112–115 *Nucleic Acids Research, 1998, Vol. 26, No. 1*

MITOMAP: a human mitochondrial genome database—1998 update

Andreas M. Kogelnik[1,2], Marie T. Lott[1], Michael D. Brown[1], Shamkant B. Navathe[2] and Douglas C. Wallace[1,*]

[1]Center for Molecular Medicine, Emory University School of Medicine, 1462 Clifton Road, Suite 420, Atlanta, GA 30322, USA and [2]Bioengineering Program–Database Group, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280, USA

Received October 7, 1997; Accepted October 8, 1997

ABSTRACT

We have continued to develop MITOMAP (http://www.gen.emory.edu/MITOMAP), a comprehensive database for the human mitochondrial DNA (mtDNA). 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. Over the past year we have increased the degree of interlinking of MITOMAP information available on the web page, by using our generalized information management system, GENOME. As increasingly larger regions of the human genome are sequenced and characterized, the need for integrating such information is growing. Consequently, MITOMAP and GENOME provide a valuable reference for the mitochondrial biologist, in addition to being a model for the development of comprehensive, information storage and retrieval systems for other components of the human genome. This paper documents the changes to MITOMAP which have been implemented over the past year.

INTRODUCTION

MITOMAP (1) is the most complete database of published data relating to the human mitochondrial genome. It is available through the World Wide Web and is managed using GENOME——the Georgia Tech Emory Networked Object Management Environment. In establishing MITOMAP in 1995, we discovered numerous shortcomings in implementing the database using commercially available relational and then object-oriented formats, the most serious were the inability to model complex biological data and relationships and the inability to easily modify our data models once in each system. Over the past year, MITOMAP has been converted to the object-relational format of GENOME which has eliminated or alleviated those deficiencies by providing mechanisms for complex data modeling and broad-based informational queries. GENOME supports complex information modeling and information exchange through a graphical user interface utilizing ASN.1 (Abstract Syntax Notation One) standard as its data definition language. MITOMAP data and information describing the relationships between mtDNA data can be stored in the system and GENOME allows simple modification of existing data and data structures.

MITOCHONDRIA BACKGROUND

The human mtDNA is a 16 569 nucleotide pair (np) closed, circular molecule located within the cytoplasmic mitochondria (Fig. 1), the first component of the human genome to be completely sequenced (2). 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 (4).

The maternally inherited mtDNA has a very high mutation rate (3). 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 unified entity which can be queried from a variety of different perspectives.

MITOMAP ARCHITECTURE

MITOMAP is currently implemented using GENOME and linked to the WWW. It is both a self-contained information system for the mitochondrial biologist and is linked to the Genome Database (4) of the international Human Genome Organization (HUGO) and to On-Line Mendelian Inheritance of Man (5), thus providing mtDNA map and clinical resources to these entities.

The architecture of the five overlapping data sets of MITOMAP remains unchanged (Fig. 2). The first set, the ‘standard’ mtDNA sequence (2), is utilized as the key unifying element for interrelating the remaining four elements: the functional genetic element data set, the clinical mutation data set, the population variation data set and the gene–gene interaction data set. The

* To whom correspondence should be addressed. Tel: +1 404 727 8368; Fax: +1 404 727 8367; Email: dwallace@gmm.gen.emory.edu

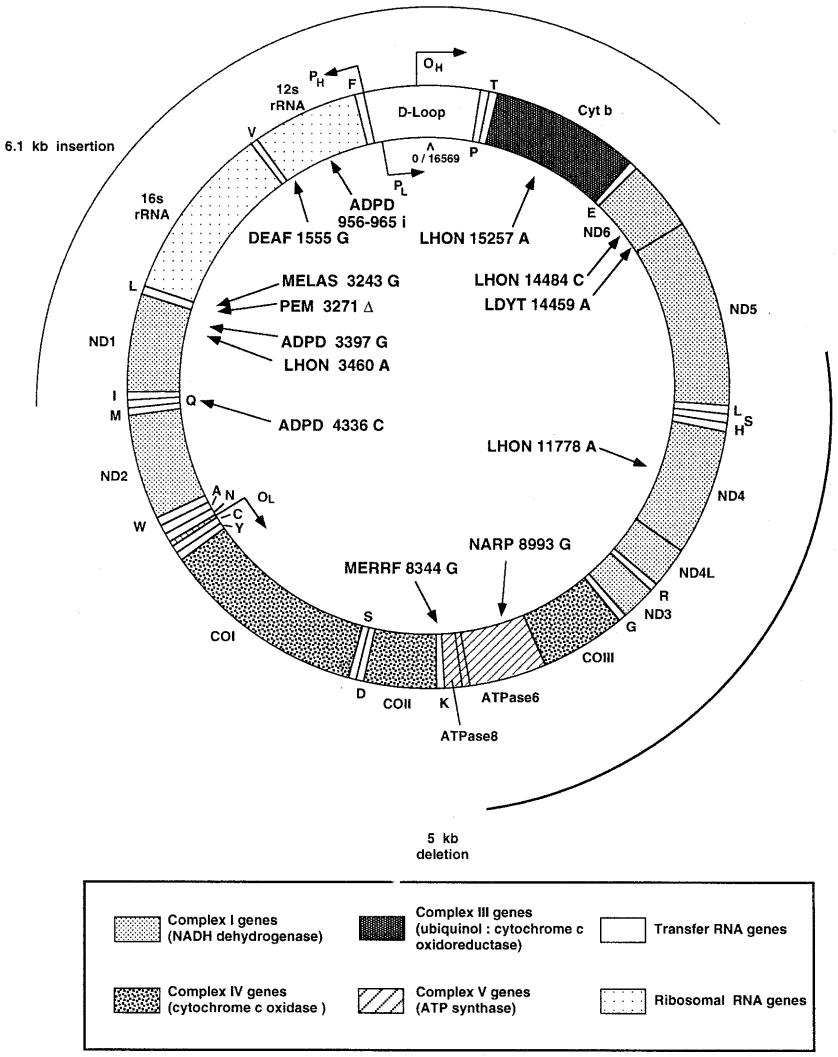

Figure 1. Human mtDNA structural and morbid maps. The tRNAs are indicated by their cognate amino acid single letter code and gene designations are defined in (3). The genes encoded by the G-rich heavy (H) strand are on the outside of the circle, while those for the C-rich light (L) strand are on the 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.

addition of several hundred aligned mtDNA sequence fragments (D-loop and others) should be completed by late 1997. Relationships from data in each set to the sequence data are established and are represented as hyperlinks on our WWW page. Portions of MITOMAP data are presented as searchable hyperlinked tables. The MITOMAP home page (http://www.gen.emory.edu/MITOMAP) provides a query form executing searches across multiple information types.

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.

Disease associated nucleotide positions and base changes are listed in the clinical mutation data set which has been extended to cover over 60 base substitutions (Table 1). In addition over 200 mtDNA rearrangements are included, documenting 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. In addition, a library of pathogenic and normal phenotypes is now available.

The population variation data set provides access to over 750 known polymorphic sites (Table 1). These include restriction site polymorphisms, small insertion–deletion variants, and identified sequence changes. The population associations of highly

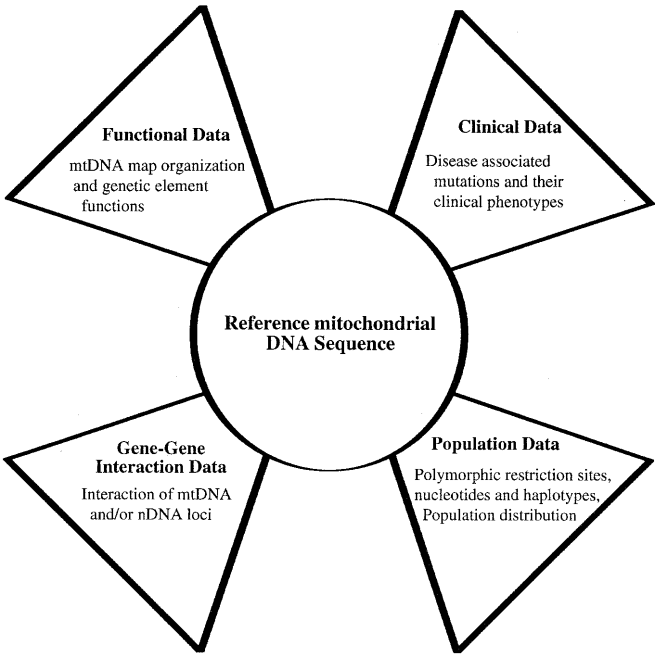

Figure 2. The primary database areas and their interrelationship with examples of the types of information which each currently documents. nDNA, nuclear DNA; mtDNA, mitochondrial DNA.

informative variants are provided through available information of mtDNA haplotypes and the continental distributions and population frequencies.

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 UTILITY

MITOMAP can be used to pose questions spanning the varied data domains outlined above. For example, the following questions regarding Leber's Hereditary Optic Neuropathy (LHON) can be addressed. What are the primary mutations associated with LHON? What secondary polymorphisms are at increased frequency in LHON? What are the continental or haplogroups associations of the LHON-associated mutations? What is the distribution of LHON mutations among the genes of the mtDNA?

MITOMAP ACCESS

The mitochondrial database is available to the general public through the WWW (http://www.gen.emory.edu/mitomap.html , Fig. 3). During the past year MITOMAP was accessed on average over 8000 times/month. The interface provides both browsing and querying capabilities. Users can browse through the database in its published flat file format (adapted from 4). In addition a query interface is provided to perform searches on specific aspects of the mitochondrial genome.

DATA ACQUISITION

Data is taken from published works on the mitochondrial genome. The committee regularly searches the literature for new publications. The database is updated as new data is obtained, WWW pages and query responses are generated on-the-fly, providing users with the most up-to-date information possible. Submissions should be send to Dr Douglas C.Wallace, Attn: Mitochondrial Genome Committee, Center for Molecular Medicine, Emory University, 1462 Clifton Road, NE Suite 420, Atlanta, GA 30322, USA or mitomap@infinity.gen.emory.edu.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Emory-Georgia Tech Center for Biotechnology (SBN, MDB and DCW), by a National Library of Medicine Fellowship (AMK), and by NIH grants GM46915, NS21328, HL30164, AG10130, AG13154, and a Muscular Dystrophy Foundation clinical research grant (D.C.W.).

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

mtDNA regions	Total no. of genes	Total no. of nucleotides	No. of mutations with disease	Total no. of polymorphisms
mRNA	13	11 388	28	393
tRNA	22	1509	32	37
rRNA	2	2513	2	64
Control region[a]	16	1121	0	390

[a]The control region spans roughly from between tRNA phenylalanine to tRNA proline.

REFERENCES

1 Kogelnik,A.M., Lott,M.T., Brown,M.D., Navathe,S.B. and Wallace,D.C. (1996) *Nucleic Acids Res.* **24**, 177–179.
2 Anderson,S., Bankier,A.T., Barrell,B.G., de Bruijn,M.H.L., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F. *et al.* (1981) *Nature* **290**, 457–465.
3 Wallace,D.C. (1992) *Annu. Rev. Biochem.* **61**, 1175–1212.
4 Fasman,K.H., Letovsky,S.I., Cottingham,R.W. and Kingsbury,D.T. (1996) *Nucleic Acids Res.* **24**, 57–63 [see also this issue (1998) *Nucleic Acids Res.* **26**, 94–99].
5 Pearson,P., Francomano,C., Foster,P., Bocchini,C., Li,P. and McKusick,V. (1994) *Nucleic Acids Res.* **22**, 3470–3473.
6 Wallace,D.C., Lott,M.T., Brown,M.D., Huoponen,K. and Torroni,A. (1995) In Cuticchia.A.J. (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).
7 Wallace,D.C. (1995) *Am. J. Hum. Genet.* **57**. 201–223.

MITOMAP v3.0

A human mitochondrial genome database
A. M. Kogelnik 1,2 , M.T. Lott 1, M.D. Brown 1, S.B. Navathe 2, D.C. Wallace 1

1. Center for Molecular Medicine, Emory University School of Medicine, Atlanta, Georgia
2. Bioengineering Program, College of Computing, Georgia Institute of Technology, Atlanta, Georgia

Search MITOMAP for information on: Function

Perform search Clear

Gene, disease, enzyme names may be abbreviated, truncated, etc.

Adapted from: The Report of the Committee on the Human Mitochondrial Genome 1995

D.C. Wallace, M. T. Lott, and M. D. Brown

- Mitochondrial DNA Functional Map
- Mitochondrial DNA Variation
- Mitochondrial DNA Mutations and Disease
- Somatic Mitochondrial DNA Mutations in Aging and Degenerative Diseases
- Figure 1: Mitochondrial DNA Function Map
- Figure 2: Eleven pathological mutations in the MTTL1 gene
- Table 1: Mitochondrial DNA Function Locations
- Table 2: Mitochondrial DNA Polypeptide Assignments
- Table 3: Polymorphic MtDNA Restriction Sites Detected by Low Resolution Screening of Human Populations
- Table 4: Polymorphic MtDNA Restriction Sites Detected by High Resolution Screening of Human Populations
- Table 5: Common Continent-Specific mtDNA Variants
- Table 6: MtDNA Control Region Sequence Polymorphisms
- Table 7: MtDNA Coding Region Sequence Polymorphisms
- Table 8: Reported Mitochondrial DNA Base Substitution Diseases: Missense mutations
- Table 9: Reported Mitochondrial DNA Base Substitution Diseases: rRNA and tRNA mutations
- Table 10: Reported mtDNA Deletions
- Table 11: Reported Multiple mtDNA Deletions Within Individuals
- Table 12: mtDNA Simple Insertions
- Table 13: mtDNA Complex Rearrangements
- References one at a time.
- All References - warning this is a rather large file!!!

MITOMAP Information

- About MITOMAP
- NAR Vol 24(1):177-9 - MITOMAP: A human mitochondrial genome database
- How to cite MITOMAP
- Diseases associated with the human mitochondrial genome
- The Human Mitochondrial Sequence - browsing tools coming soon!!
- List of mitochondrial Internet resources - coming soon!!

Go to the Molecular Medicine homepage

Figure 3. The MITOMAP homepage on the World Wide Web. http://www.gen.emory.edu/mitomap.html