

MitoRes: a bio-sequences resource of nuclearly encoded mitochondrial genes and products in metazoa

Catalano D., Licciulli F., Turi A., Grillo G., D'Elia D.

Istituto di Tecnologie Biomediche - Sezione di Bari, CNR, Via Amendola 122/D, 70126 Bari - Italy

Motivation

The incredible explosion of “knowledge production” in Biology in the past two decades has created a critical need for bioinformatic instruments able to manage data and facilitate their retrieval and analysis. Molecular sequences and biological data on nuclear mitochondrial genes and their products are publicly available from a wide variety of mitochondrial specialized databases. Some are species-specific, mainly human dedicated, others contain only few species and for most of them the only sequences data reported concern proteins. We have developed MitoRes database to collect and integrate information on nuclearly encoded proteins and genes targeting the mitochondrion for all metazoan species and to provide a flexible and efficient tool for the export of bio-sequences in support of researchers interested in functional characterization of gene, transcript and amino acid sequences related to biogenesis, metabolism and pathological dysfunctions of mitochondria.

Methods

MitoRes is an interconnected knowledge management system based on a MySQL relational database, ensuring data consistency and integrity, and on a Web Graphical User Interface (GUI), developed in Seagull PHP Framework, offering a wide range of search and sequence extraction facilities. The database is compiled extracting and integrating information from public resources and data generated by the MitoRes team. In order to build up the database informative content, we implement an automated annotation procedure, a suite of Bio-Perl and C programs, which retrieves data from external data sources and populates database records automatically. Data extracted from external databases include: proteins information and GO annotation from UniProt, genes and genomic information from ENSEMBL, transcript information from RefSeq and UTRdb databases. MitoRes contains also a section named CLUSTERS which is obtained through an “all versus all” global pair-wise alignment of protein sequences (EMBOSS Stretcher program) followed by an automated clustering based on a threshold similarity value of at least 60%. The high level of automatization of the annotation procedure allows efficient and frequent updates of the database content.

Results

The MitoRes database consists of comprehensive sequence entries whose key data are represented by the sequences of protein, transcript(s) and gene (when available) and by the taxonomic information describing the biological source of the protein. Additional information includes: bio-sequences structure and location, biological function of protein product and dynamic links to databases used as data resources and to other databases reporting complementary information. A graphical representation of gene and transcript sequences is provided through the dynamic construction of their physical maps. The Cluster Section contains data derived from the clustering of MitoRes proteins collection based on their sequence similarity. Cluster entries are composed of the list of “Cluster Components” and for each one of

them is reported: the MitoRes entry name, taxonomic information, sequence similarity percentage, gene name and map. The MitoRes web interface has been designed to give graphical access to the database informative content in a consistent and comprehensive way. Through this interface it is possible to run multi-record queries combining different search criteria and to customize the search results view. Furthermore, MitoRes includes a powerful sequence export manager for the massive extraction of sequences and sub-sequences (protein, transcript, gene, flanking gene regions, CDS, signal peptide, UTRs, intron, exon) in different file formats (FASTA, GenBank, EMBL). The current version of MitoRes is based on the UniProt Release 4 and includes 3180 entries and 579 Clusters from about 64 different metazoan species.

Contact email: flavio.licciulli@ba.itb.cnr.it

URL: <http://www2.ba.itb.cnr.it/MitoRes/>