148–152 Nucleic Acids Research, 2000, Vol. 28, No. 1 © 2000 Oxford University Press

MitBASE : a comprehensive and integrated mitochondrial DNA database. The present status

M. Attimonelli*, N. Altamura[1], R. Benne[2], A. Brennicke[3], J. M. Cooper[4], D. D'Elia[5], A. de Montalvo[6], B. de Pinto[1], M. De Robertis, P. Golik[7], V. Knoop[3], C. Lanave[1], J. Lazowska[7], F. Licciulli[5], B. S. Malladi[8], F. Memeo, M. Monnerot[7], R. Pasimeni, S. Pilbout[7], A. H. V. Schapira[9], P. Sloof[2] and C. Saccone

Dipartimento di Biochimica e Biologia Molecolare, Università degli Studi di Bari, 70126 Bari, Italy, [1]Centro di Studio sui Mitocondri e Metabolismo Energetico, CNR, 70126 Bari, Italy, [2]Vakgroep Biochemie, Academic Medical Center, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands, [3]Ulm University, D-89069 Ulm, Germany, [4]Department of Clinical Neurosciences, Royal Free Hospital, London, UK, [5]Area di Ricerca, CNR, 70126 Bari, Italy, [6]Departamento de Biología Molecular, Hospital Universitario 'Marqués de Valdecilla', Universidad de Cantabria, 39011 Santander, Spain, [7]Centre de Genetique Moleculaire, CNRS, 91198 Gif-sur-Yvette, France, [8]EBI Hinxton Hall, Hinxton, Cambridge, UK and [9]Department of Clinical Neurology, Institute of Neurology, London, UK

Received October 4, 1999; Accepted October 6, 1999

ABSTRACT

MitBASE is an integrated and comprehensive database of mitochondrial DNA data which collects, under a single interface, databases for Plant, Vertebrate, Invertebrate, Human, Protist and Fungal mtDNA and a Pilot database on nuclear genes involved in mitochondrial biogenesis in *Saccharomyces cerevisiae*. MitBASE reports all available information from different organisms and from intraspecies variants and mutants. Data have been drawn from the primary databases and from the literature; value adding information has been structured, e.g., editing information on protist mtDNA genomes, pathological information for human mtDNA variants, etc. The different databases, some of which are structured using commercial packages (Microsoft Access, File Maker Pro) while others use a flat-file format, have been integrated under ORACLE. Ad hoc retrieval systems have been devised for some of the above listed databases keeping into account their peculiarities. The database is resident at the EBI and is available at the following site: http://www3.ebi.ac.uk/Research/Mitbase/mitbase.pl . The impact of this project is intended for both basic and applied research. The study of mitochondrial genetic diseases and mitochondrial DNA intraspecies diversity are key topics in several biotechnological fields. The database has been funded within the EU Biotechnology programme.

INTRODUCTION

Mitochondrial DNA (mtDNA) is an essential component of all eukaryotic cells. It ensures consistency of function (cellular respiration and oxidative phosphorylation) despite the great diversity of genome organisation (1). The advent of sequencing technologies and their recent improvements has resulted in the production of mtDNA sequences for a large number of species and variants. Hence a lot of information is available, but, as in a puzzle, the dispersed pieces need to be assembled.

The vast differences in mitochondrial genome organisation and mode of gene expression observed between the taxonomic groups have made this assembling task very difficult. Ever more problematic is the storage of all data in such a way that it can be retrieved and analysed.

To alleviate some of these problems, a project was initiated as a collaborative effort of several European groups (henceforth the nodes) aiming at the construction of a Comprehensive and Integrated Database, collecting information on mtDNA from all organisms containing mitochondria in their cells resulting in the database MitBASE (2).

The present paper describes the current status of the MitBASE project.

MitBASE NETWORK AND STRUCTURE

MitBASE is defined as an integrated database because it is a collection of specialised sub-databases, one per node, differentiated by the specificity of the data each node has to manage, the guiding thread being the mtDNA.

One of the special features of MitBASE is the presence in it of 'variants', i.e., any fragment where nucleotide differences have been detected as compared with a reference sequence which can be associated to a real sample or a synthetic fragment resulting from the consensus between a set of sequences. The presence of variants in MitBASE distinguishes it from GOBASE (3), the Organelle Genomes Database. Other salient features of MitBASE are the storage of data related with the editing process occurring in plants and protists species, the

*To whom correspondence should be addressed. Tel: +39 080 548 2130; Fax: +39 080 548 4467; Email: marcella@area.ba.cnr.it

annotation of Fungi Mutants, the annotation of Nuclear Genes involved in mitochondrial biogenesis, the definition of a standard gene names classification based on GOBASE gene names with some modifications and on the KEYnet structure (4). MitBASE gene names classification includes fungal intronic open reading frames names as they have been assigned by the fungal node team. The whole database is centrally structured under ORACLE at the EBI Oracle server. For each subdatabase an ORACLE database structure has been designed and implemented. Moreover each sub-database is locally managed by the annotators supported by commercial Database Management Systems (Microsoft Access for the Human and Vertebrates Nodes, and for the Fungal Mutant data; FileMaker Pro for the Pilot Node) or through simple editor systems working on the flat-file (ff) data extracted from primary databases. The Integrated MitBASE database can be queried at the EBI WWW site through different approaches or can be released in a ff format suitable for management with any other biological database query system [i.e., SRS (5)].

DATA SOURCES

Data sources in MitBASE are the primary databases [EMBL Data Library (6) and GenBank (7)], literature through bibliographic databases (MEDLINE, Current Contents and Current Advances in Genetics and Molecular Biology) and personal communications. Data extracted from the above data sources are accurately revised, thus assuring the quality of the data stored in MitBASE. Revised data are enriched with value added information derived from other sources.

THE MitBASE HOME PAGE

The integrated database and any other information related to it, and more generally to mitochondrial DNA, is available through the MitBASE home page http://www3.ebi.ac.uk/Research/Mitbase/mitbase.pl developed at the EBI. Besides a brief description of the project and its goals, a list of functions has been implemented on the home page as follows:

- Collaborators and publications http://www3.ebi.ac.uk/Research/Mitbase/coll.html
- Submission of data http://www3.ebi.ac.uk/Research/Mitbase/mit_subm.pl
- Browser of mitochondrial keywords classification http://tonic.ebi.ac.uk:8889/mitbase/plsql/mitbase_sqs.gene_tree?gename=MITOCHONDRIAL_GENOME
- Mitochondrial translation tables http://www3.ebi.ac.uk/Research/Mitbase/trans_tabs.html
- Simple query systems http://tonic.ebi.ac.uk:8889/mitbase/plsql/mitbase_sqs.start_screen
- Elaborated query systems http://tonic.ebi.ac.uk:8889/mitbase/plsql/mitbase_eqs.get_pass_word
- Multiply aligned data http://bigarea.area.ba.cnr.it:8000/BioWWW/#AMMTDB
- Pointers to mitochondrion related information on the WWW http://www3.ebi.ac.uk/Research/Mitbase/mitrel.html
- Complete genomes in MitBASE http://www3.ebi.ac.uk/Research/Mitbase/flavio/fulgnom.html
- Entries' status of MitBASE http://www3.ebi.ac.uk/Research/Mitbase/dbstat.html

MitBASE NODES DESCRIPTION

A description of the traits of each node database is reported below.

THE HUMAN MitBASE DATABASE

A huge quantity of papers have been published so far on human mtDNA variability and its association with disease and diversity studies. On account of the great interest that the scientific community is demonstrating towards the availability in public databases of single nucleotide polymorphisms (SNP) (8) in order to study association between genotype and phenotype, a comprehensive and well structured database of the human mtDNA covering SNP can be useful. The primary focus of the Human MitBASE database (9) is to collect in a structured set all the data available worldwide.

Data in Human MitBASE are structured in relational tables.

Each entry in the Human MitBASE database is related to a variant, which is defined as a mtDNA fragment extracted from a tissue of an individual with a different pattern of variation events with respect to the reference sequence, which is the complete human mtDNA sequence published by Anderson *et al*. in 1981 (10). The information annotated in the molecular human MitBASE tables include: analysed mtDNA region, experimental method used for the analysis, tissue or cell lines used for the molecular studies, sex, age and population data of the subject and information about his/her geographical and linguistic origin. Information about the type of variation occurring (substitution, deletion, insertion), the variation location, possible restriction site gain or loss are also reported. In the clinical tables the following sub-groups have been defined: clinical, histopathological, analyte and biochemistry features (http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/hum_molecular.html and http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/hum_clinic. html)

A ff format has been fully designed and implemented as reported at http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/human_ff.html

An Elaborated Query System for human data is under definition in order to allow the end-users to fully take advantage of the whole range of added information present in the Human MitBASE database. At present, data can be retrieved through the MitBASE Simple Query system and SRS. Examples of retrieval of the Human MitBASE data through the MitBASE Simple Query System (http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/example1) and through the SRS server under the Mutation section both at the EBI and at the CNR Research Area (http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/example2) are reported on the Web.

Human MitBASE database reports at present 4319 variants associated to 14 153 subjects derived from 254 papers on disease studies and 2984 polymorphic variants associated to 4367 subjects derived from 2917 EMBL Data Library entries related to D-loop studies. Data in Human MitBASE reports at present 663 SNP (http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/polymorphisms.html).

THE FUNGAL MitBASE DATABASE

The Fungal MitBASE database collects fungal wild-type mitochondrial sequences including those of ascomycetes and filamentous ascomycetes and fungal mitochondrial mutations/variants. This database allows for the first time to explore the structure–function relationship of the mitochondrial gene sequence variants. Both classes of data are linked in the final database and their biologically relevant relations can be explored by the end-user. Data collected in Fungal MitBASE report, in addition to the standard information found in primary databases, new information concerning mainly the complex structure of some fungal genes. A unified format for annotating the split (intron-containing) genes, including a novel standardized nomenclature of intronic open reading frames (ORFs), has been implemented. The information concerning biologically relevant features of mitochondrial genes (introns, ORFs, overlapping genes, etc.) has been structured according to a standardized dataset and added to the existing data. Primary databases (EMBL Data Library, GenBank) have been the main data sources for the sequence section of Fungal MitBASE, some entries have also been created *de novo* based on personal communications. Entries from primary databases have been restructured to fit the standardized dataset and extensively validated. In this process many redundant and erroneous entries have been removed or corrected. Information from publications and personal communication, absent in the primary database entries, have also been added. The resulting database contains highly structured and uniformed data comprising 944 non-redundant entries, including six complete genomes, searchable using the MitBASE Simple Query System. The output is in ff format that will soon be implemented in the SRS system.

The Fungal Variant Database

Data contained in this section refer mostly to the 20 years of studies on the mitochondrial genetics of yeast *Saccharomyces cerevisiae*. Yeast is the main model organism of mitochondrial genetics and several hundred mutants have been found and characterized genetically. From the literature, data on 348 variants related to 465 single variation events have been collected. This constitutes a unique resource for analyzing the structure–function relationship of mitochondrial genes, since in yeast, differently from most mammals, the phenotypic effect of a given sequence change can be determined experimentally. In addition, it is now possible to introduce site-directed mutations into yeast mitochondrial genes by ballistic transformation. The wealth of information concerning yeast mitochondrial genetics has so far been scattered throughout many papers, some of the data even being unpublished. We have, for the first time, entered this data into a biological database. A structure for the Fungal Variants Database has been defined and implemented in ORACLE at the EBI. The ff output has been defined and will soon be implemented in the SRS system. The database contains entries on polymorphisms (not related to a phenotypic effect) and mutations having an associated phenotype. Thirty-four different phenotypes have been found, including respiratory deficiency, resistance, splicing, regulatory defects, etc.

The variant entries are linked to the relevant sequence entries, enabling the user to navigate through all the information concerning a particular sequence region. A specialized query system (http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/Fungi) for variant data has also been constructed. It allows the user to search the database by gene and subgenic region, sequence class (exon, intron, promoter, etc.) and phenotype thus providing the means of asking biologically relevant questions that would take enormous effort to answer using only primary literature data. The variant database can therefore be considered as one of the key achievements of the Fungal node.

THE MitBASE Pilot DATABASE

MitBASE Pilot is a yeast specialised database related to nuclear genes involved in mitochondrial biogenesis and its regulation (11). MitBASE Pilot is a relational database and is available at the WWW site: http://www3.ebi.ac.uk/Research/Mitbase/mitbiog.pl

The aim of this database is to provide the scientific community with a basic plan of the nuclear control of mitochondrial biogenesis such as to constitute a reference model with which to compare other organisms, including human. MitBASE Pilot contains nuclear genes encoding mitochondrial proteins as well as genes encoding products, which are localised in other sub-cellular compartments but nevertheless interact with mitochondrial functions. MitBASE Pilot includes selected information on the mitochondrial phenotype resulting from gene deletion studies. The database can be accessed through an intuitive visual query system providing hypertext navigation in predefined menus for retrieval of specific sets of genes. A more elaborate query system, using multiple criteria, allows retrieval of information related to single fields independently or in different combination with any other field in the database. This query system provides additional options to select and combine fields of interest to be displayed in a tabular format alongside the list of the genes resulting from the query (http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/Pilot). These unique features of MitBASE Pilot allow overviews of the distribution of genes among chromosomes, the expression level, the knockout mitochondrial phenotype or the mitochondrial pathway and constitute a useful tool for comparative analysis. MitBASE Pilot has been updated and further improved in both the structure and query options. Links to other yeast databases are now available (12–14). In line with the conceptual model and the pilot nature of this database, information related to the regulation of nuclear genes that interact with mitochondrial functions have been included. A regulation table has been implemented which collects nuclear genes whose expression is regulated by various factors (oxygen, heme, carbon source, transcriptional activators, etc.). Fields have been defined to contain information on both the regulatory factors and regulation type. A specific query system has been developed to retrieve genes regulated by a given factor. Data on regulation consolidate and integrate the dataset originally defined for this database. Thus the basic genetic plan of the nuclear control of the mitochondrial biogenesis in the simplest eukaryote begins to emerge.

THE PLANT AND ALGAE MitBASE DATABASE

Typical features of plant mitochondrial sequences include the presence of RNA editing changing the information of the DNA, the presence of introns of the group I or group II classes, the disruption of gene continuity by trans-splicing introns of

the latter class, promiscuous sequences originating from the chloroplast and nuclear DNA, and frequent recombination events leading to alternative or co-existing gene arrangements occasionally associated with phenotypes such as cytoplasmic male sterility (CMS). The plant node of the MitBASE project has aimed at a standardised annotation of these features to allow their querying in the database via a WWW interface.

Particular features of the plant mitochondrial sequence entry are described at http://www.biologie.uni-ulm.de/bio2/knoop/mitbase/plant_mt_gene.gif . Differently from animal mitochondrial genomes which are available as complete sequences in dozens of cases, sequences of large plant mitochondrial genomes are accumulating in bits and pieces. Only two complete plant mitochondrial genomes are available to date (15,16). The Plant query system has been designed to allow complex questions to be formulated that extend the possibilities to ask for the presence of certain features and their combination by boolean logic. The added information of intron classes, promiscuous DNA sequences and RNA editing sites is fully accessible for querying in various combinations. Alternative to the presentation of parallel query forms as implemented in GOBASE (3), the database has an integrated approach on a single query page; this allows the combination of several features and the narrowing of the search depending on sequence parameters given simultaneously. Sequence features can be searched with restriction to size and sequence motifs within or bordering them. Exons and introns can be searched at the same time and all queried regions can be used for output or scanned for the presence of any type of RNA editing sites and with further restrictions on the latter. Editing sites are highlighted in the sequence output. In the search output identified entries can be individually selected for further display by checkboxes. The user can choose whether entire identified entries or only the searched-for features should be displayed. The latter option is of particular use for piping the information into alignment programs, e.g., as a prerequisite for phylogenetic analyses. In parallel to the specifically designed query, the entries will also be searchable after ff import via the well established SRS. The prototypical WWW search interface is so far provisionally accessible at http://tonic.ebi.ac.uk:8889/mitbase/plsql/pla_qry.pla_show_qry_opts/

THE PROTIST MitBASE DATABASE

The protist group is not phylogenetically equal to that of plants, animals and fungi because of the enormous cytological, organisational and molecular diversity of these organisms. This diversity is also reflected in protist mitochondrial DNAs, which can be either linear or circular, sometimes, as in kinetoplastids, even consisting of a network of concatenated circles, and which are extremely heterogeneous in gene content and size. Data have been derived from primary databases by screening with every imaginable keyword, such as names of universal and species-specific mitochondrial genes, generic names of protist that contain mitochondria, and keywords derived from the specific features of the mitochondrial genetic system of individual protists or from RNA editing terminology. In total, this has resulted in the incorporation of 451 separate entries in MitBASE as listed in http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/protist_table.html . In a small but significant number of cases (~5%), errors in existing database entries were remarked, and after consultation of the authors and/or GenBank, corrections were introduced, both in the existing databases and in MitBASE entries. An additional complication is the occurrence of different forms of RNA editing. Gene-encoded sequences can be altered by insertion/deletion of Us (in kinetoplastids), insertion of C, U and various dinucleotides, substitution of C by U (in *Physarum*), and by various other nucleotide substitutions (in *Acanthamoeba*). This means that the genomic sequences as such are cryptic and cannot be used to infer the corresponding mRNA (and protein!), rRNA or tRNA sequences. In addition, in kinetoplastids the editing for the U-insertion/deletion patterns of the mRNAs is provided by small guide (g)RNAs, encoded by gRNA genes which in many cases reside in mtDNA molecules ('minicircles') different from the ones encoding the corresponding pre-edited RNAs ('maxicircles'). None of the existing large databases takes this into account, thus hampering access to edited RNA sequences. The protist MitBASE data have been organised with the specific purpose to interconnect cryptogenes with corresponding edited RNAs, gRNAs and gRNA genes. In entries containing cryptic genes ('cryptogenes'), the file containing the corresponding edited RNA sequence is listed and clickable. In the edited RNA file, the RNA coordinates and the sequence of the known corresponding gRNAs have been annotated, and (if known) the files that contain the gRNA genes are listed and clickable. Last but not least, alignments of DNA, edited RNA and gRNA sequences are provided by clicking on 'cryptogene' in the cryptic gene file. This type of interconnection has been provided for all of the 77 cryptic genes, the corresponding edited RNAs and the 234 gRNAs and 65 gRNA genes included in the protist section of MitBASE. The implemented organisation of data on edited protist RNAs is unique among databases. It not only complements the large existing general databases (GenBank, EMBL Data Library, GOBASE) but it also combines valuable features of smaller, more specialized databases, such as the U-insertion/deletion database (17), which contains gRNA:mRNA alignments for five kinetoplastid species, including some (but not all) alignments with the corresponding cryptogenes; the Guide RNA database (18), which compiles all available gRNA sequences; the minicircle database (http://www.ebi.ac.uk/parasites/parasite-genome.html), which lists all known minicircle sequences.

THE VERTEBRATE DATABASE

A great interest has recently emerged in studies on inter- and intra-species diversity based on mtDNA. In particular many studies have been published on fish mtDNA following the economic interest of stock identification in fisheries, and the related sequence data have been submitted to the primary databases. mtDNA vertebrate data available in the primary databases amount to 19 029 entries (May 1999) of which ~40% are composed of variant sequences and the remaining are non-variant data. Therefore, two Microsoft Access databases have been defined: one each for the non-variant and variant sequences.

Non-variant data, duly purified from redundancies, have been retrieved and extracted in lists according to different taxonomic classes: Mammals, Fish, Aves, Amphibia and Reptilia. These lists are available at http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/vertebrate.html ; organism names, gene names, sequence length and cross-reference to the EMBL

Data Library are reported. About 10% of these data have been fully re-annotated in MitBASE according to a dataset including a general block reporting identifiers and taxonomic data, a citations block, an experimental method block where data on experiments performed to obtain the sequences are stored, individual block and features block describing the function of any fragment. Data related to the general block and to the citation block, beside the nucleotide sequence itself are submitted to the EBI ORACLE database through the web submission interface available at the EBI MitBASE site. Data derived from the 'local Microsoft Access not variant database' are submitted in tabular format to the EBI ORACLE database.

The 38 complete genomes available among the vertebrate data are listed under the Complete Genome option of the EBI MitBASE site. The table reported at http://bio-www.ba.cnr.it:8000/Tutorials/MitBASE/vertebrate_table.html summarises the content of the Vertebrate MitBASE database. Vertebrate data already in MitBASE can be queried through the Simple Query System available at the EBI MitBASE site.

THE INVERTEBRATE MitBASE DATABASE

The Invertebrate database shares its aims with the Vertebrate database, i.e. check the accuracy of information already present in the EMBL Data Library and GenBank databases and add information (mainly: geographic origin, details on organisms, methods used for establishing sequences). When this information is not available from the literature, the authors are directly asked to provide it. Special attention is given to gene recognition and location frequently liable to error in the primary databases. Details are added for tRNA and rRNA gene description to facilitate subsequent alignments. Data relevant to polymorphisms are also enclosed as variants. All the entries related to variants of a given species are presented individually. Mention of the name of the sequence chosen as the master one is added in each one (the reverse is true for the reference sequence). This should allow any researcher to get information as a whole or on a set chosen according to her/his own requirements. At present the Invertebrate database reports 5732 entries updated to January 1999. It is worth noting that the numerous errors or inconsistencies for invertebrates, present in primary databases, have been amended in MitBASE. A query on invertebrate mtDNA submitted to primary databases will lead to false answers in ~20% of the cases. This is to stress the great added value of the MitBASE project. The development of an ad hoc query system for invertebrate MitBASE data or their implementation in SRS system is absolutely required.

ACKNOWLEDGEMENTS

This work has been partially supported by the EU-Biotechnology Programme (Contract n. BIO4-CT95-0160), by MPI (Italy), by 'Programma Biotecnologie legge 95/95 (MURST 5%)' and by CNR Research Area of Bari (Italy).

REFERENCES

1. Saccone,C. (1994) *Curr. Opin. Gen. Dev.*, **4**, 875–881.
2. Attimonelli,M., Altamura,N., Benne,R., Boyen,C., Brennicke,A., Carone,A., Cooper,J.M., D'Elia,D., De Montalvo,A., de Pinto,B., De Robertis,M., Golik,P., Grienenberger,J.M., Knoop,V., Lanave,C., Lazowska,J., Lemagnen,A., Malladi,B.S., Memeo,F., Monnerot,M., Pilbout,S., Schapira,A.H.V., Sloof,P., Slonimski,P., Stevens,K. and Saccone C. (1999) *Nucleic Acids Res.*, **27**, 128–133.
3. Gray,M.W., Lang,B.F., Cedergren,R., Golding,G.B., Lemieux,C., Sankoff,D., Turmel,M., Brossard,N., Delage,E., Littlejohn,T.G., Plante,I., Rioux,P., Saint-Louis,D., Zhu,Y. and Burger,G. (1998) *Nucleic Acids Res.*, **26**, 865–878.
4. Licciulli,F., Catalano,D., D'Elia,D., Lorusso,V. and Attimonelli M. (1999) *Nucleic Acids Res.*, **27**, 365–367. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 372–373.
5. Etzold,T., Ulyanov,A. and Argos,P. (1996) *Methods Enzymol.*, **266**, 114–128.
6. Stoesser,G., Tuli,M.A., Lopez,R. and Sterk,P. (1999) *Nucleic Acids Res.*, **27**, 18–24. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 19–23.
7. Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J., Ouellette,B.F.F., Rapp,B.A. and Wheeler,D.L. (1999) *Nucleic Acids Res.*, **27**, 12–17. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 15–18.
8. Masood,E. (1999) *Nature*, **398**, 545–546.
9. Attimonelli,M., Cooper,J.M., D'Elia,D., de Montalvo,A., De Robertis,M., Lehväslaiho,H., Malladi,S.B., Memeo,F., Stevens,K., Schapira,A.H.V. and Saccone,C. (1999) *Nucleic Acids Res.*, **27**, 143–146.
10. 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) *Nature*, **290**, 457–465.
11. de Pinto,B., Malladi,S.B. and Altamura N. (1999) *Nucleic Acids Res.*, **27**, 147–149.
12. Mewes,H.W., Heumann,K., Kaps,A., Mayer,K., Pfeiffer,F., Stocker,S. and Frishman,D. (1999) *Nucleic Acids Res.*, **27**, 44–48. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 37–40.
13. Hodges,P.E., McKee,A.H.Z., Davis,B.P., Payne,W.E. and Garrels,J.I. (1999) *Nucleic Acids Res.*, **27**, 69–73. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 73–76.
14. Chervitz,S.A., Hester,E.T., Ball,C.A., Dolinski,K., Dwight,S.S., Harris,M.A., Juvik,G., Malekian,A., Roberts,S., Roe,T., Scafe,C., Schroeder,M., Sherlock,G., Weng,S., Zhu,Y., Cherry,J.M. and Botstein,D. (1999) *Nucleic Acids Res.*, **27**, 74–78. Updated article in this issue: *Nucleic Acids Res.* (2000), **28**, 77–80.
15. Oda,K., Yamato,K., Ohta,E., Nakamura,Y., Takemura,M., Nozato,N., Akashi,K., Kanegae,T., Ogura,Y., Kohchi,T. *et al.* (1992), *J. Mol. Biol.*, **223**, 1–7.
16. Unseld,M., Marienfeld,J.R., Brandt,P. and Brennicke,A. (1997) *Nature Genet.*, **15**, 57–61.
17. Simpson,L., Wang,S.H., Thiemann,O.H., Alfonzo,J.D., Maslov,D.A., Avila,H.A. (1998) *Nucleic Acids Res.*, **26**, 170–176.
18. Hinz,S. and Göringer,H.U. (1999) *Nucleic Acids Res.*, **27**, 168.