Understanding Mitomap Variants

Modified from (3)

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 December 2013 this data set contained 18363 sequences. The size of the sequence set is expected to increase with quarterly scans of GenBank and possibly other public sequence repositories. Caution is advised when considering the frequency of any particular variant in GenBank. GenBank sequences may not be of equal quality (4), published results may have sequencing errors, partial data, and analysis mistakes that, even if corrected, still might not be downloadable as a corrected sequences directly from GenBank and might only be found as published erratum to the corresponding publications. Please note that pre-loaded sequences from GenBank have not been individually reviewed by MITOMAP. Also it is important to keep in mind that any human mitochondrial sequence has the potential to have past, current or future disease-relevant variants. More and more sequences from patients with known mitochondrial disease are being published.

If a mitochondrial DNA variant has been found in a published paper with reports of possible disease-association (validated or not), it is listed in the "disease reports" section of Mitomap. A researcher who is interested in this report should click on the "references" link to get the citation for the original publication for more information.

There are a few cases where variants are listed in both the "disease reports" table and the "population variants" table. This is simply because someone has reported the variant in question in a patient, while other investigators have found it in a general or non-defined population sample. Of interest will be the frequencies of this particular variant in the GenBank database, which contains sequences from the general population as well as some individuals with well-defined disease. For an extreme example, variant 750 A-G is found in 18,084 out of the 18,363 sequences as of January 2014, indicating that 750G can be considered a "consensus" allele, found in most human mtDNA samples; 750A is a rare allele which happens to be in the reference mtDNA sequence (the "rCRS", GenBank sequence number NC_012920 (1, 2)). The 663 A-G variant is

found in many samples (531 of the 18363) but is not as widespread as A750G. Clicking on the GB link (currently showing "531") for the 663G sequence information reveals that all but three of these sequences are of the haplogroup A type. [Note: it takes several minutes for this analysis to run, so please wait for the data to appear.] Reports of patients with this variant could simply indicate that they might be a member of haplogroup A.

Finally, mitochondrial DNA is highly variable in and between populations and new variants findings will be made with more and more sequencing studies performed around the world. 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") is not an assignment of pathogenicity by MITOMAP 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 and to avoid false reports on "novel" and "pathogenic" mtDNA variants.

1. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., and Young, I.G. 1981. Sequence and organization of the human mitochondrial genome. Nature 290(5806):457-465. http://www.ncbi.nlm.nih.gov/pubmed?term=7219534
2. Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics 23(2):147. http://www.ncbi.nlm.nih.gov/pubmed?term=10508508
3. Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V., and Wallace, D.C. 2013. mtDNA variation and analysis using MITOMAP and MITOMASTER (http://www.mitomap.org and http://mitomaster.mitomap.org). Current Protocols in Bioinformatics 44:1.23.1-1.23.26. http://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi0123s44/full
4. Yao, Y.G., Salas, A., Logan, I., and Bandelt, H.J. 2009. mtDNA data mining in GenBank needs surveying. American Journal of Human Genetics 85(6):929-933; author reply 933. http://www.ncbi.nlm.nih.gov/pubmed?term=20004768