

MITOMAP: a human mitochondrial genome database

Andreas M. Kogelnik^{1,2}, Marie T. Lott¹, Michael D. Brown¹, Shamkant B. Navathe² and Douglas C. Wallace^{1,*}

¹Department of Genetics and Molecular Medicine, Emory University School of Medicine, 1462 Clifton Road, Suite 446, Atlanta, GA 30322, USA and ²Bioengineering Program, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280, USA

Received September 12, 1995; Accepted September 20, 1995

ABSTRACT

We have developed a comprehensive database (MITOMAP) for the human mitochondrial DNA (mtDNA), the first component of the human genome to be completely sequenced [Anderson *et al.* (1981) *Nature* 290, 457–465]. MITOMAP uses the mtDNA sequence as the unifying element for bringing together information on mitochondrial genome structure and function, pathogenic mutations and their clinical characteristics, population associated variation, and gene–gene interactions. As increasingly larger regions of the human genome are sequenced and characterized, the need for integrating such information will grow. Consequently, MITOMAP not only provides a valuable reference for the mitochondrial biologist, it may also provide a model for the development of information storage and retrieval systems for other components of the human genome.

INTRODUCTION

The human mtDNA is a 16 569 nucleotide pair (np) closed, circular molecule located within the cytoplasmic mitochondria (Fig. 1). Each of the several thousand mtDNAs per cell encodes a control region encompassing a replication origin and the promoters, a large (16S) and small (12S) rRNA, 22 tRNAs and 13 polypeptides. All of the mtDNA polypeptides are components of the mitochondrial energy generating pathway, oxidative phosphorylation (OXPHOS), which is functionally essential and evolutionarily constrained (3).

The maternally inherited mtDNA has a very high mutation rate (2). This has resulted in a wide variety of pathologic mutations

and neutral polymorphisms. MITOMAP attempts to integrate the broad spectrum of available molecular, genetic, functional and clinical information into a single unified entity which can be queried from a variety of different perspectives.

MITOMAP architecture

MITOMAP is currently implemented using the Sybase relational database management software. It is both a self-contained information system for the mitochondrial biologist and also provides the mtDNA map and clinical information resources for the Genome DataBase (GDB, Johns Hopkins University) of the international Human Genome Organization (HUGO) and for On-Line Mendelian Inheritance of Man (OMIM, Johns Hopkins University and the National Library of Medicine).

MITOMAP is divided into five interrelated elements: the 'standard' mtDNA sequence (1), the functional genetic element data set, the clinical mutation data set, the population variation data set, and the gene–gene interaction data set (Fig. 2 and Table 1).

The functional genetic element data set provides the genomic location of the known functional domains of the mtDNA, defined by nucleotide position. It also provides information on the amino acid sequence of proteins, structure of RNAs and sequences of the regulatory elements.

The clinical mutation data set provides the nucleotide positions and base changes of the over 50 base substitutions that have been associated with disease. It also encompasses information on over 100 mtDNA rearrangements that have been characterized, including nucleotide positions of breakpoint junctions and sequences of associated repeat elements. The clinical characteristics associated with the mutations are accessible both through associated data sets of MITOMAP as well as through linkage to OMIM.

Table 1. Summary of mutations among the major regions of mtDNA

mtDNA Regions	Total no. genes	Total no. nucleotides	No. of mutations with disease	Total no. of polymorphisms
mRNA	13	11 388	25	183
tRNA	22	1509	28	33
rRNA	2	2513	2	56
Control region ^a	16	1121	0	310

^aThe control region spans roughly from between tRNA phenylalanine and tRNA proline

* To whom correspondence should be addressed

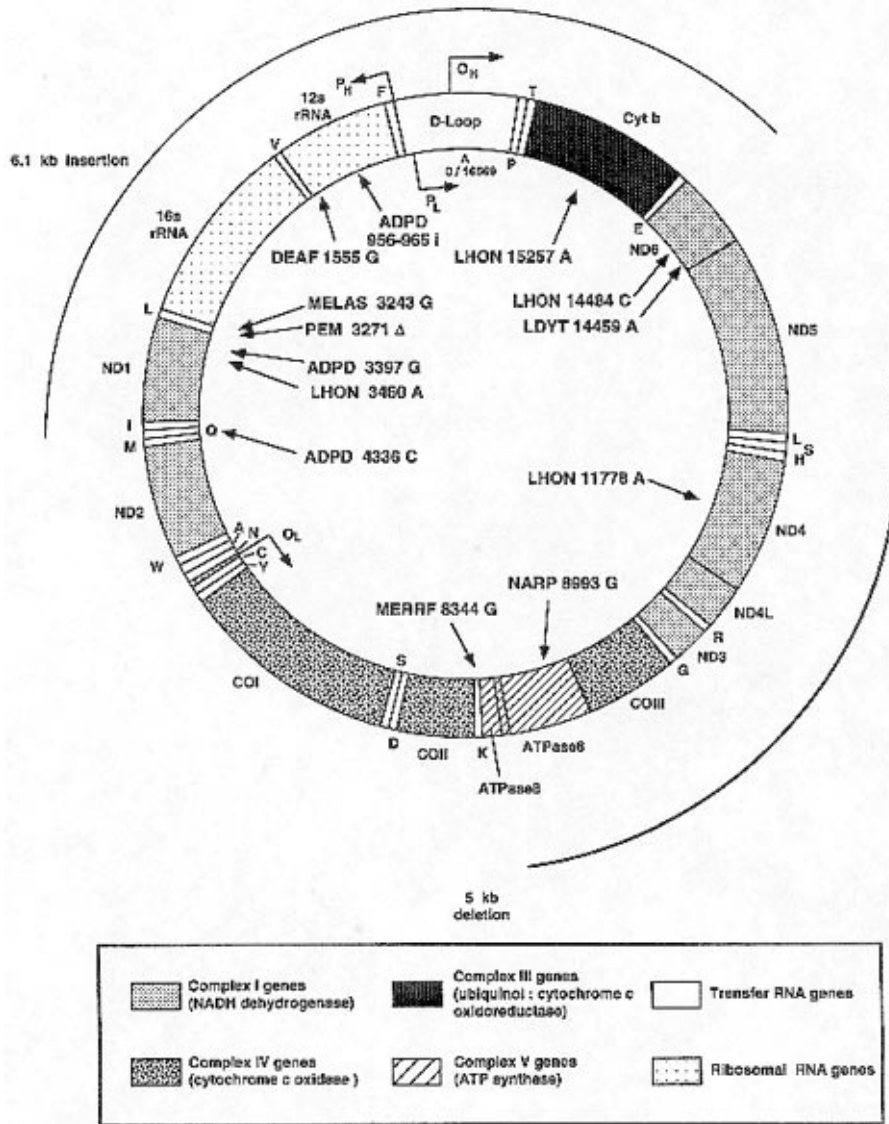


Figure 1. Human mtDNA structural and morbid map. The tRNAs are indicated by their cognate amino acid single letter code, and gene designations are defined in (3). The genes encoded by the G-rich heavy (H) strand are on the outside of the circle, while those for the C-rich light (L) strand are on the inside. The H- and L-strand origins (O_H and O_L) and promoters (P_H and P_L) are shown. The positions of representative pathologic base substitutions are shown on the inside of the circle. ADPD, Alzheimer and Parkinson disease; DEAF, neurosensory hearing loss; LHON, Leber's hereditary optic neuropathy; LDYT, Leber's hereditary optic neuropathy and dystonia; MELAS, mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes; MERRF, myoclonic epilepsy and ragged red muscle fibers; NARP, neurogenic muscular weakness, ataxia and retinitis pigmentosa. Two representative rearrangements are shown outside the circle, the common 5 kb deletion associated with ocular myopathy and Pearson marrow/pancreas syndrome and the 6.1 kb insertion observed in the maternally inherited diabetes and deafness family harboring the trimolecular heteroplasmy (1,3,4). Reproduced with permission of the American Society for Human Genetics.

The population variation data set provides access to known polymorphic sites. These include restriction site polymorphisms, small insertion-deletion variants, and identified sequence changes. The population associations of these variants are provided through available information of mtDNA haplotypes and the continental distributions and population frequencies of the more informative markers.

The gene-gene interactions data set catalogs known information on the polypeptide associations within the OXPHOS enzymes. It also provides information on nuclear genes which impinge on mtDNA structure and function.

MITOMAP access

The mitochondrial database is available to the general public through the World Wide Web (<http://www.gen.emory.edu/mitomap.html>). The interface provides both browsing and querying capabilities. Users can browse through the database in its published flat file format [adapted from (3)]. A query interface is provided to perform searches on specific aspects of the mitochondrial genome. For example, the database can be used to answer questions such as 'Are there any polymorphisms or mutations

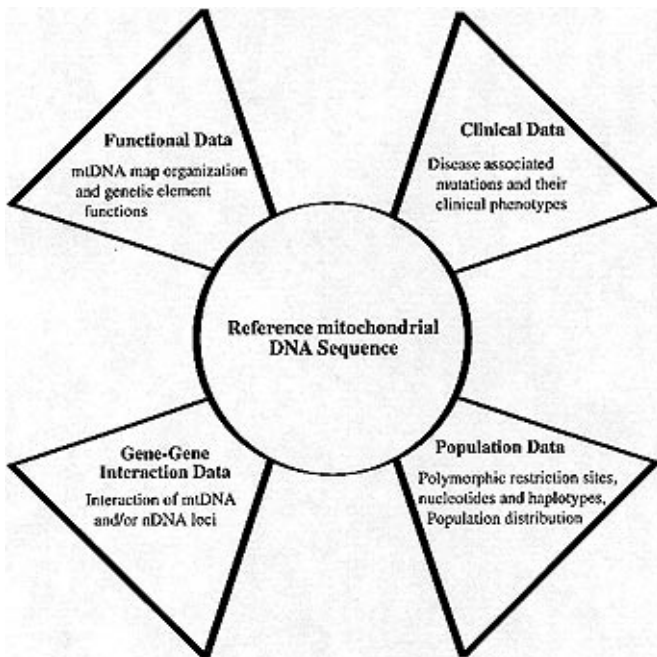


Figure 2. The primary database areas and their interrelationship with examples of the types of information which each currently documents. nDNA, nuclear DNA and mtDNA, mitochondrial DNA.

within a particular gene, and if so, have they been reported to cause disease?

Data acquisition

Data is taken from published works on the mitochondrial genome. The committee (see ref. 3) regularly searches the literature for new publications. The database is updated as new data is obtained, flat files are generated daily from the database, and queries are performed on-the-fly for the most up-to-date results. Submissions should be sent to Dr Douglas C. Wallace, Attn: Mitochondrial Genome Committee, Department of Genetics and Molecular Medicine, Emory University, 1462 Clifton Rd. NE, Atlanta, GA 30322, USA or mitomap@www.gen.emory.edu.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Emory-Georgia Tech Center for Biotechnology (SBN, MDB and DCW) and by NIH grants GM46915, NS21328, HL30164, NS30164, AG10130, DK45215 and a Muscular Dystrophy Foundation clinical research grant (DCW).

REFERENCES

- 1 Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., *et al* (1981) *Nature* **290**, 457–465.
- 2 Wallace, D.C. (1992) *Ann. Rev. Biochem.* **61**, 1175–1212.
- 3 Wallace, D.C., Lott, M.T., Brown, M.D., Huoponen, K., Torroni, A. (1995) Report of the committee on human mitochondrial DNA. In: Cuticchia AJ (ed) *Human Gene Mapping 1995: A compendium*. Johns Hopkins University Press, Baltimore, pp. 910–954 (also available at <http://www.gen.emory.edu/mitomap.html>).
- 4 Wallace, D.C. (1995) *Am. J. Hum. Genet.* **57**, 201–223.