

PeroxisomeDB 2.0: an integrative view of the global peroxisomal metabolome

Agatha Schlüter^{1,2}, Alejandro Real-Chicharro³, Toni Gabaldón⁴,
Francisca Sánchez-Jiménez^{2,3} and Aurora Pujol^{1,2,5,*}

¹Neurometabolic Disease Lab, Institut de Neuropatologia de Bellvitge (IDIBELL), Gran Via n 199, 08907 L'Hospitalet de Llobregat, Barcelona, ²CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ³Biología Molecular y Bioquímica, Universidad de Málaga, Málaga, ⁴Bioinformatics Department, Centre de Regulació Genòmica (CRG), Barcelona and ⁵Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain

Received August 19, 2009; Revised October 7, 2009; Accepted October 9, 2009

ABSTRACT

Peroxisomes are essential organelles that play a key role in redox signalling and lipid homeostasis. They contain a highly diverse enzymatic network among different species, mirroring the varied metabolic needs of the organisms. The previous PeroxisomeDB version organized the peroxisomal proteome of humans and *Saccharomyces cerevisiae* based on genetic and functional information into metabolic categories with a special focus on peroxisomal disease. The new release (<http://www.peroxisomeDB.org>) adds peroxisomal proteins from 35 newly sequenced eukaryotic genomes including fungi, yeasts, plants and lower eukaryotes. We searched these genomes for a core ensemble of 139 peroxisomal protein families and identified 2706 putative peroxisomal protein homologs. Approximately 37% of the identified homologs contained putative peroxisome targeting signals (PTS). To help develop understanding of the evolutionary relationships among peroxisomal proteins, the new database includes phylogenetic trees for 2386 of the peroxisomal proteins. Additional new features are provided, such as a tool to capture kinetic information from Brenda, CheBI and Sabio-RK databases and more than 1400 selected bibliographic references. PeroxisomeDB 2.0 is a freely available, highly interactive functional genomics platform that offers an extensive view on the peroxisomal metabolome across lineages, thus facilitating comparative genomics and systems analysis of the organelle.

INTRODUCTION

Peroxisomes were discovered in 1954 with electron microscopy (1) and isolated by De Duve and Baudhuin in 1966 (2). Only recently, their evolutionary and ontogenetic origin has been related to the endoplasmic reticulum (3–5). Peroxisomes contribute to many crucial metabolic processes such as fatty acid oxidation, biosynthesis of ether lipids and free radical detoxification (6–8). These organelles are essential for normal development and growth, and their deficiencies lead to severe and often fatal inherited peroxisomal disorders in humans (9). With the aim of gaining knowledge of peroxisome biology, we have created PeroxisomeDB (<http://www.peroxisomeDB.org>), a relational database that compiles and integrates functional and genomic information about peroxisomal proteomes of *Homo sapiens* and *Saccharomyces cerevisiae*, organized in metabolic pathways (10). Key features of the database included a peroxisomal targeting signal (PTS) predictor and a peroxisome predictor, useful to identify peroxisomes and candidate peroxisomal proteins in newly sequenced genomes (10).

Peroxisomes differ substantially among species with respect to their enzyme content and specific metabolic functions. Some examples of the remarkable metabolic plasticity of peroxisomes include: (i) photorespiration in photosynthetic organisms; (ii) conversion of fatty acids into carbohydrates through the glyoxylate cycle in plants, protozoa and some yeasts (therefore designated as glyoxysomes); and (iii) glycolysis, purine salvage and pyrimidine biosynthesis, and pentose phosphate pathway in the glycosomes of trypanosomatids (11–13). To better compile the diversity of peroxisomes, we have introduced PeroxisomeDB 2.0. We generated this comprehensive catalogue by searching 37 taxonomically diverse

*To whom correspondence should be addressed. Tel: +34 93 2607343; Fax: +34 93 2607414; Email: apujol@idibell.cat

full-sequenced genomes and identified the global peroxisomal metabolome. The database includes high-quality trees to make it easier to understand the evolutionary history of peroxisome proteins. PeroxisomeDB 2.0 aims at organising curated and updated, biochemical, genomic and phylogenetic information on peroxisomal proteomes and their interactions.

RESULTS AND DISCUSSION

Capturing the peroxisomal proteome from 37 eukaryotic genomes

With a comparative genomics approach, we predicted peroxisomal proteins *in silico* based on homology (Figure 1). First, we obtained 85 *H. sapiens* and 61 *S. cerevisiae* peroxisomal proteins from PeroxisomeDB (10), the proteomes of glycosomes from the trypanosomatids *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania major*, manually curated according to literature (14), and the glyoxysomal proteome from the *Arabidopsis thaliana* genome contained in the Araperox database (15). Second, to identify peroxisomal protein homologs from 19 additional eukaryotic genomes in ENSEMBL, we used orthology predictions from ENSEMBL COMPARA (16) database version 47 (<http://oct2007.archive.ensembl.org/>). Third, from the resulting peroxisomal ensemble, we designed 139 peroxisomal protein families or core functional units, using a tool based on the profile Hidden Markov Models, HMMER (17). Finally, to identify the homologs

to these peroxisomal protein family units, we calculated a HMM best hit score for searches against 37 eukaryotic genomes (35 new genomes, *H. sapiens* and *S. cerevisiae*). A putative peroxisomal homolog had an HMM score of at least 10^{-3} . This search yielded a set of 2706 predicted peroxisomal homologous proteins, which were deposited into PeroxisomeDB 2.0.

We sampled genomes with the goal of encompassing eukaryotic phylogenetic diversity, from mammals to protists or lower eukaryotes, including plants and fungi. The 37 eukaryotes included trypanosomatids (*L. major*, *T. brucei* and *T. cruzi*); the red algae *Cyanidioschyzon merolae*, a unicellular organism that contains a single peroxisome; the diatom *Thalassiosira pseudonana*, which acquired a plastid through a secondary endosymbiotic process; the ciliate *Tetrahymena thermophila* which is closely related to the Apicomplexa that have no peroxisomes (4,18); and the social amoeba *Dictyostelium discoideum* from a eukaryotic lineage prior to the split of metazoans and yeasts. A list of all species included in the database, and their genome sources is available at PeroxisomeDB 2.0 (<http://www.peroxisomedb.org/show.php?action=OrganismCatalogue>).

Fishing for PTS *in silico*

Our previously developed PTS predictor (10) is a powerful tool that provides PTS1, PTS2 and PEX19BS predictions by using a sequence multiple alignment without gaps or BLOCK (19). We performed high-scale motif searches using Blimps software (20) applied

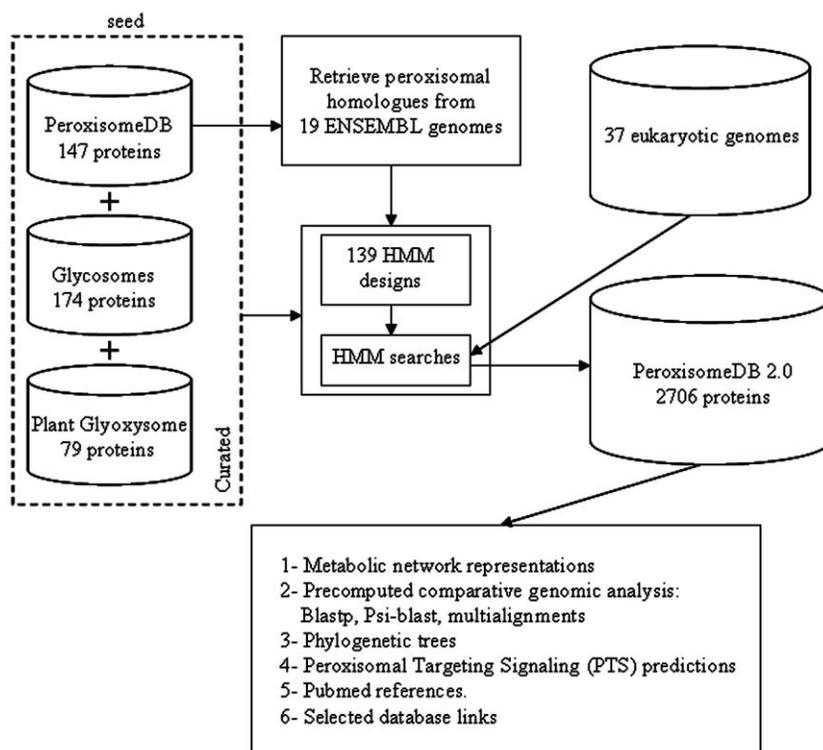


Figure 1. An outline on how the PeroxisomeDB 2.0 was generated. The PeroxisomeDB 2.0 includes proteins from yeast and human (PeroxisomeDB previous release), the glycosome proteome from trypanosomatids and the glyoxysome proteome from *Arabidopsis thaliana*. Using a protein consensus designed by HMM, we have scanned 37 eukaryotic genomes to identify 2706 putative peroxisomal homologs.

to each PTS BLOCK to screen the 2706 identified peroxisomal proteins. We identified 1126 PTSs from 944 proteins of the 2706 entries. Most proteins had either a PTS1 signal (571) or a PEX19BS signal (482). Only 73 proteins had PTS2 signals, although this motif is only known to be a part of 3-ketoacyl-CoA thiolase (Acaa1), Phytanoyl-CoA dioxygenase (Phyh) and Alkylidihydroxyacetonephosphate synthase (Agps) protein families. We confirmed a PTS2 signal in several known peroxisomal enzymes, such as Acyl-coenzyme A oxidase (Acox) and Citrate synthase (Cs) of *D. discoideum* and *A. thaliana* (21,22), and the Glycerol-3-phosphate dehydrogenase (GPD) of yeasts (23). We also predicted a PTS2 in the Glycosyl hydrolase (PEN2) of *A. thaliana*, experimentally identified as peroxisomal protein (24).

The global peroxisomal metabolome at a glance

Each of the 2706 proteins has a database entry that includes information on the assigned protein family, functional category and subcellular localisation, as well as links to reference databases and Pubmed. The entry also summarizes the PTS predictions and homology computations, including BLASTP, PSI-BLAST, multialignments and phylogenetic trees. Finally, enzymes entries are linked through the ontology-based mediator used by SBMM Assistant (www.smmm.uma.es) (25) to the reference databases Brenda, CheBI and Sabio-RK, which provide updated kinetic and biochemical information for peroxisomal enzymes. This assistant works by recognizing the respective EC codes from Gene Bank and/or Ensemble and subsequently fetches the relevant data. A summary of the novel functions of PeroxisomeDB 2.0 as compared to the previous PeroxisomeDB is available as Supplementary Table S1.

A central feature of the database is a global Peroxisome Metabolome Network composed of the most relevant known peroxisomal metabolic pathways, and it includes the putative peroxisomal homologs assigned to their respective protein families after identification in the current analysis (Figure 2). The integration into this comprehensive and complete metabolomic picture is of particular relevance for organisms such as *Anopheles gambiae* and *Tetraodon nigroviridis*, to cite but two, whose proteins lack curated peroxisomal annotation at NCBI and, to the best of our knowledge, experimental peroxisomal localisation (see Supplementary Table S2). We have developed a function for viewing the metabolome networks in a species-specific manner (Figure 2). Beyond these species-specific metabolic pathways, PeroxisomeDB 2.0 emerges as the integrative tool for the study of the global peroxisomal metabolome.

Molecular phylogenetic analysis of the global peroxisomal metabolome

Using MUSCLE 3.6 (26), TrimAl v1.2 (27), PhyML aLRT version (28,29) and Mr Bayes (30) as described in PeroxisomeDB (http://www.peroxisomedb.org/peroxHelp/PROGRAMS_3.php), we built 2386 phylogenetic trees. An example of a phylogenetic tree

for the Malate dehydrogenase (Mdh) enzyme is shown in Figure 3. For the remaining 320 peroxisomal homologs, we did not build a tree because fewer than four homologs were identified in the eukaryotic genomes. For some model species, such as *S. cerevisiae* and humans, links to additional phylogenomic resources are available through PhylomeDB database (31).

This molecular phylogenetic approach yielded important findings with functional implications. Beyond the well-known peroxisomal core metabolic pathways that confer identity to the organelle, such as catalase and beta-oxidation of fatty acids (6–8), we identified additional ancient peroxisomal acquisitions in taxonomically diverse organisms. For example, Peroxiredoxin 1 (Prdx1) in mammals (32) and Tryparedoxin peroxidase (Tpx) of trypanosomatids were closely related, suggesting a common peroxisomal origin for the peroxiredoxin system. Similarly, the trypanosomatid NADP⁺-dependent dehydrogenase Glucose-6-phosphate 1-dehydrogenase (G6pdh) of the pentose phosphate pathway of glycosomes is present in all eukaryotic organisms analysed here, although the peroxisomal localisation of the pathway has only been described in trypanosomatids to date. This finding might provide a rationale for previous experimental data showing 10% enzymatic activity of these enzymes in rat liver peroxisomes (33), raising the possibility of a partial peroxisomal location of the pathway in mammals. Additionally, the phylogenies revealed a common origin for metazoan Coenzyme A diphosphatase or nucleoside diphosphate-linked moiety X motif 7 (Nudt7 or Nudix motif 7) and the Peroxisomal nudix pyrophosphatase with specificity for coenzyme A and CoA derivatives (Pcd1) of yeasts, reinforcing previous suggestions (34).

Not all currently known peroxisomal enzymes were part of the original peroxisome, examples being some enzymes of beta-oxidation of fatty acids route (3). Peroxisomes of different lineages have also acquired enzymes as the metabolic needs of the organisms changed. Interestingly, we detected independent events of the same homologous protein from different species recruited into the peroxisome. Some enzymes of the glyoxylate pathway have been independently incorporated into the peroxisome of yeasts, plants and trypanosomatids (35). This is the case for the Mdh (see tree Figure 3), which evolved from duplications of homologous proteins in another cellular location, which occurred after eukaryote divergence. For Cs, duplications occurred in yeasts, but a horizontal transfer from eubacteria was at the origin of the peroxisomal enzyme in plants (35). This is an excellent example of the metabolic constraints of the cell acting as the driving force in the recruitment of a given key function to peroxisomes, several times independently.

Future challenges

As organelle-specific proteomic technologies and whole genome sequencing projects progress, the number of known peroxisomal proteins is exponentially growing. It is challenging to keep up with these protein discoveries while maintaining a high-quality curated database.

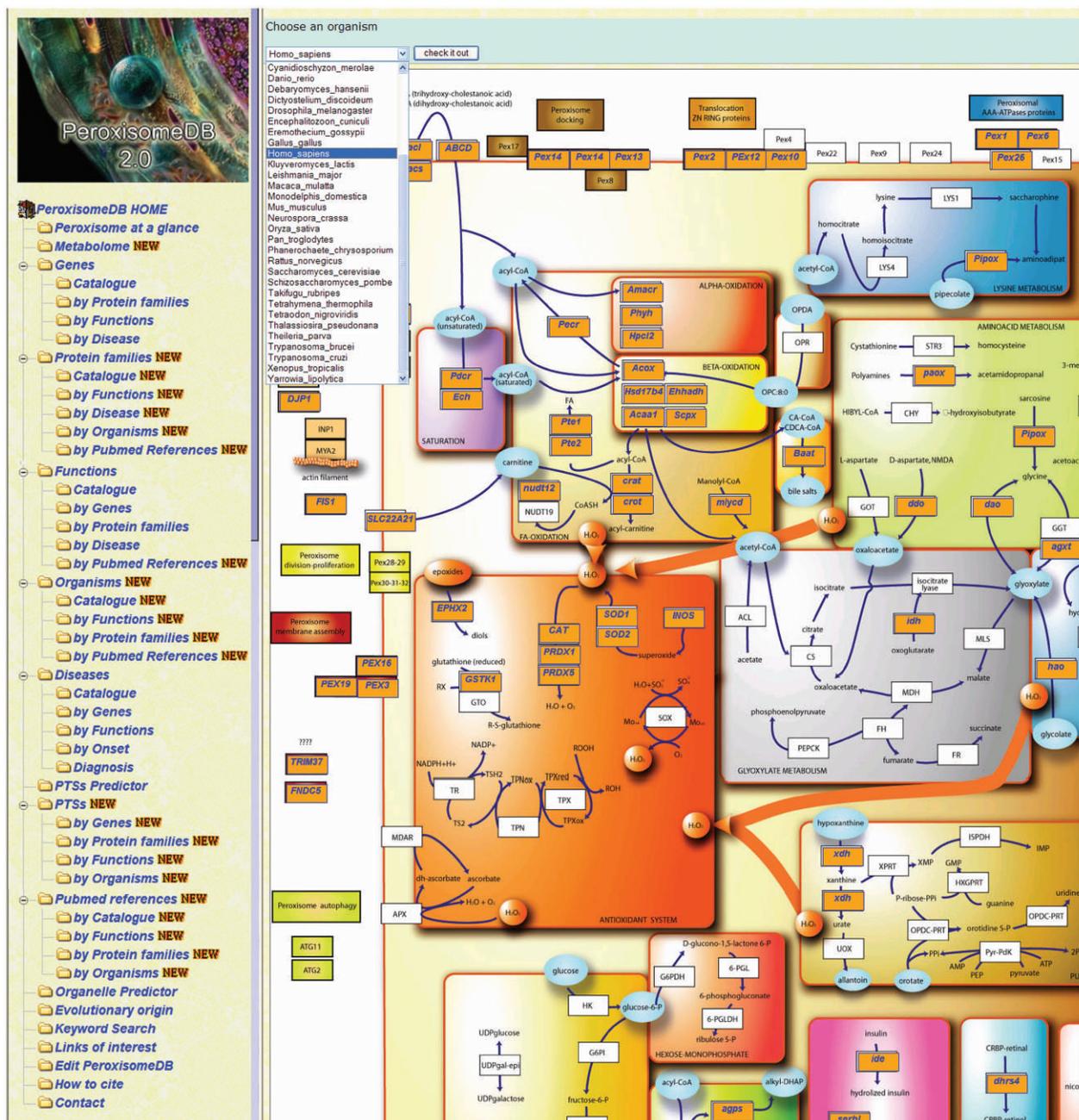


Figure 2. The Metabolome page displays the global peroxisomal metabolome network for the 37 eukaryotic genomes. The tool allows the user to focus on a species of interest by clicking on 'choose an organism'.

The PeroxisomeDB 2.0 currently includes manually curated human, *S. cerevisiae*, *A. thaliana* and trypanosomatid peroxisomal proteins with the peroxisomal proteins from the remaining species assigned based on an automatic sequence homology. We plan to replace this predicted peroxisomal annotation to fetch experimental organelle localisation with a semi-automatic curation system for semantic data integration and text mining from literature, which will be relaunched every two months. Precomputed homology searches using BLAST and PSI-BLAST will also be relaunched bimonthly. A collaborative interface will be developed to encourage user input to complement these

routes and benefit from the collective knowledge of the scientific community. Finally, peroxisomes are dynamic compartments metabolically linked to other subcellular structures, such as mitochondria and endoplasmic reticulum (36). Extension of the peroxisomal metabolome beyond the organelle borders to encompass these shared metabolic pathways is an exciting perspective.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

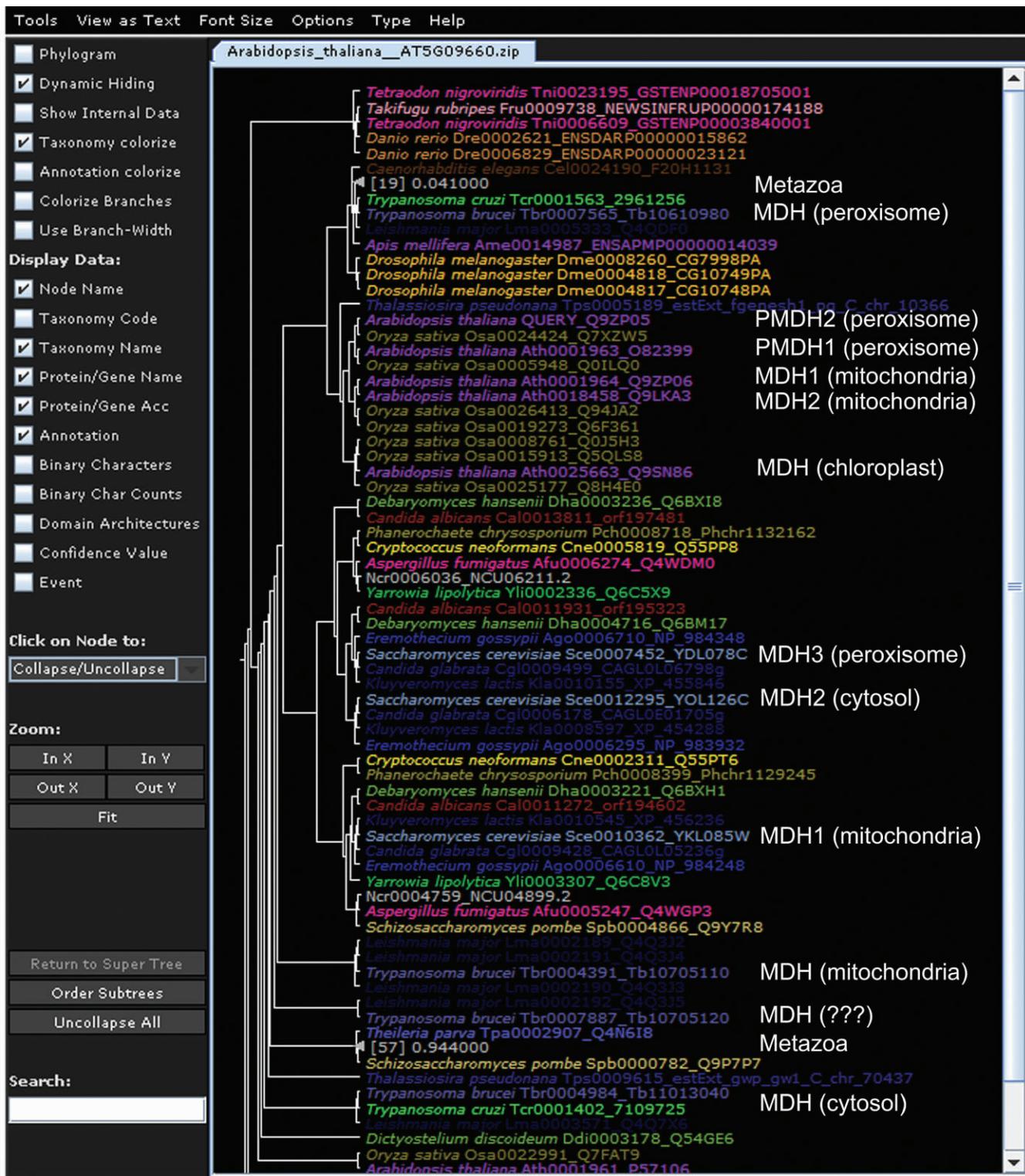


Figure 3. The maximum likelihood tree (ML) of malate dehydrogenase (Mdh) protein family, taking the peroxisomal Mdh from *Arabidopsis thaliana* as a seed. Different species names are given in different colours. Phylogenetic trees are available in newick standard format and associated with taxonomic data via phyloXML, an XML language for the analysis, exchange, and storage of phylogenetic trees and associated data. The Archaeopteryx applet software was used for the visualisation of phylogenetic trees (37).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the computer resources, technical expertise and assistance provided by the Barcelona Supercomputing Center—Centro Nacional de Supercomputación, and The Spanish National Bioinformatics Institute—Instituto Nacional de Bioinformática (INB). Part of the simulations have been done at the freely available Biportal (www.biportal.uio.no).

FUNDING

Grant FIS PI080991 and 2009SGR85 (to A.P.); SAF2008-2522 (to F.S.-J.); Grant CVI-2999 (to A.R.-C.); FIS (ECA07/055 to A.S.). The work was developed under the COST action BM0604 (to A.P.). CIBERER is an initiative of the Instituto de Salud Carlos III. Funding for open access charge: Grant (FIS PI080991).

Conflict of interest statement. None declared.

REFERENCES

- Rhodin (1954) Correlation of ultrastructural organization and function in normal experimentally changed convoluted tubule cells of the mouse kidney. *Thesis*. Karolinska Institutet, Stockholm, Aktiebolaget Godvil.
- De Duve, C. and Baudhuin, P. (1966) Peroxisomes (microbodies and related particles). *Physiol. Rev.*, **46**, 323–357.
- Gabalton, T., Snel, B., van Zimmeren, F., Hemrika, W., Tabak, H. and Huynen, M.A. (2006) Origin and evolution of the peroxisomal proteome. *Biol. Direct*, **1**, 8.
- Schluter, A., Fourcade, S., Ripp, R., Mandel, J.L., Poch, O. and Pujol, A. (2006) The evolutionary origin of peroxisomes: an ER-Peroxisome Connection. *Mol. Biol. Evol.*, **23**, 838–845.
- Hoepfner, D., Schildknecht, D., Braakman, I., Philippsen, P. and Tabak, H.F. (2005) Contribution of the endoplasmic reticulum to peroxisome formation. *Cell*, **122**, 85–95.
- Bonekamp, N.A., Volkl, A., Fahimi, H.D. and Schrader, M. (2009) Reactive oxygen species and peroxisomes: struggling for balance. *Biofactors*, **35**, 346–355.
- Schrader, M. and Fahimi, H.D. (2008) The peroxisome: still a mysterious organelle. *Histochem. Cell Biol.*, **129**, 421–440.
- Wanders, R.J. and Waterham, H.R. (2006) Biochemistry of mammalian peroxisomes revisited. *Annu. Rev. Biochem.*, **75**, 295–332.
- Wanders, R.J. (2004) Metabolic and molecular basis of peroxisomal disorders: a review. *Am. J. Med. Genet.*, **126A**, 355–375.
- Schluter, A., Fourcade, S., Domenech-Estevé, E., Gabalton, T., Huerta-Cepas, J., Berthommier, G., Ripp, R., Wanders, R.J., Poch, O. and Pujol, A. (2007) PeroxisomeDB: a database for the peroxisomal proteome, functional genomics and disease. *Nucleic Acids Res.*, **35**, D815–D822.
- Kunze, M., Pracharoenwattana, I., Smith, S.M. and Hartig, A. (2006) A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochim. Biophys. Acta*, **1763**, 1441–1452.
- Michels, P.A., Bringaud, F., Herman, M. and Hannaert, V. (2006) Metabolic functions of glycosomes in trypanosomatids. *Biochim. Biophys. Acta*, **1763**, 1463–1477.
- Reumann, S. and Weber, A.P. (2006) Plant peroxisomes respire in the light: some gaps of the photorespiratory C2 cycle have become filled—others remain. *Biochim. Biophys. Acta*, **1763**, 1496–1510.
- Opperdoes, F.R. and Szikora, J.P. (2006) In silico prediction of the glycosomal enzymes of *Leishmania major* and trypanosomes. *Mol. Biochem. Parasitol.*, **147**, 193–206.
- Reumann, S., Ma, C., Lemke, S. and Babujee, L. (2004) AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes. *Plant Physiol.*, **136**, 2587–2608.
- Hubbard, T.J., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T. *et al.* (2007) Ensembl 2007. *Nucleic Acids Res.*, **35**, D610–D617.
- Krogh, A., Brown, M., Mian, I.S., Sjolander, K. and Haussler, D. (1994) Hidden Markov models in computational biology. Applications to protein modeling. *J. Mol. Biol.*, **235**, 1501–1531.
- Ding, M., Clayton, C. and Soldati, D. (2000) Toxoplasma gondii catalase: are there peroxisomes in toxoplasma? *J. Cell Sci.*, **113**(Pt 13), 2409–2419.
- Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.
- Henikoff, J.G., Pietrokovski, S., McCallum, C.M. and Henikoff, S. (2000) Blocks-based methods for detecting protein homology. *Electrophoresis*, **21**, 1700–1706.
- Huang, Y.C., Chen, Y.H., Lo, S.R., Liu, C.I., Wang, C.W. and Chang, W.T. (2004) Disruption of the peroxisomal citrate synthase CshA affects cell growth and multicellular development in *Dictyostelium discoideum*. *Mol. Microbiol.*, **53**, 81–91.
- Pracharoenwattana, I., Cornah, J.E. and Smith, S.M. (2005) Arabidopsis peroxisomal citrate synthase is required for fatty acid respiration and seed germination. *Plant Cell*, **17**, 2037–2048.
- Valadi, A., Granath, K., Gustafsson, L. and Adler, L. (2004) Distinct intracellular localization of Gpd1p and Gpd2p, the two yeast isoforms of NAD⁺-dependent glycerol-3-phosphate dehydrogenase, explains their different contributions to redox-driven glycerol production. *J. Biol. Chem.*, **279**, 39677–39685.
- Lipka, V., Dittgen, J., Bednarek, P., Bhat, R., Wiermer, M., Stein, M., Landtag, J., Brandt, W., Rosahl, S., Scheel, D. *et al.* (2005) Pre- and postinvasion defenses both contribute to nonhost resistance in Arabidopsis. *Science*, **310**, 1180–1183.
- Reyes-Palmares, A., Montanez, R., Real-Chicharro, A., Chniber, O., Kerzazi, A., Navas-Delgado, I., Medina, M.A., Aldana-Montes, J.F. and Sanchez-Jimenez, F. (2009) Systems biology metabolic modeling assistant: an ontology-based tool for the integration of metabolic data in kinetic modeling. *Bioinformatics*, **25**, 834–835.
- Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
- Capella-Gutierrez, S., Silla-Martinez, J.M. and Gabalton, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
- Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.*, **55**, 539–552.
- Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**, 696–704.
- Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Huerta-Cepas, J., Bueno, A., Dopazo, J. and Gabalton, T. (2008) PhylomeDB: a database for genome-wide collections of gene phylogenies. *Nucleic Acids Res.*, **36**, D491–D496.
- Immschuh, S., Baumgart-Vogt, E., Tan, M., Iwahara, S., Ramadori, G. and Fahimi, H.D. (2003) Differential cellular and subcellular localization of heme-binding protein 23/peroxiredoxin I and heme oxygenase-1 in rat liver. *J. Histochem. Cytochem.*, **51**, 1621–1631.
- Antononkov, V.D. (1989) Dehydrogenases of the pentose phosphate pathway in rat liver peroxisomes. *Eur. J. Biochem.*, **183**, 75–82.
- Gasmi, L. and McLennan, A.G. (2001) The mouse Nudt7 gene encodes a peroxisomal nudix hydrolase specific for coenzyme A and its derivatives. *Biochem. J.*, **357**, 33–38.
- Schnarrenberger, C. and Martin, W. (2002) Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. *Eur. J. Biochem.*, **269**, 868–883.
- Schrader, M. and Yoon, Y. (2007) Mitochondria and peroxisomes: are the ‘big brother’ and the ‘little sister’ closer than assumed? *Bioessays*, **29**, 1105–1114.
- Zmasek, C.M. and Eddy, S.R. (2001) ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics*, **17**, 383–384.