 Nucleic Acids Research, 1999, Vol. 27, No. 1 147–149

MitBASE Pilot: a database on nuclear genes involved in mitochondrial biogenesis and its regulation in *Saccharomyces cerevisiae*

B. de Pinto, S. B. Malladi[1] and N. Altamura*

Consiglio Nazionale delle Ricerche, Centro di Studio sui Mitocondri e Metabolismo Energetico, presso Università di Bari, via Amendola 165/A, I-70126 Bari, Italy and [1]EMBL Outstation, the EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Received September 3, 1998; Revised October 1, 1998; Accepted October 9, 1998

ABSTRACT

In the framework of the EU BIOTECH PROGRAM and within the 'MITBASE: a comprehensive and integrated database on mtDNA' project, we have prepared a pilot database (MitBASE Pilot) on nuclear genes involved in mitochondrial biogenesis and its regulation in *Saccharomyces cerevisiae*. MitBASE Pilot includes nuclear genes encoding mitochondrial proteins as well as nuclear genes encoding products which are localised in other sub-cellular compartments but nevertheless interact with mitochondrial functions. Genes have been classified on the basis of the mitochondrial process in which they participate and the mitochondrial phenotype of the gene knockout. The structure of the MitBASE Pilot database has been conceived for a flexible organisation of the information. An intuitive visual query system has been developed which allows users to select information in different combinations, both in the query and the output format, according to their needs. MitBASE Pilot is a relational database, is maintained at the EMBL-European Bioinformatics Institute (EBI) and is available at the World Wide Web site http://www3.ebi.ac.uk/Research/Mitbase/mitbiog.pl

INTRODUCTION

The central importance of the yeast *Saccharomyces cerevisiae* as a model organism to study the basic genetic plan in eukaryotes has been exalted since the sequence of the entire genome has been released (1). This simple eukaryote, as an aerobic facultative organism, is the organism of excellence in the study of nucleo–mitochondrial intergenomic signalling, which controls the biogenesis of mitochondria (2). This process requires a few gene products encoded by the mitochondrial genome (mtDNA) and hundreds of products encoded by nuclear genes, which are specifically assembled within the different mitochondrial compartments. Mutations in these nuclear genes can result in a respiratory deficiency (*PET* genes) (3). In recent years a great number of these *PET* genes have been identified and assigned a function. Their products play essential roles in various processes: mitochondrial protein synthesis, mitochondrial lipid or heme synthesis, TCA cycle and oxidative phosphorylation, and in mtDNA expression.

In man, some disorders characterised by the presence of mtDNA abnormalities are inherited as Mendelian traits suggesting lesions in nuclear genes likely to be involved in the replication or expression of mtDNA (4). However, identification of nuclear genes involved in the maintenance and propagation of functional mitochondria in higher eukaryotes is a hard task due to the difficulty in dealing with mutants. The yeast *S.cerevisiae* constitutes a unique source in this field and a compilation of these currently known genes is of general interest for studies on basic functions, evolutionary processes and disorders.

In the framework of the EU BIOTECH PROGRAM and within the 'MITBASE: a comprehensive and integrated database on mtDNA' project, started in 1996 (5), we have prepared a pilot database (MitBASE Pilot) on nuclear genes involved in mitochondrial biogenesis and its regulation in *S.cerevisiae*. The aim of this project is to provide the scientific community with a basic plan of the nuclear contribution to the eukaryotic mitochondrial biogenesis such as to constitute a reference model with which to compare other organisms including man.

DESCRIPTION

Structure and content of the database

MitBASE Pilot contains a non-redundant compilation of nuclear genes accurately selected on the basis of their involvement in the biogenesis of functional mitochondria and its regulation. MitBASE Pilot includes not only nuclear genes encoding mitochondrial proteins, but also those products that are localised in sub-cellular compartments other than mitochondria and nevertheless interact with mitochondrial function. The information data set has been defined to integrate and complement basic molecular and genetic data, which are essentially based on YPD (6) and MIPS (7) sources, with selected information relating to peculiar aspects of the mitochondrial research as a result of extensive screening of the literature. These genes have been

*To whom correspondence should be addressed. Tel: +39 080 544 3315; Fax: +39 080 544 3317; Email: n.altamura@area.ba.cnr.it

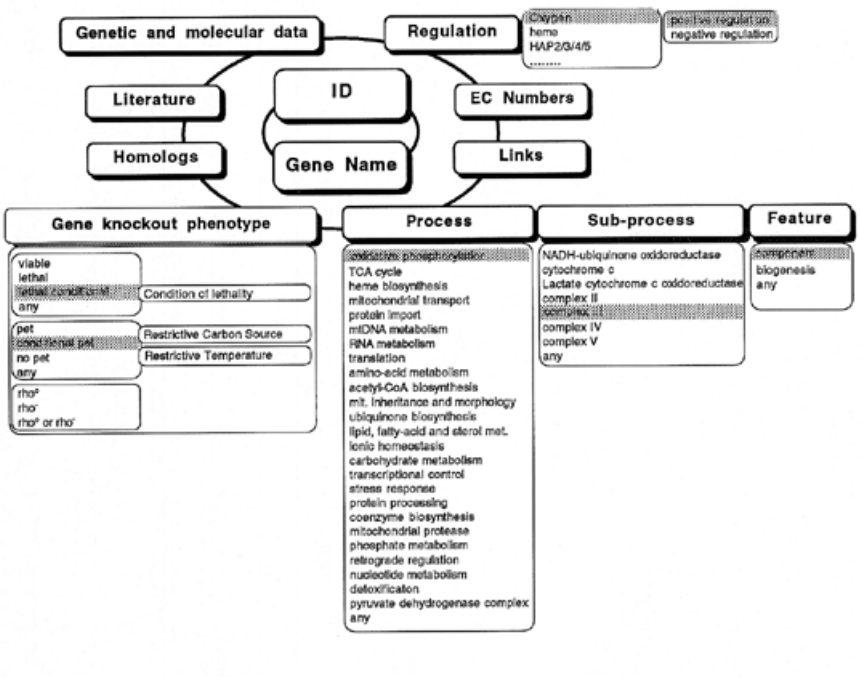

Figure 1. The conceptual model of the MitBASE Pilot database. Some of the most relevant fields, tables and relationships are reported.

classified according to the mitochondrial process in which their product participates and the mitochondrial phenotype resulting from gene knockout. The latter includes the respiratory deficiency phenotype (*PET* genes), conditional *PET* phenotypes depending on temperature and/or carbon source (*PET* cond), and the mtDNA status (rho type). All information has been structured in fields and organised in tables in a relational model (Fig. 1). Links to the primary databases EMBL/GenBank (8,9), SWISS-PROT (10), PIR (11) and the bibliographic database MEDLINE have been implemented. In the present status MitBASE Pilot contains data related to almost 500 genes.

Query system

A flexible query system has been developed to give users intuitive access to the database. It provides three classes of database retrieval operations: (i) direct retrieval by gene name/synonyms or accession numbers; (ii) retrieval through hypertext navigation in predefined menus; (iii) retrieval by multiple criteria.

They are all directly available from the MitBASE Pilot home page on the World Wide Web. Retrieval by gene name or accession number promptly displays a report containing selected and integrated information on the gene and its product. Retrieval by hypertext navigation provides specific sets of genes such as those coding for the respiratory complexes II–V subunits (succinate dehydrogenase, ubiquinol cytochrome c oxidoreductase, cytochome c oxidase and ATP synthase) with the complete set of genes whose products are components of each enzyme complex. Additional options in predefined menus allow retrieval of genes encoding products which are not components of the enzyme complexes but are specifically involved in the synthesis or assembly of a given respiratory complex. Similarly, genes encoding ribosome constituents, tRNA synthetases, initiation or elongation factors etc., can be retrieved. A distinguished feature of the MitBASE Pilot query system is retrieval by using multiple criteria. The latter includes query options by which virtually any field of the database can be accessed either independently or in different combinations with other fields. As an example the query by process ‘protein import’ can be combined with the query by knockout phenotype ‘lethal’ or the respiratory deficiency phenotype ‘*PET*’ in order to retrieve genes that are involved in the protein import system and are essential for cell viability or respiratory function.

The query by multiple criteria system also includes output options by which users can define the output format that best fits their needs. Thus, users can either retrieve information resulting from the query in a comprehensive report for each gene or select and combine solely fields of interest to be displayed alongside the list of the genes as a tabular format. The latter is indeed a very useful tool for comparative analysis since it allows overviews of the distribution of genes among chromosomes, the expression level, the knockout phenotype or the mitochondrial pathway.

DATABASE AVAILABILITY AND CITATION

MitBASE Pilot is maintained in the relational management system ORACLE at the EMBL-European Bioinformatics Institute (EBI). It has been prepared within the ‘MitBASE: a comprehensive and integrated database on mtDNA’ project and can be accessed either through the MitBASE home page at the WWW site: http://www.ebi.ac.uk/htbin/Mitbase/mitbase.pl or directly at the WWW site: http://www3.ebi.ac.uk/Research/Mitbase/mitbiog.pl

Users of the MitBASE Pilot database are requested to cite the present article in their publications.

CONCLUSION AND PERSPECTIVES

Current development and future enhancement of the MitBASE Pilot database include the implementation of a relational querying for the regulation of genes by various factors such as oxygen, heme, carbon source etc. Genes will be selectively available either for positive or negative regulation exerted by the selected factor.

Our efforts are also focused on identifying the human counterpart of yeast genes on the basis of their functional homology. Knowledge of the basic eukaryotic plan of the nuclear control of mitochondrial biogenesis would offer the scientific community an invaluable tool to help in the study of basic functions, evolutionary processes and diseases.

ACKNOWLEDGEMENTS

We would like to thank Prof. P. P. Slonimski for encouragement and helpful discussions, and G. Dujardin for her critical reading of the manuscript. We thank Prof. C. Saccone for her efforts in coordinating the MitBASE project. This work has been funded within the EU Biotech Program under the contract BIO4 CT95 0160.

REFERENCES

1 Johnston,M. (1996) *Curr. Biol.*, **6**, 500–503.
2 Grivell,L.A. (1995) *Crit. Rev. Biochem. Mol. Biol.*, **30**, 121–164.
3 Tzagoloff,A. and Dieckmann,C.L. (1990) *Microbiol Rev.*, **54**, 211–225.
4 Tiranti,V., Savoia,A., Forti,F., D'Apolito,M.F., Centra,M., Rocchi,M. and Zeviani,M. (1997) *Hum. Mol. Genet.*, **6**, 615–625.
5 Attimonelli,M., Altamura,N., Benne,R., Boyen,C., Brennicke,A., Carone,A., Cooper,J.M., D'Elia,D., De Montalvo,A., de Pinto,B. *et al.* (1999) *Nucleic Acids Res.*, **27**, 128–133.
6 Hodges,P.E., Payne,W.E. and Garrels,J.I. (1998) *Nucleic Acids Res.*, **26**, 68–72.
7 Mewes,H.W., Hani,J., Pfeiffer,F. and Frishman,D. (1998) *Nucleic Acids Res.*, **26**, 33–37.
8 Stoesser,G., Moseley,M.A., Sleep,J., McGowran,M., Garcia-Pastor,M. and Sterk,P. (1998) *Nucleic Acids Res.*, **26**, 8–15.
9 Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J. and Ouellette,B.F. (1998) *Nucleic Acids Res.*, **26**, 1–7.
10 Bairoch,A. and Apweiler,R. (1998) *Nucleic Acids Res.*, **26**, 38–42.
11 Barker,W.C., Garavelli,J.S., Haft,D.H., Hunt,L.T., Marzec,C.R., Orcutt,B.C., Srinivasarao,G.Y., Yeh,L.S.L., Ledley,R.S., Mewes,H.W., Pfeiffer,F. and Tsugita,A. (1998) *Nucleic Acids Res.*, **26**, 27–32.