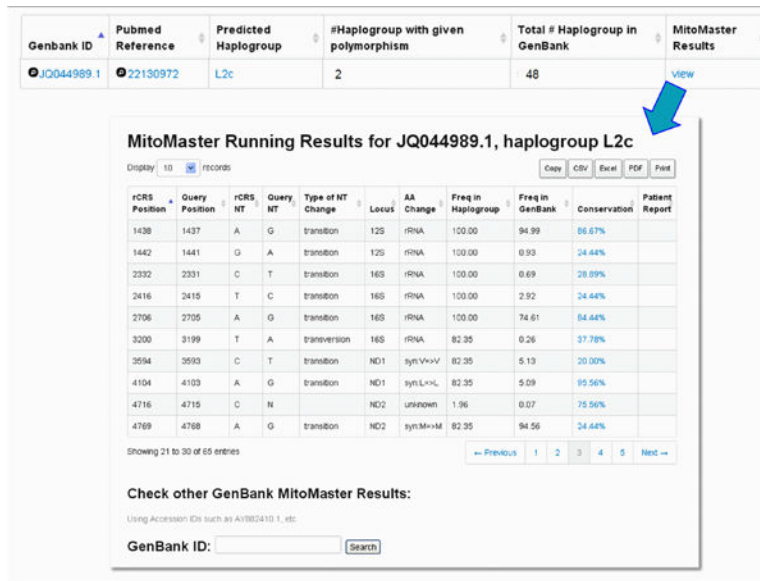


Figure 20. MITOMASTER analysis report on each sequence

MITOMAP: Reported Mitochondrial DNA Base Substitution Diseases: Coding and Control Region Point Mutations

Last Edited Feb 18, 2013

Locus	Disease	Allele	Nucleotide Position	Nucleotide Change	Amino Acid Change	Homo-plasmy	Hetero-plasmy	Status	References
MT_CO1	MM & Rhabdomyolysis	G6708A	6708	G-A	G-Ter	-	+	Reported	references
MT_CO1	Acquired Idiopathic Sideroblastic Anemia	T6721C	6721	T-C	M-T	-	+	Reported	references
MT_CO1	Acquired Idiopathic Sideroblastic Anemia	T6742C	6742	T-C	L-T	-	+	Reported	references
MT_CO1	Multisystem Disorder	G6930A	6930	G-A	G-Ter	-	+	Reported	references
MT_CO1	Mid EXIT and MR	G6955A	6955	G-A	G-D	+	+	Reported	references
MT_CO1	MELAS-like syndrome	G7023A	7023	G-A	V-M	-	+	Reported	references
MT_CO1	Prostate Cancer	G7041A	7041	G-A	V-I	+	-	Reported	references
MT_CO1	Prostate Cancer	T7090C	7090	T-C	F-L	+	-	Reported	references
MT_CO1	Prostate Cancer	A7083G	7083	A-G	L-V	+	-	Reported	references
MT_CO1	Prostate Cancer	A7158G	7158	A-G	L-V	+	-	Reported	references
MT_CO1	Prostate Cancer	A7305C	7305	A-C	M-L	+	-	Reported	references
MT_CO1	DEAF	A7443G	7443	A-G	Ter-G	+	-	Reported	references
MT_CO1	LHONSNHLDEAF	G7444A	7444	G-A	Ter-K	+	-	Reported	references
MT_CO1	DEAF	A7445C	7445	A-C	Ter-S	+	-	Reported	references
MT_CO1	SNHL	A7445G	7445	A-G	Ter-Ter	+	+	Cfm	references
MT_CO2	Mitochondrial Encephalomyopathy	T7587C	7587	T-C	M-T	-	+	Reported	references
MT_CO2	Possible LHON helper variant	G7598A	7598	G-A	A-T	-	+	Reported	references

Figure 21. Reported Coding and Control Region mutations found in patients

MITOMAP: Reported Mitochondrial DNA Base Substitution Diseases: rRNA/tRNA mutations

Last Edited: Nov 25, 2012

Locus	Disease	Allele	RNA	Homo- plasmy	Hetero- Plasmy	Status	References
MT-TE	Mitochondrial myopathy	T582C	tRNA Phe	-	+	Reported	references
MT-TE	MELAS / MM & EXIT	G583A	tRNA Phe	-	+	Cfm	references
MT-TE	Extrapyramidal disorder with akinesia-rigidity, psychosis and SNHL	G586A	tRNA Phe	-	+	Reported	references
MT-TE	Axial myopathy with encephalopathy	C602T	tRNA Phe	-	+	Reported	references
MT-TE	Myoglobinuria	A606G	tRNA Phe	+	+	Unclear	references
MT-TE	Tubulo-interstitial nephritis	A608G	tRNA Phe	+	-	Reported	references
MT-TE	MERRF	G611A	tRNA Phe	-	+	Reported	references
MT-TE	Maternally inherited epilepsy	T616C/G	tRNA Phe	+	+	Reported	references
MT-TE	MM	T618C	tRNA Phe	-	+	Reported	references
MT-TE	Ptosis CPEO MM & EXIT	T618G	tRNA Phe	-	+	Reported	references
MT-TE	EXIT & Deafness	G622A	tRNA Phe	-	+	Reported	references
MT-TE	SNHL & Epilepsy	G625A	tRNA Phe	-	+	Reported	references
MT-TE	DEAF	A636G	tRNA Phe	+	-	Reported	references

Figure 22. Reported rRNA and tRNA mutations found in patients

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Limits Advanced Search Help

Display Settings: GenBank Send: Change region shown Customize view Analyze this sequence Run BLAST Pick Primers Highlight Sequence Features Find in this Sequence Related information Related Sequences Full text in PMC PopSet Protein PubMed PubMed (Weighted) Taxonomy

Homo sapiens isolate ND6/J1 (Tor578) mitochondrion, complete genome
GenBank EU915478.1
[FASTA](#) [Graphics](#) [PopSet](#)

Seq ID: [+](#)

LOCUS EU915478 16569 bp DNA circular F01 12-DEC-2008
DEFINITION Homo sapiens isolate ND6/J1 (Tor578) mitochondrion, complete genome.
ACCESSION EU915478
VERSION EU915478.1 GI:194441041
KEYWORDS .
SOURCE mitochondrion Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 16569)
AUTHORS Pello, R., Martin, M.A., Carelli, V., Nijtmans, L.G., Achilli, A., Pala, M., Torroni, A., Gomez-Duran, A., Buit-Pesini, E., Martinuzzi, A., Smetnik, J.A., Arenas, J. and Ugalde, C.
TITLE Mitochondrial DNA background modulates the assembly kinetics of OXPHOS complexes in a cellular model of mitochondrial disease
JOURNAL Hum. Mol. Genet. 17 (24), 4001-4011 (2008)
PubMed 18806272

Figure 23. Sample Sequence EU915478 at GenBank

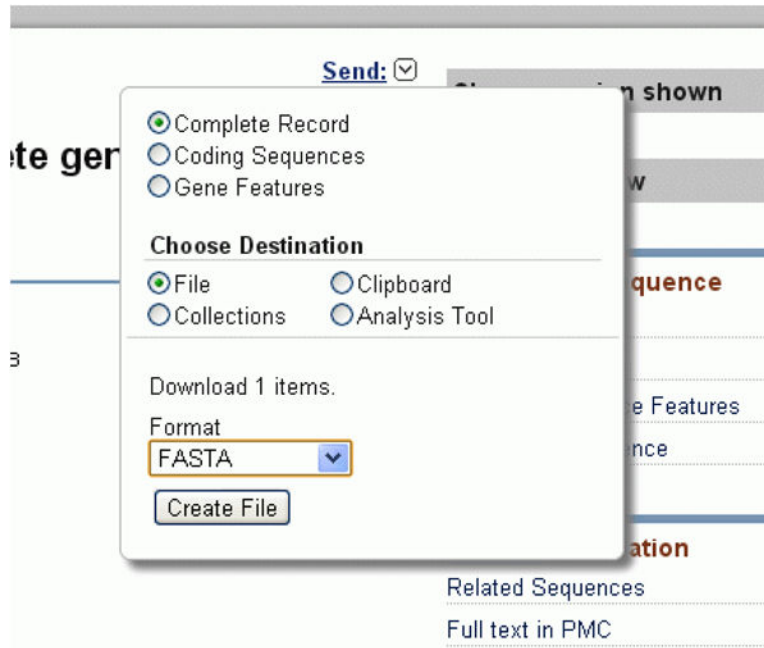


Figure 24. Send FASTA dialog box

Mitomaster **Sequence Query** GenBank Query SNV Query Species Use Web Service

Submit sequence as file upload or paste into text area

Step 1 [Select conservation species of interest](#)

Step 2 **Fasta file** use this for more than one sequence

Option 1

Option 2 **Copy-paste**

Paste a single raw unheadered sequence here. Minimum length: 100bp

Figure 25. Submit Sequence screen

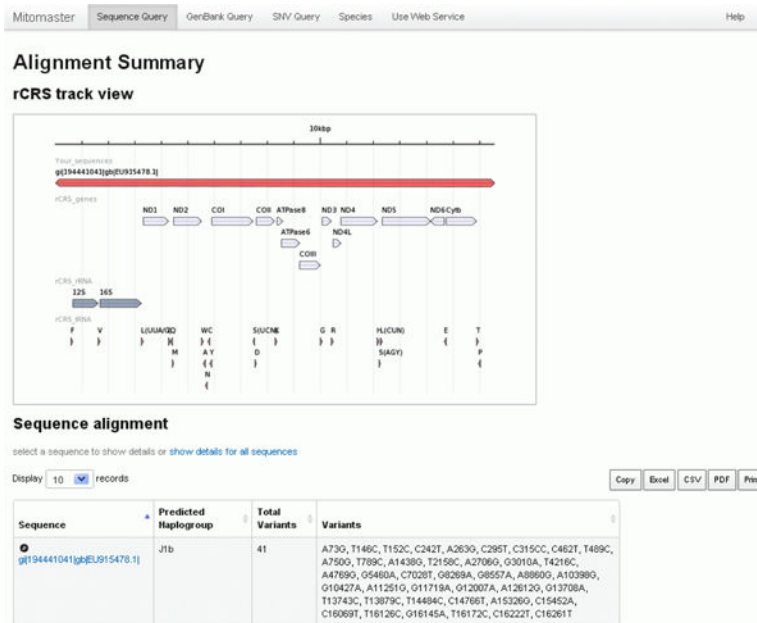


Figure 26. Alignment Summary

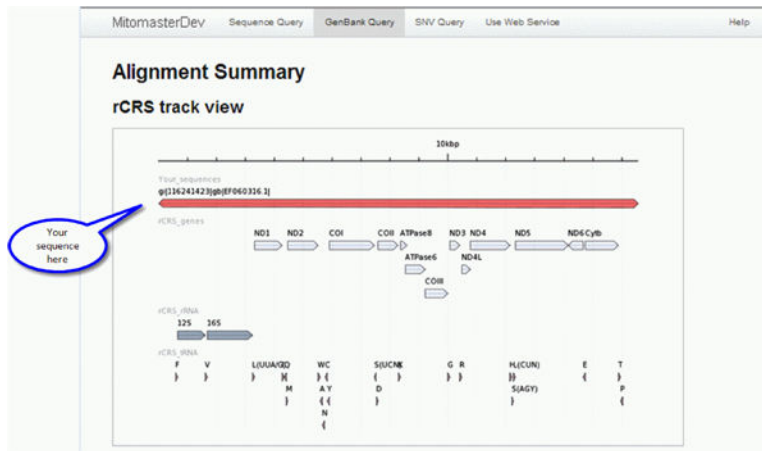


Figure 27. rCRS Track View

Sequence	Predicted Haplogroup	Total Variants	Variants
g[194441041]g[EU915478.1]	-J1b	41	A73G, T148C, T152C, C242T, A263G, C295T, C3150C, C462T, T489C, A750G, T789C, A1438G, T2156C, A2706G, G3010A, T4216C, A4766G, G5469A, C7028T, G8591A, G8557A, A8860G, A10398G, G10427A, A11251G, G11719A, G12007A, A12612G, G13708A, T13743C, T13878C, T14484C, C14796T, A15326G, C15452A, C18069T, T16126G, G16145A, T16172C, C16222T, C16261T

Figure 28. Sequence Alignment, areas of interest circled

Mitomaster Sequence Query GenBank Query SNV Query Species Use Web Service Help

Sequence run 23981 / g[194441041]g[E915478.1]

Alignment Details

Display 10 records Copy Excel CSV PDF Print

rCRS Position	Query Position	rCRS NT	Query NT	Mut type	Locus	tx Effect	Freq % in GB	Freq % in J1b	Refs	Conservation	Patient Report
73	73	A	G	transition	HVS2/ATT/D-Loop/7S DNA	non-coding	71.33	98.31	refs		
146	146	T	C	transition	HVS2/OH/ATT/D-Loop/7S DNA	non-coding	19.68	10.17	refs		
152	152	T	C	transition	HVS2/OH/ATT/D-Loop/7S DNA	non-coding	20.73	9.32	refs		
242	242	C	T	transition	HVS2/OH/TFX/ATT/D-Loop	non-coding	0.35	50.85	refs		
263	263	A	G	transition	HVS2/OH/ATT/D-Loop	non-coding	90.50	98.31	refs		
295	295	C	T	transition	HVS2/OH/TFY/ATT/D-Loop	non-coding	4.56	97.46	refs		
315	315	C	CC	insertion	HVS2/OH/ATT/D-Loop	non-coding	80.95	88.98			
462	463	C	T	transition	ATT/D-Loop	non-coding	3.58	98.31	refs		
489	490	T	C	transition	ATT/D-Loop	non-coding	24.79	98.31	refs		
750	751	A	G	transition	12S	rRNA	98.48	100.00	refs	97.78%	

Showing 1 to 10 of 41 entries ← Previous 1 2 3 4 5 Next →

Figure 29. Alignment Detail

Mitomaster [Sequence Query](#) [GenBank Query](#) [SNV Query](#) [Species](#) [Use Web Service](#) [Help](#)

Cerbank run 34000 : p[134441041]p[EV91547B.1]

Alignment Details

Display 10 records Copy Excel CSV PDF Print

rCRS Position	Query Position	rCRS NT	Query NT	Mut type	Locus	tx Effect	Freq % in GB	Freq % in J1B	Refs	Conservation	Patient Report
10398	10398	A	G	transition	ND3	non-syn T->A	43.41	98.31	refs	91.11%	PD protective factor/longevity/alterred cell ph/metabolic syndrome/breast cancer risk
4216	4216	T	C	transition	ND1	non-syn Y->H	10.11	98.31	refs	24.44%	LHON/Insulin Resistance
13708	13708	G	A	transition	ND5	non-syn A->T	7.16	99.15	refs	33.33%	LHON/increased MS risk/higher freq in PD-ADS
14484	14484	T	C	transition	ND6	non-syn M->Y	0.19	0.85	refs	31.11%	LHON
3010	3011	G	A	transition	16S	rRNA	16.93	98.31	refs	20.00%	Cyclic Vomiting Syndrome with Migraine
5460	5460	G	A	transition	ND2	non-syn A->T	5.47	66.95	refs	4.44%	ADRPD
750	751	A	G	transition	12S	rRNA	98.48	100.00	refs	97.78%	
11719	11719	G	A	transition	ND4	syn G->G	74.29	100.00	refs	97.78%	
11251	11251	A	G	transition	ND4	syn L->L	9.51	99.15	refs	83.33%	
789	790	T	C	transition	12S	rRNA	0.15	0.85	refs	91.11%	

Showing 1 to 10 of 41 entries ← Previous 1 2 3 4 5 Next →

Figure 30. Alignment Details, areas of interest circled

gij194441041|gb|EU915478.1 conservation grid

Conservation denotes the percentage of residues that match rCRS among selected species. This query sequence mutation is depicted in isolation - i.e. other mutations in your sequence are not shown - and flanked by rCRS reference for clarity. Loci are shown in their transcribed orientation.

Species	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Mutation at rCRS pos 10398 (locus pos 114) of ND3						A					
Homo sapiens	K	G	L	D	W	T	E	-	-	-	-
Cebus albifrons	K	G	L	D	W	V	D	-	-	-	-
Gorilla gorilla	K	G	L	D	W	T	E	-	-	-	-
Hylobates lar	K	G	L	D	W	V	E	-	-	-	-
Macaca sylvanus	K	G	L	D	W	A	E	-	-	-	-
Nycticebus coucang	K	G	L	E	W	Q	E	-	-	-	-
Pan paniscus	K	G	L	D	W	A	E	-	-	-	-
Pan troglodytes	K	G	L	D	W	T	E	-	-	-	-
Papio hamadryas	K	G	L	D	W	T	-	-	-	-	-
Pongo abeli	K	G	L	D	W	A	E	-	-	-	-
Pongo pygmaeus	K	G	L	D	W	T	E	-	-	-	-
Tarsius bancanus	K	G	L	E	W	T	E	-	-	-	-
Boa taurus	K	G	L	E	W	T	E	-	-	-	-
Hippopotamus amphibius	K	G	L	E	W	T	E	-	-	-	-
Ovis aries	K	G	L	E	W	T	E	-	-	-	-
Sus scrofa	K	G	L	E	W	A	E	-	-	-	-
Viverra zibetha	Q	G	L	E	W	T	E	-	-	-	-
Canis lupus familiaris	K	G	L	E	W	T	E	-	-	-	-
Felis catus	K	G	L	E	W	T	E	-	-	-	-
Halichoerus grypus	K	G	L	E	W	T	E	-	-	-	-
Phoca vitulina	K	G	L	E	W	T	E	-	-	-	-
Balaenoptera musculus	E	G	L	E	W	A	E	-	-	-	-
Balaenoptera physalus	E	G	L	E	W	A	E	-	-	-	-
Ceratotherium simum	K	G	L	E	W	A	E	-	-	-	-
Equus asinus	K	G	L	E	W	T	E	-	-	-	-
Equus caballus	K	G	L	E	W	T	E	-	-	-	-
Bison bonasus	K	G	L	E	W	T	E	-	-	-	-

Figure 31A. Conservation Grid

Mitomaster Sequence Query GenBank Query SNV Query **Species** Use Web Service Help

Conservation index species

All species are used by default. Select a subset of species for the conservation index calculation in genes, tRNAs, and rRNAs.
 Mouse over to view common name.

Mammals

Primates

<input checked="" type="checkbox"/> Homo sapiens (rCRS)	<input checked="" type="checkbox"/> Cebus albifrons	<input checked="" type="checkbox"/> Gorilla gorilla	<input checked="" type="checkbox"/> Hylobates lar
<input checked="" type="checkbox"/> Macaca sylvanus	<input checked="" type="checkbox"/> Nycticebus coucang	<input checked="" type="checkbox"/> Pan paniscus	<input checked="" type="checkbox"/> Pan troglodytes
<input checked="" type="checkbox"/> Papio hamadryas	<input checked="" type="checkbox"/> Pongo abelli	<input checked="" type="checkbox"/> Pongo pygmaeus	<input checked="" type="checkbox"/> Tarsius bancanus

Artiodactyla

<input type="checkbox"/> Bos taurus	<input type="checkbox"/> Hippopotamus amphibius	<input type="checkbox"/> Ovis aries	<input type="checkbox"/> Sus scrofa
<input type="checkbox"/> Vicugna pacos			

Figure 31B. Conservation Grid of Species

The screenshot shows the Mitomaster web interface. At the top, there is a navigation bar with the following links: Mitomaster, Sequence Query, GenBank Query (which is highlighted), SNV Query, Species, Use Web Service, and Help. Below the navigation bar, the main heading is "Submit GenBank Identifiers". Underneath, there are two steps: "Step 1 Select conservation species of interest" and "Step 2 Input Accession(s) or GI(s)". Step 2 is the active step. Below the heading for Step 2, there is a text input field containing "EF060316" and a "Search" button. Above the input field, there is a small text box that reads: "Allowed values: singletons (EF060316 93116889), a comma-separated list (EF060316,302376313,DQ112752), or a range (AF346963-AF346968)".

Figure 32. Submitting GenBank Identifiers

Mitomaster [Sequence Query](#) [GenBank Query](#) **[SNV Query](#)** [Species](#) [Use Web Service](#) [Help](#)

Submit variant list

 as file upload or paste into text area

 Step 1 **Select conservation species of interest**

 Compute haplotype (only recommended if supply a full set of SNVs)

 Step 2 **Copy-paste**

 Option 1

sample	pos	ref	var
Sample1	73	A	G
Sample1	146	T	C
Sample2	263	A	:
Sample2	709	:	G

The variant list can follow either of these formats:

- Tabular ([load example](#))
 - Either no header or the following:

sample	pos	ref	var
--------	-----	-----	-----
- Four tab-delimited columns:
 - sample_name (with no spaces)
 - mitochondrial position
 - reference base
 - variant base
- Insertions are designated using a '*' or empty cell for the reference
- Deletions are designated using a '.' or empty cell for the variant
- Lines beginning in # are ignored for user comments
- Compact ([load example](#))
 - One variant per line, may be tab-prefaced with a sample column
 - Substitutions: 9028T or C9028T
 - Insertions: C573CC or 573insC or 573_1C or :573C
 - Deletions: 291d or A291d or A291del or A291- or A291

Figure 33. Single Nucleotide Variant Submission