MmtDB: a Metazoa mitochondrial DNA variants database

D. Calò*, A. De Pascali, D. Sasanelli, F. Tanzariello, M. Tommaseo Ponzetta[1], C. Saccone and M. Attimonelli

Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, 70125 Bari, Italy and [1]Istituto di Zoologia e Anatomia Comparata, Università di Bari 70126 Bari, Italy

Received August 5, 1996; Revised and Accepted October 8, 1996

ABSTRACT

The present paper describes the structure of MmtDB—a specialized database designed to collect Metazoa mitochondrial DNA variants. Priority in the data collection is given to the Metazoa species for which a large amount of variants is available, as it is the case for human variants. Starting from the sequences available in the Nucleotide Sequence Databases, the redundant sequences are removed and new sequences from other sources are added. Value-added information are associated to each variant sequence, e.g. analysed region, experimental method, tissue and cell lines, population data, sex, age, family code and information about the variation events (nucleotide position, involved gene, restriction site's gain or loss). Cross-references are introduced to the EMBL Data Library, as well as an internal cross-referencing among MmtDB entries according to their tissual, heteroplasmic, familiar and aplotypical correlation. MmtDB can be accessed through the World Wide Web at URL http://WWW.ba.cnr.it/~areamt08/MmtDBWWW.htm.

INTRODUCTION

Mitochondria are subcellular organella under the control of both nuclear and mitochondrial genomes. The mitochondrion is the only organelle in Metazoa that contains its own DNA (1). The idea to create a Metazoa mtDNA variants specialized database (MmtDB) originated from the awareness that a large mass of information associated to mtDNA sequences is not stored in the primary databases, which instead contain redundant information (e.g. bibliographic and taxonomic). Therefore MmtDB has been designed and implemented as a subset of the primary databases, accurately revised and enriched with specific information pertaining to the biological features of each entry, aimed to provide new data structure and to generate new cross-references between sets of data completely unlinked until now.

MmtDB is characterized as being a collection of variants and not simply a collection of Metazoa mtDNA sequences. A variant is therefore, for each species of the class Metazoa, a fragment where nucleotide differences (variations) are detected as compared with a reference sequence.

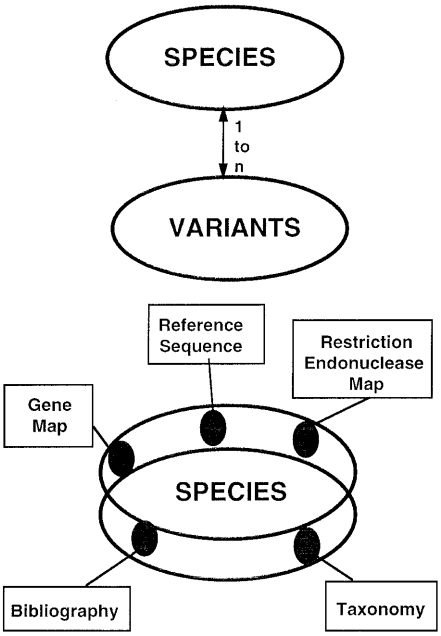


Figure 1. MmtDB structure. MmtDB is structured into two large classes: SPECIES and VARIANTS. Each of these classes is further organized into subclasses. *For each species n variants* are possible. Each variant is an entry in MmtDB database.

MmtDB DATA

MmtDB was originally designed as a Metazoan database, and our group started with the management of vertebrate and particularly human data. Yet we realized a much greater effort was needed to cover all Metazoa and some collaboration was seeked. So presently MmtDB is part of the MitBase project, a comprehensive and integrated mitochondrial database recently funded by the EU BIOTECHNOLOGY Programme. MitBase will be developed by a network of six nodes, each collecting and editing data on different groups of organisms (protists, plants, fungi, vertebrates, invertebrates and humans), by a bioinformatic node (EBI) and a node dealing with a pilot project on nuclear genes related to mitochondria. The role of our group in the project is to continue the work started on vertebrates.

*To whom correspondence should be addressed. Tel: +39 80 548 2130; Fax: +39 80 548 4467; Email: areadc13@area.ba.cnr.it

Table 1. Nucleotide variation code matrix. Each figure in the matrix indicates the numeric code associated in MmtDB to each variation event in the variant compared with the reference sequence

VARIANT

REFERENCE SEQUENCE

→	A	C	G	T	-	N	M	R	W	S	Y	K	V	H	D	B
A	**0**	**5**	**1**	**7**	**13**	**23**	33	34	35	36	37	38	39	40	41	42
C	**6**	**0**	**9**	**3**	**14**	**24**	43	44	45	46	47	48	49	50	51	52
G	**2**	**10**	**0**	**11**	**15**	**25**	53	54	55	56	57	58	59	60	61	62
T	**8**	**4**	**12**	**0**	**16**	**26**	63	64	65	66	67	68	69	70	71	72
-	**17**	**18**	**19**	**20**	**0**	**27**	73	74	75	76	77	78	79	80	81	82
N	**28**	**29**	**30**	**31**	**32**											
M	83	84	85	86	87											
R	88	89	90	91	92											
W	93	94	95	96	97											
S	98	99	100	101	102											
Y	103	104	105	106	107											
K	108	109	110	111	112											
V	113	114	115	116	117											
H	118	119	120	121	122											
D	123	124	125	126	127											
B	128	129	130	131	132											

21 = String deletion

22 = String insertion

Codes in bold indicate the variation among nucleotides A, C, G and T and the insertion (17, 18, 19 and 20) or the deletion (13, 14, 15 and 16) of a nucleotide. Codes 23–27 and 28–32 mark a variation into or from the undefined nucleotide N. Codes not in bold, 33–132, are related to the variation into uncertain and ambiguous nucleotides according to the IUB codification.

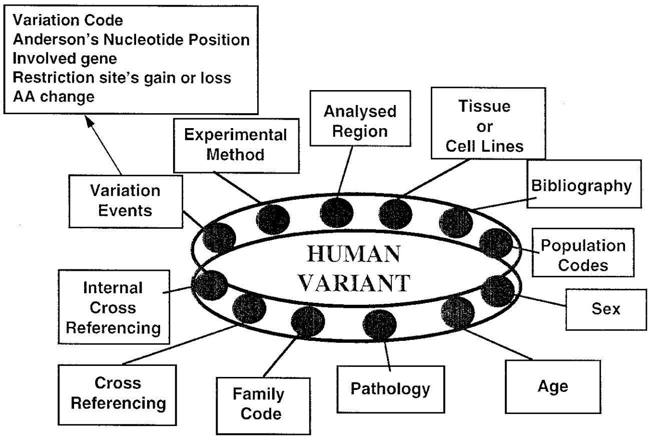


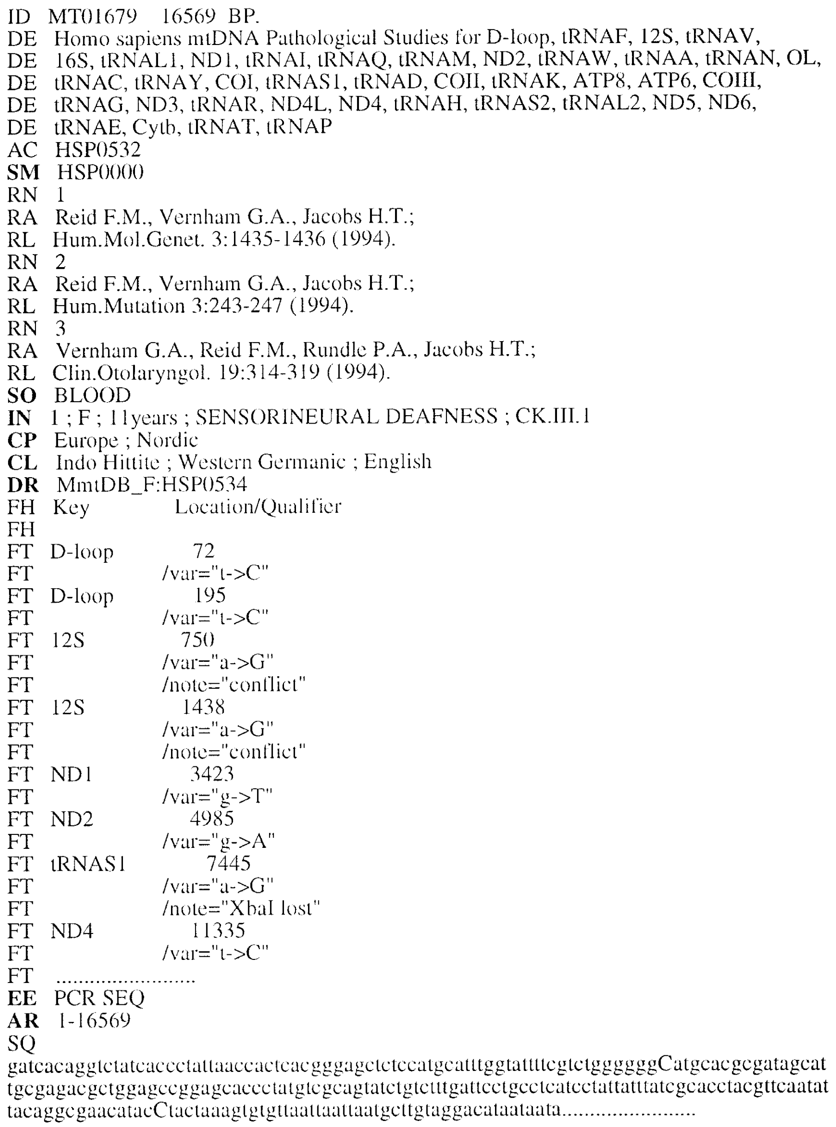
Figure 2. Each class is organized into subclasses. The subclasses associated to each human variant are shown.

The vertebrates mitochondrial genome is a closed round molecule having a size from 15 000 to 20 000 bp (2). The peculiar features of this molecule are its high compactness and simplicity; mitochondrial DNA (mtDNA) is dense with information, does not have introns and is generally composed of 13 genes coding for proteins, 22 tRNA genes and two rRNA genes. It shows a major non-coding region, called the D-loop in vertebrates, which is involved in regulatory processes. The D-loop is the most variable part of the genome and is thus used as a marker for human diversity studies (3,4).

The mtDNA has an unusual genetics. It is maternally inherited (5) and is polyploid, which means it is present, in both the cell and organelle, with a high copy number. Therefore the mtDNA in a same organelle, cell, tissue, organ or individual may not be homogeneous, indeed when a new mutation occurs it creates a mixed intracellular population of mutant and normal mtDNAs known as heteroplasmy. When a heteroplasmic cell divides, the mutant and the normal DNAs are randomly distributed in the daughter cells (mitotic segregation process) (6). At some stages in oogenesis, the amount of mtDNA molecules is reduced to a relatively small number (bottleneck hypothesis) (7). Following these stages over- replication brings the amount of molecules in each DNA cell to its normal high level and this can lead to a relatively pure population of each genome that pre-existed in the original parental organelle (8).

Moreover, the location of mtDNA, which is attached to the mitochondrial inner membrane and close to the respiratory chain, its lack of protective proteins (e.g. histones) and a very poor DNA repair system (9) make mtDNA particularly prone to mutation.

Because of its reduced size and the fact that mtDNA is used in several fields of applied biology (10), the number of sequences



```
ID   MT01679   16569 BP.
DE   Homo sapiens mtDNA Pathological Studies for D-loop, tRNAF, 12S, tRNAV,
DE   16S, tRNAL1, ND1, tRNAI, tRNAQ, tRNAM, ND2, tRNAW, tRNAA, tRNAN, OL,
DE   tRNAC, tRNAY, COI, tRNAS1, tRNAD, COII, tRNAK, ATP8, ATP6, COIII,
DE   tRNAG, ND3, tRNAR, ND4L, ND4, tRNAH, tRNAS2, tRNAL2, ND5, ND6,
DE   tRNAE, Cytb, tRNAT, tRNAP
AC   HSP0532
SM   HSP0000
RN   1
RA   Reid F.M., Vernham G.A., Jacobs H.T.;
RL   Hum.Mol.Genet. 3:1435-1436 (1994).
RN   2
RA   Reid F.M., Vernham G.A., Jacobs H.T.;
RL   Hum.Mutation 3:243-247 (1994).
RN   3
RA   Vernham G.A., Reid F.M., Rundle P.A., Jacobs H.T.;
RL   Clin.Otolaryngol. 19:314-319 (1994).
SO   BLOOD
IN   1 ; F ; 11years ; SENSORINEURAL DEAFNESS ; CK.III.1
CP   Europe ; Nordic
CL   Indo Hittite ; Western Germanic ; English
DR   MmtDB_F:HSP0534
FH   Key            Location/Qualifier
FH
FT   D-loop            72
FT                  /var="t->C"
FT   D-loop            195
FT                  /var="t->C"
FT   12S               750
FT                  /var="a->G"
FT                  /note="conflict"
FT   12S               1438
FT                  /var="a->G"
FT                  /note="conflict"
FT   ND1               3423
FT                  /var="g->T"
FT   ND2               4985
FT                  /var="g->A"
FT   tRNAS1            7445
FT                  /var="a->G"
FT                  /note="XbaI lost"
FT   ND4               11335
FT                  /var="t->C"
FT   ........................
EE   PCR SEQ
AR   1-16569
SQ
gatcacaggtctatcaccctattaaccactcacgggagctctccatgcatttggtattttcgtctggggggCatgcacgcgatagcat
tgcgagacgctggagccggagcaccctatgtcgcagtatctgtctttgattcctgcctcatcctattatttatcgcacctacgttcaatat
tacaggcgaacatacCtactaaagtgtgttaattaattaatgcttgtaggacataataata........................
```

Figure 3. Example of MmtDB entry in flatfile format. The fields in bold contain classes of information usually not stored in the flatfile of the primary databases, e.g. **SM** = reference sequence accession number. **SO** = source, i.e. tissue or cell lines from which the mtDNA was extracted. **IN** = information on the individuals, i.e. number of individuals for which the same type of source has produced the same variant with the same analysis; sex; age; pathological or normal status; family code, the first two letters identify the generation, the Roman number indicates the generation, the Arabic number marks the individual in the generation, **CP** = classification of the population to which the individuals belong, continental groups and population groups. **CL** = linguistic classification to which the individuals belong. **DR** = cross-referencing to the primary databases or/and to MmtDB. In the entry the DR line refers to a family correlation with another MmtDB entry (MmtDB_F:HSP0534) **EE** = experimental technique used (PCR SEQ = sequencing by PCR). **AR** = analysed region based on the reference sequence.

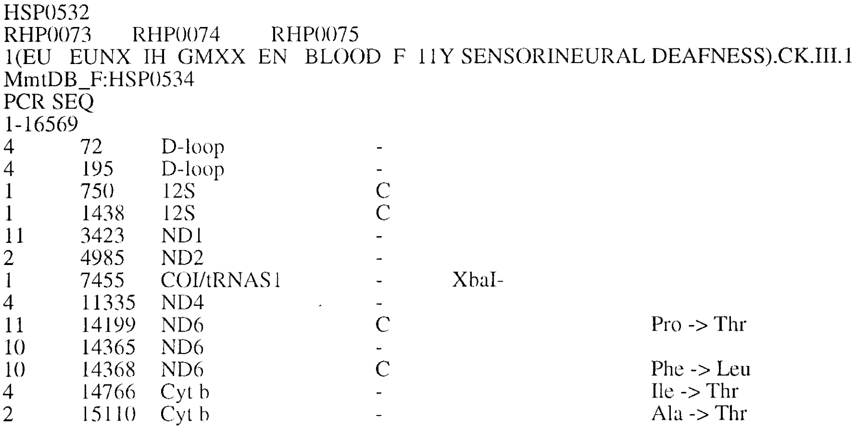
of mitochondrial genes and complete genomes is growing exponentially. In particular, much information on polymorphic regions of mtDNA is now available in literature. For human mt DNA, several of the sequences managed in MmtDB are related to evolutionary studies and are relevant to the hypervariable segments (HVI, HVII) of the D-loop (11–14), and as many others are in connection with pathology studies on alterations of the mtDNA (deletions, insertions and point mutations) (15,16).

The human data are coded using as reference the nucleotide sequence published by Anderson *et al.* in 1981 (17), which, despite being a hybrid (was derived from placenta mtDNA and in part from HeLa cell mtDNA), represents an important reference in human variability studies.

MmtDB DATA SOURCE

The Metazoan mitochondrial sequences are retrieved from the EMBL (18) and GenBank (19) primary databases.

The data are extracted from the primary databases using the GCG (20), ACNUC (21) and SRS (22) packages. The comparison between a reference sequence and each potential variant is performed by applying the GCG program BESTFIT. The published sequence data which are not included in the primary databases are extracted from bibliographic databases (Medline, Current Contents), from Entrez or other information systems. Congress acta and unpublished data kindly provided by the authors are also included.



```
HSP0532
RHP0073     RHP0074     RHP0075
1(EU  EUNX  IH  GMXX  EN  BLOOD  F  11Y SENSORINEURAL DEAFNESS).CK.III.1
MmtDB_F:HSP0534
PCR SEQ
1-16569
4      72      D-loop          -
4      195     D-loop          -
1      750     12S             C
1      1438    12S             C
11     3423    ND1             -
2      4985    ND2             -
1      7455    COI/tRNAS1      -      XbaI-
4      11335   ND4             -
11     14199   ND6             C                        Pro -> Thr
10     14365   ND6             -
10     14368   ND6             C                        Phe -> Leu
4      14766   Cyt b           -                        Ile -> Thr
2      15110   Cyt b           -                        Ala -> Thr
```

Figure 4. Compact format of the same MmtDB entry whose flatfile is shown in Figure 3. From top: HSP0532 = entry number. RHP0073, RHP0074, RHP0075 = reference number codes. In the following four lines information on the individuals, cross-referencing, experimental technique and analysed region are reported. The other lines report: first column = nucleotide variation codes; second column: nucleotide variation position; third column = involved genes or regions; fourth column = pathogenicity of the variation event (– marks a variation not associated to a pathology by the author, + marks a variation associated to a pathology by the author, C conflicting variation compared with the reference sequence); fifth column = possible enzyme name and restriction sites lost (–) or gained (+); sixth column = possible amino acid change.

MmtDB STRUCTURE

The data in MmtDB are organized into two large classes: **SPECIES** and **VARIANTS** (Fig. 1). Each of these classes is further organized into subclasses or objects (an example of subclasses associated to each human variant is shown in Fig. 2). To *each species, n variants* are associated and **each variant is an entry in the MmtDB** database. The SPECIES class refers to the items in the database which can be associated to a biological species of METAZOA and of which mtDNA variants are available.

To the class SPECIES the following objects are associated: the Reference Sequence, the Gene and Restriction Endonuclease Maps, the Taxonomic Classification and the Bibliography. The reference sequence(s) is represented by the nucleotide sequence of the complete mitochondrial genome, if the genome of that species has been fully sequenced; otherwise if one or more fragments have been sequenced, the reference is either the longest sequence of each fragment or a virtual sequence made up by the combination of overlapping fragments from different sources. Such a reference sequence can be regarded as a tool which allows compactness of information. For each species the entire sequence can thus be reported only once and for each variant it is sufficient to code only the differences in the *'pattern of the variation events'*.

The VARIANTS class includes as objects information items specific of the fragment under consideration, such as: (i) the location of the fragment in the reference sequence (*analysed region*); (ii) the *experimental method* used for the detection of the variant, e.g. Sanger, Maxam and Gilbert, RFLP, Southern or PCR; (iii) the *pattern of the variation events* with respect to the reference sequence, i.e. the nucleotide position in the reference sequence where the variation occurs, the type of variation (point mutations, deletions or insertions), according to the codes reported in Table 1, the involved gene and the loss or gain of a restriction site following the variation; (iv) *bibliographic* references; (v) the *tissue or cell lines* from which the DNA was extracted; (vi) *population data*, relevant to the geographic and linguistic origin of the individuals from which the DNA was extracted; (vii) the *age* and the *sex* of the individuals and their *pathological* or *normal status*.

Population data are coded according to geographical (Continental groups) and anthropological (Population groups) classifications. A linguistic classification according to M. Ruhlen (23) (Linguistic phylum, Language group and Language) has also been added. These classifications are often limited by the scarcity of information reported in the original papers.

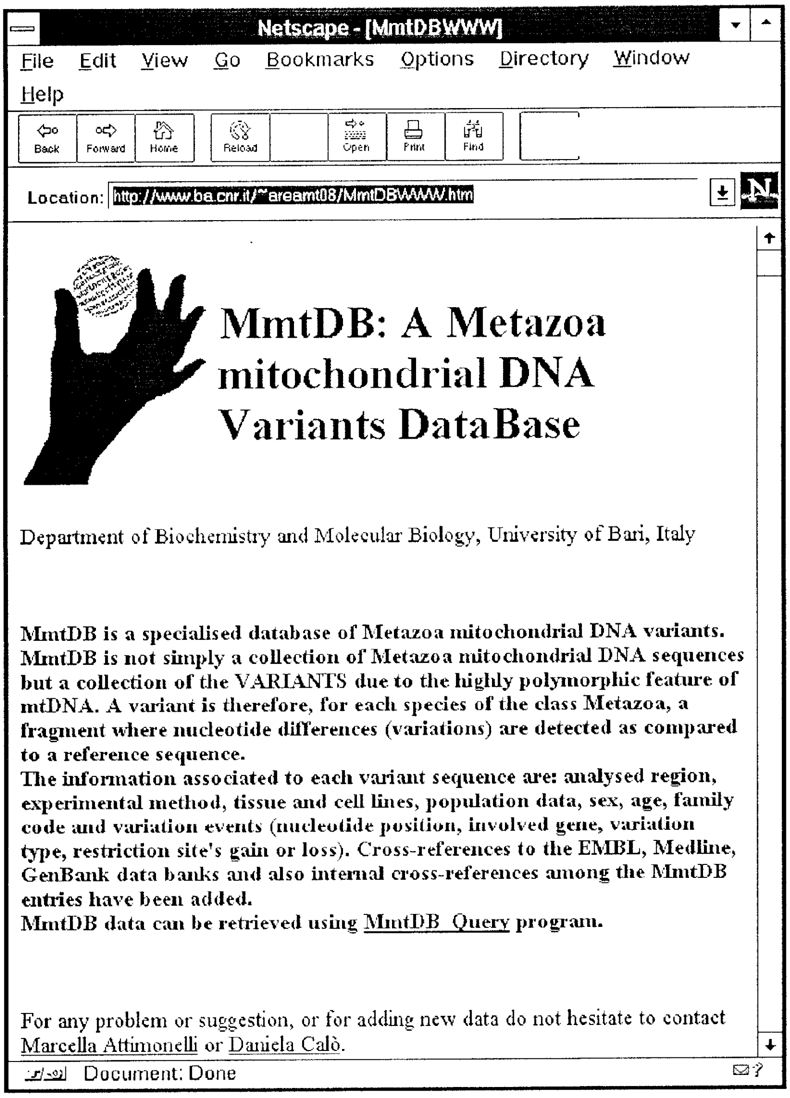
When the variant has been extracted from the primary databases it is *cross-referenced* through the Accession Number and the Entry Name which univocally identify each data-entry in the primary databases (DR field in Fig. 3). Then the entries are *internally cross-referenced* through their accession number in MmtDB (AC fields in Fig. 3) in order to link different but correlated entries. The correlation can be based on the tissue type (T), the aplotype (A), the family (F) or the heteroplasmic status (H).

The family correlation pertains mainly to human data related to mitochondrial disease studies. Indeed when complete pedigrees have been analysed a *family code* is defined in MmtDB composed of two letters for the family, followed by a roman number identifying the generation and an Arabic number identifying the individual in the generation. Figure 3 shows an example of MmtDB entry in the flatfile format, which is commonly used by most of the biological databases, as in the EMBL data library, the GenBank, the Swissprot (24) and many others.

DATA INPUT PROGRAM

Data at present are stored by annotators skilled in biology. We are planning to prepare a submission form allowing the authors to submit data directly. The data in MmtDB are stored in a compact format as shown in Figure 4 using the program MITO-INS.

MITO-INS is written in C language and interactively structured to allow the annotator to insert, modify and display the data of an entry in the compact format. It is provided with a menu for the choice of operation by the annotator. The input data are saved into a database made up of a set of files relationally structured to allow the complete and fully flexible interrogation and retrieval of data.



http://www.ba.cnr.it/~areamt08/MmtDBWWW.htm

MmtDB: A Metazoa mitochondrial DNA Variants DataBase

Department of Biochemistry and Molecular Biology, University of Bari, Italy

MmtDB is a specialised database of Metazoa mitochondrial DNA variants. MmtDB is not simply a collection of Metazoa mitochondrial DNA sequences but a collection of the VARIANTS due to the highly polymorphic feature of mtDNA. A variant is therefore, for each species of the class Metazoa, a fragment where nucleotide differences (variations) are detected as compared to a reference sequence.
The information associated to each variant sequence are: analysed region, experimental method, tissue and cell lines, population data, sex, age, family code and variation events (nucleotide position, involved gene, variation type, restriction site's gain or loss). Cross-references to the EMBL, Medline, GenBank data banks and also internal cross-references among the MmtDB entries have been added.
MmtDB data can be retrieved using MmtDB Query program.

For any problem or suggestion, or for adding new data do not hesitate to contact Marcella Attimonelli or Daniela Calò.

Figure 5. MmtDB home page. Underlined words can be clicked to navigate in MmtDB.

MmtDB DATA DISTRIBUTION AND DATABASE ACCESS

A World Wide Web site has been developed to allow an easy access to the information in the MmtDB at the following address: http://WWW.ba.cnr.it/~areamt08/MmtDBWWW.htm . MmtDB Home Page is shown in Figure 5.

MmtDBWWW is an interrogation system which allows a point-and-click interface for the selection of lists of entries on which the following tasks can be performed: **flatfile generation** for each variant entry (Fig. 3); **nucleotide sequence extraction** of variant sequences based on a reference sequence with nucleotide variations in capital letters; **analysis of variation events** (Fig. 6).

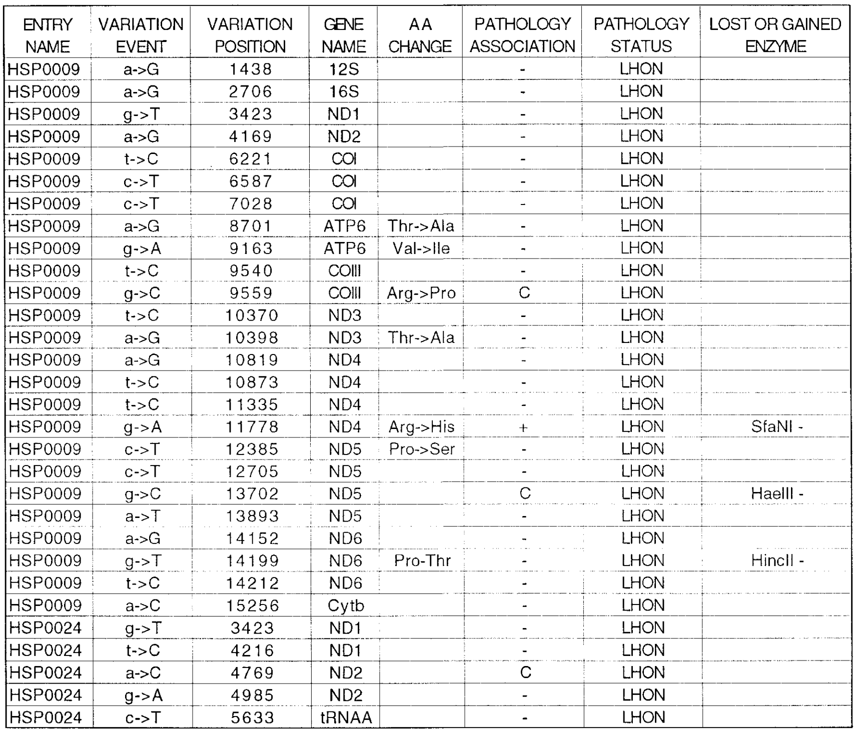
The selection can be performed using at maximum the Boolean combination of four of the following criteria: gene code, source, technique name, continental group, linguistic family, linguistic group, language, family code, sex code, age, pathology acronym, variation event code, variation position, analysed region, restriction enzyme name and all text, that is search for a specific word in the entire flatfile entry.

CONCLUSIONS AND DEVELOPMENTS

MmtDB is not the only specialized mitochondrial DNA database available to the scientific community, two other databases contain similar data, GOBASE and Mitomap (25,26).

The major purpose of GOBASE is to organize, collect and cross-reference the dispersed information concerning organelle genomes, both mitochondria and chloroplasts, with the aim to make them available to the scientific community as an organized set, free from errors.

Mitomap is a comprehensive collection of human mitochondrial data and includes essentially clinical data on pathogenic mutations. Mitomap is the one similar in content to MmtDB, but it does not offer the same flexibility in structure and potential for data correlation. Indeed, Mitomap is made up of text tables which do not allow an easy and prompt correlation between data. No cross-reference is provided in Mitomap to other databases. Furthermore, selection is possible for only one of four set fields: function, disease, polymorphisms and restriction sites. Despite the good clinical information provided in Mitomap, it does not include information on the analysed region and sequence, on the



ENTRY NAME	VARIATION EVENT	VARIATION POSITION	GENE NAME	AA CHANGE	PATHOLOGY ASSOCIATION	PATHOLOGY STATUS	LOST OR GAINED ENZYME
HSP0009	a->G	1438	12S		-	LHON	
HSP0009	a->G	2706	16S		-	LHON	
HSP0009	g->T	3423	ND1		-	LHON	
HSP0009	a->G	4169	ND2		-	LHON	
HSP0009	t->C	6221	COI		-	LHON	
HSP0009	c->T	6587	COI		-	LHON	
HSP0009	c->T	7028	COI		-	LHON	
HSP0009	a->G	8701	ATP6	Thr->Ala	-	LHON	
HSP0009	g->A	9163	ATP6	Val->Ile	-	LHON	
HSP0009	t->C	9540	COIII		-	LHON	
HSP0009	g->C	9559	COIII	Arg->Pro	C	LHON	
HSP0009	t->C	10370	ND3		-	LHON	
HSP0009	a->G	10398	ND3	Thr->Ala	-	LHON	
HSP0009	a->G	10819	ND4		-	LHON	
HSP0009	t->C	10873	ND4		-	LHON	
HSP0009	t->C	11335	ND4		-	LHON	
HSP0009	g->A	11778	ND4	Arg->His	+	LHON	SfaNI -
HSP0009	c->T	12385	ND5	Pro->Ser	-	LHON	
HSP0009	c->T	12705	ND5		-	LHON	
HSP0009	g->C	13702	ND5		C	LHON	HaeIII -
HSP0009	a->T	13893	ND5		-	LHON	
HSP0009	a->G	14152	ND6		-	LHON	
HSP0009	g->T	14199	ND6	Pro-Thr	-	LHON	HincII -
HSP0009	t->C	14212	ND6		-	LHON	
HSP0009	a->C	15256	Cytb		-	LHON	
HSP0024	g->T	3423	ND1		-	LHON	
HSP0024	t->C	4216	ND1		-	LHON	
HSP0024	a->C	4769	ND2		C	LHON	
HSP0024	g->A	4985	ND2		-	LHON	
HSP0024	c->T	5633	tRNAA		-	LHON	

Figure 6. Result of MmtDB_WWW query for the analysis of variation events. For each entry name the variation event, its position, the gene name, amino acid change, the associated pathology, pathological status and lost or gained enzyme restriction sites (see legend for Fig. 4) are reported.

used analysis technique, the source of the mtDNA, the restriction enzymes as well as on the individual.

Therefore, MmtDB is a more complete database and its structure is suitable for the flexible organization of several different information items, which can then be easily retrieved.

In MmtDB, redundancy has been minimized by comparing each new sequence against the whole set of stored sequences before it is entered as a new entry. Every effort has been made to ensure the accuracy of the data.

Database users are constantly encouraged to provide comments and possibly new data to include in the database. We believe that the contribution of the mitochondrion community, of biochemists, clinicians, geneticists and taxonomists is essential to allow the implementation and growth of the project.

ACKNOWLEDGEMENTS

We are grateful to Professor Alberto Mioni of the University of Padova for his help in population linguistic classifications. This work has been partially supported by MPI (Italy), by the ‘Comitato Biotecnologie e Biologia Molecolare’ of the CNR, Italy and by the EU-Biotechnology Programme.

REFERENCES

1 Nass,S. and Nass,M.M.K. (1983) *J. Cell. Biol.,* **19**, 593–628.
2 Wolstenholme,D.R. (1992) *Int. Rev. Cytol.*, **141**, 173–216.
3 Vigilant,L., Stoneking,M., Harpending,H., Hawkes,K. and Wilson,A. (1991) *Science*, **253**, 1503–1507.
4 Ward,R.H., Frazier,B.L., Dew-Jager,K. and Paabo,S. (1991) *Proc. Natl Acad. Sci. USA,* **88**, 8720–8724.
5 Giles,R.E., Blanc,H., Cann,H.M. and Wallace,D.C. (1980) *Proc. Natl Acad. Sci. USA,* **77**, 6715–6719.
6 Wallace,D.C. (1986) *Somatic Cell Mol. Genet.*, **12**, 41–49.
7 Haurswirth,W.W. and Laipis,P.J. (1982) *Proc. Natl Acad. Sci. USA,* **79**, 4686–4690.
8 Haurswirth,W.W. and Laipis,P.J. (1985) In Quagliariello,E. *et al.* (eds), *Achievements and Perspectives of Mitochondrial Research.* Elsevier Science Publishers B.V, Vol.II: Biogenesis, pp. 49–59.
9 Schon,E.A. (1993) In DiMauro,S. and Wallace,D.C. (eds), *Mitochondrial DNA in Human Pathology.* Raven Press, New York, pp. 1–7.
10 Saccone,C. (1994) *Curr. Opin. Genet. Dev.*, **4**, 875–881.
11 Cann,R.L., Stoneking,M. and Wilson,A.C. (1987) *Nature,* **325**, 31–36.
12 Stoneking,M., Jorde,L.B., Bhatia,K. and Wilson,A.C. (1990) *Genetics,* **124**, 717–733.
13 Torroni,A. and Wallace,D.C. (1994) *J. Bioenergetics Biomembranes,* **26**, 261–271.
14 Ayala,F.J. and Escalante,A.A. (1995), *Mol. Phylogenet. Evolution,* **5**, 188–201.
15 Wallace,D.C. (1992) *Annu. Rev. Biochem*., **61**, 1172–1212.
16 Wallace,D.C. (1992) *Science,* **256**, 628–632.
17 Anderson,S., Bankier,A.T., Barrell,B.G., Debrujin,M.H., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., *et al.* (1981) *Nature*, **290**, 457–465.
18 Rice,C.M., Fuchs,R., Higgins,D.G., Stoehr,P.J. and Cameron,G.N. (1993) *Nucleic Acids Res.,* **21**, 2967–2971.
19 Benson,D., Lipman,D.J. and Ostell,J. (1993) *Nucleic Acids Res.,* **21**, 2963–2965.
20 Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res*., **12**, 387–395.
21 Gouy,M., Gautier,C., Attimonelli,M., Lanave,C. and DiPaola,G. (1985) *CABIOS,* **1**, 167–172
22 Etzold,T. and Argos,P. (1993) *CABIOS,* **9**, 49–57.
23 Ruhlen,M. (1991) In Arnold,E. (ed.), *A Guide to the World’s Languages,* Vol. I.
24 Bairoch,A. and Boeckmann,B. *Nucleic Acids Res.,* **19**, 2247–2249.
25 GOBASE: The Organelle Genome Database Project (http://megasun.bch.umontreal.ca/gobase/gobase.html).
26 Kogelnik,A.M., Lott,M.T., Brown,M.D., Navathe,S.B. and Wallace,D. (1996) *Nucleic Acids Res.,* **24**, 177–179.