Journal of Experimental Botany, Vol. 60, No. 2, pp. 453–459, 2009
doi:10.1093/jxb/ern317 Advance Access publication 26 December, 2008

Journal of Experimental Botany
www.jxb.oxfordjournals.org

REVIEW PAPER

PeroxiBase: a powerful tool to collect and analyse peroxidase sequences from Viridiplantae

Michele Oliva[1,*], Grégory Theiler[1], Marcel Zamocky[2], Dominique Koua[1,†], Marcia Margis-Pinheiro[3], Filippo Passardi[1] and Christophe Dunand[4,‡]

[1] Department of Plant Biology, University of Geneva, Quai Ernest-Ansermet 30, CH-1211 Geneva 4, Switzerland
[2] Department of Chemistry, University of Natural Resources and Applied Life Sciences, Vienna, Austria
[3] Federal University of Rio Grande do Sul, Department of Genetics, Institute of Biology, Brazil
[4] Plant Cell Surfaces and Signaling Laboratory, University of Toulouse, UPS, CNRS, 24 Chemin de Borderouge, F-31326 Castanet-Tolosan, France

Received 15 August 2008; Revised 13 November 2008; Accepted 17 November 2008

Abstract

Peroxidases are enzymes that are implicated in several biological processes and are detected in all living organisms. The increasing number of sequencing projects and the poor quality of annotation justified the creation of an efficient tool that was suitable for collecting and annotating the huge quantity of data. Started in 2004 to collect only class III peroxidases, PeroxiBase has undergone important updates since then and, currently, the majority of peroxidase sequences from all kingdoms of life is stored in the database. In addition, the web site (http://peroxibase.isb-sib.ch) provides a series of bioinformatics tools and facilities suitable for analysing these stored sequences. In particular, the high number of isoforms in each organism makes phylogenetic studies extremely useful to elucidate the complex evolution of these enzymes, not only within the plant kingdom but also between the different kingdoms. This paper provides a general overview of PeroxiBase, focusing on its tools and the stored data. The main goal is to give researchers some guidelines to extract classified and annotated sequences from the data base in a quick and easy way in order to perform alignments and phylogenetic analysis. The description of the database is accompanied by the updates we have recently carried out in order to improve its completeness and make it more user-friendly.

Key words: Database, evolution, peroxidases, phylogenetic analyses, Viridiplantae.

Introducing PeroxiBase: classification of peroxidases in the database

Peroxidases (EC 1.11.1.x) are enzymes able to carry out a reaction in which peroxide is reduced and a substrate is oxidized. Although they have been found in all kingdoms, in plants they assume fundamental roles in all tissues and during the whole life cycle. In particular, class III peroxidases are plant-specific enzymes located in cell walls: their ability to cleave cell wall polysaccharides and to form diferulic bonds, makes them key players in the regulation of cell wall formation and thus in the cell expansion process (Liszkay *et al.*, 2003; Passardi *et al.*, 2005). The large number of isoforms detected in plant genomes (e.g. 73 in *Arabidopsis thaliana*) made the building of a database, that centralized the records of the sequences and related information about these cell wall proteins, necessary (Bakalovic *et al.*, 2006). In a second step, the database was enlarged to include the other types of plant peroxidases as well as peroxidases from the other kingdoms (Passardi *et al.*, 2007*b*). This exhaustive data mining allowed new classes and subclasses to be defined. All the peroxidases have been put into two large sets: Haem peroxidases and Non-Haem peroxidases (Table 1). The

* Present address: BioQuant Center, Heidelberg Institute for Plant Sciences, University of Heidelberg, D-69120 Heidelberg, Germany.
† Present address: Swiss Institute of Bioinformatics, Swiss-Prot Group, CMU-1, rue Michel Servet, CH-1211 Geneva 4, Switzerland.
‡ To whom correspondence should be addressed: E-mail: dunand@scsv.ups-tlse.fr

Table 1. Peroxidase families and superfamilies included in the database with their distributions across the major kingdoms

The presence or the absence of the different families in the various taxonomic groups has been identified. (*) Marginal presence probably due to a gene transfer.

	Prokaryotes	Plants	Fungi	Animals	Other eukaryotes	Total sequences
Haem peroxidase						
Peroxidase-Cyclooxygenase superfamily (Animal peroxidase)						
Alpha-dioxygenase (DiOx)		√				**28**
Other animal peroxidase (12 subfamilies)	√		√	√	√	**323**
Catalase (Kat)	√	√	√	√	√	**206**
Di-haem peroxidase superfamily	√					**115**
DyP-type peroxidase	√		√		√	**80**
Haloperoxidase			√		√	**49**
Non-Animal peroxidase						
Class I peroxidase						
Ascorbate peroxidase (APx)		√		*		**385**
Catalase peroxidase	√		*		*	**347**
Cytochrome *c* peroxidase		*	√		√	**103**
Class II peroxidase			√			**120**
Class III peroxidase		√				**2656**
Non-haem peroxidase						
Alkylhydroperoxidase D-like superfamily	√		*			**190**
Haloperoxidase	√		√		√	**54**
Manganese catalase	√					**93**
NADH peroxidase, oxidase and dehydrogenase	√					**76**
Thiol peroxidase						
Glutathione peroxidase (GPx)	√	√	√	√	√	**408**
Peroxiredoxin						
1-Cysteine peroxiredoxin (1CysPrx)	√	√	√	√	√	**109**
Typical 2-Cysteine peroxiredoxin (2CysPrx/AhpC)	√	√	√	√	√	**230**
Atypical 2-Cysteine peroxiredoxin (PrxII, PrxV, PrxGrx)	√	√	√	√	√	**91**
PrxII-glutoredoxin fusion	√					**51**
Atypical 2-Cysteine peroxiredoxin (PrxQ, BCP)	√	√	*		√	**115**
Thioredoxin-dependent thiol peroxidase	√				√	**22**
AhpE like peroxiredoxin	√					**23**
NADPH oxidase	√	√	√	√	√	**150**

former is composed of six groups: Animal peroxidase/peroxidase-cyclo-oxygenases, Catalases, Di-haem peroxidases, DyP-type peroxidases, Haloperoxidases, and Non-animal peroxidases. The Non-Haem peroxidases category contains five groups: Alkylhydroperoxidases D-like, Haloperoxidases, Mn catalases, NADH peroxidases, and Thiol peroxidases. Some of them are referred to as 'superfamily' because of the large number of subgroups. Despite the lack of peroxidase domains, the NAD(P)H oxidase group has recently been included in the database for its homology with the second domain of Dual oxidase proteins (a member of the Animal peroxidase/peroxidase-cyclo-oxygenase superfamily). Finally, an effort has been made in the data entry to increase the number of sequences represented in the various Viridiplantae groups, especially those that were poorly represented in the database. A benefit of this update is that multiple alignments may be built up from more closely related sequences, thus making broader phylogenetic studies possible. In addition, the continuous recruitment of sequences in the same group will allow reliable prediction of the duplication processes occurring within one class in a single organism to be performed.

Extracting data from PeroxiBase: new insights into the evolution of peroxidase-encoding sequences in Viridiplantae

The availability of classified, well-annotated sequences helps in performing global analyses on many living organisms. BLAST tools, as well as multicriteria search tools, allow results to be obtained in Fasta, a format also suitable for alignments and phylogenetic analyses (Fig. 1). The continuous recruitment of sequences in the Viridiplantae class has recently led to interesting findings. For instance, the following peroxidase families and superfamilies could not be found in Viridiplantae: Di-haem peroxidase superfamily, DyP-type peroxidase, Class II peroxidase, Alkylhydroperoxidase D-like superfamily, Haloperoxidase, Manganese catalase, NADH peroxidase, oxidase and dehydrogenase (Table 1). Alpha-dioxygenase (DiOx), identified exclusively in Viridiplantae, is the only member of the large animal peroxidase superfamily present in plants (Table 1). Surprisingly, no DiOx-encoding sequences have been detected in the well-sequenced Chlorophyta group (complete genome of *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*) (Table 2). Among the different peroxidase classes, the

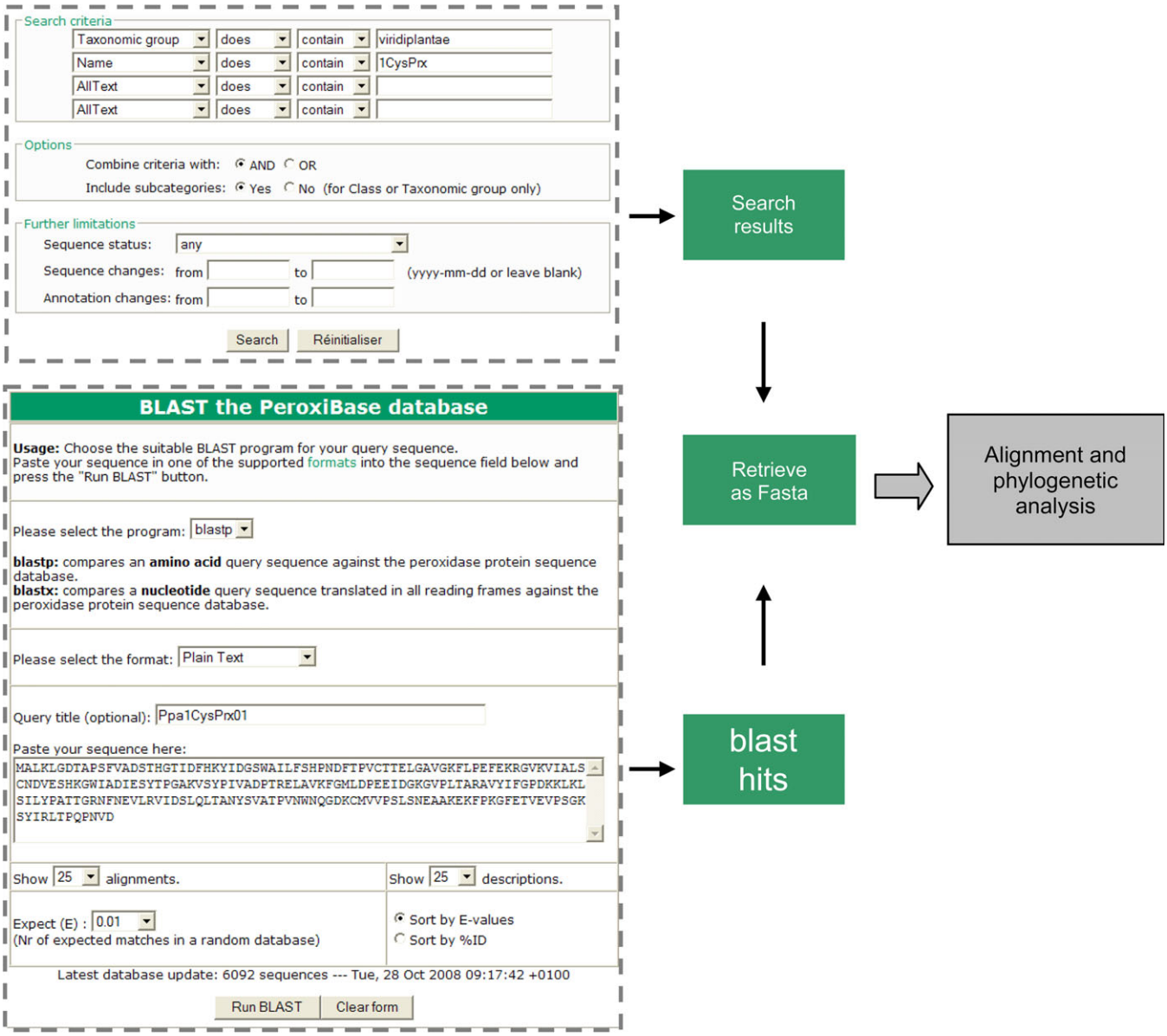

Fig. 1. Typical procedures to prepare a set of selected sequences before alignment and phylogenetic analysis. Search with multicriteria allows class sequences to be selected from a taxonomic group. By using one selected sequence, BLAST allows homologous sequences to be obtained. In both cases, the results can be retrieved in Fasta format before being exported to bioinformatics software for alignment and phylogenetic analyses.

peroxiredoxins containing 1 cysteine (1CysPrx) and the atypical peroxiredoxins containing 2 cysteines (2CysPrx) can be detected in all kingdoms (Table 1). A phylogenetic analysis of representative 1CysPrx and 2CysPrx sequences coming from Viridiplantae has been performed to illustrate the potential source of information available in the database. As expected, the tree shows a well-supported divergence between 1CysPrx and 2CysPrx branches that originated from the same ancestral sequence (Fig. 2). Contrary to the 2CysPrx sequences that exist in all Viridiplantae, and with extensive data mining, 1CysPrx sequences were not detected in Chlorophyta. The absence of Class III peroxidase [only a hybrid sequence has been found in *C. reinhardtii* (Passardi *et al.*, 2007*a*)] and 1-cysteine peroxiredoxin (1CysPrx) from the Chlorophyta group seems to confirm the evolutionary divergence between multicellular green organisms and land plants. Since their separation more than 500 MY ago, the land plants have acquired actual class III peroxidase sequences and the green algae have lost the 1CysPrx sequence. More sequences from liverworts or more basal Viridiplantae need to be found to confirm this hypothesis. However, additional analyses of Viridiplantae sequences available in the database show a clear redundancy of genes of the same class within the genome of a single species. The evolution, the spread, and the maintenance of such a high number of isoforms have been discussed in several recent works. For instance, the global comparison among all glutathione

Table 2. Peroxidase-encoding sequence detected in the various Viridiplantae groups

The presence or the absence of the different families in the various taxonomic groups has been identified. !: one sequence has been detected for only one organism. ?: no sequence has been detected in an organism with small EST library. ns: no sequence has been detected in an organism completely sequenced or with a large EST library. The values in brackets stand for the number of isoforms detected. List of organisms used to find representative sequences of the different classes: Chlorophyta (complete genome from *Chlamydomonas reinhardtii*, *Ostreococcus lucimarinus*), Cryptogam (complete genome from *Selaginella moellendorffii*, *Physcomitrella patens*), Gymnospermae (Coniferophyta (900 000 ESTs), Cycadophyta (40 000 ESTs), Ginkgophyta (19 000 ESTs), Gnetophyta (20 000 ESTs)), Monocotyledons (complete genome from *Oryza sativa*), Eudicotyledons (complete genome from *Arabidopsis thaliana*, *Populus trichocarpa*), Other Angiospermae [basal Magnoliophyta (47 000 ESTs), Magnoliids (69 000 ESTs)] and Other Streptophyta. [Zygnemophyceae, Mesostigmatophyceae (19 000 ESTs)].

		Streptophyta					
				Angiospermae			
	Chlorophyta	**Cryptogam**	**Gymnospermae**	**Monocotyledons**	**Eudicotyledons**	**Other Angiospermae**	**Other Streptophyta**
Haem peroxidase							
Peroxidase-Cyclooxygenase superfamily (Animal peroxidase)							
Alpha-dioxygenase (DiOx)	ns	√ (1)	√ (2)	√ (1)	√ (2)	√ (1?)	?
Catalase (Kat)	√ (2)	√ (2)	√ (2)	√ (3)	√ (3)	√ (2)	?
Non Animal peroxidase							
Class I peroxidase							
Ascorbate peroxidase (APx)	√ (2)	√ (3)	√ (4?)	√ (10)	√ (9)	√ (2?)	√ (1?)
Cytochrome *c* peroxidase	!	ns	ns	ns	Ns	ns	?
Class III peroxidase	!	√	√	√	√	√	√
Non-haem peroxidase							
Thiol peroxidase							
Glutathione peroxidase (GPx)	√ (4)	√ (2?)	√ (2?)	√ (5)	√ (8)	√ (?)	√ (?)
Peroxiredoxin							
1-Cysteine peroxiredoxin (1CysPrx)	ns	√ (2)	√ (1)	√ (2)	√ (1)	√ (1)	?
Typical 2-Cysteine peroxiredoxin (2CysPrx/AhpC)	√ (2)	√ (1)	√ (1)	√ (1)	√ (2)	√ (1?)	√ (1?)
Atypical 2-Cysteine peroxiredoxin (PrxII, PrxV, PrxGrx)	√ (3)	√ (1)	√ (2?)	√ (4)	√ (6)	√ (2)	√ (2?)
Atypical 2-Cysteine peroxiredoxin (PrxQ, BCP)	√ (1)	√ (1)	√ (1)	√ (1)	√ (1)	?	?
NADPH oxidase	√ (2)	√ (4)	√ (2?)	√ (9)	√ (10)	√ (1?)	?
Total sequence/groupe	40	113	213	975	2113	62	6

peroxidase (GPx) sequences available in PeroxiBase (Table 1) showed that plant GPxs form an independent cluster, suggesting that an ancestral gene led to the origin of all plant *GPx* genes. According to this analysis, all plant *GPx* genes were generated by four major duplication events, which occurred before the divergence of Monocotyledons and Dicotyledons (Margis *et al.*, 2008).

GPxs, as well as other classes, are members of multigenic families. A quick specialization of these enzymes has been suggested to play an important role for their retention in the genome (Margis *et al.*, 2008; Passardi *et al.*, 2004). Additional analyses of PeroxiBase sequences showed that redundancy is also present in the taxonomic group of Bacteria. In particular, catalases seem to be over-represented in several species; even though the reason is still not clear, it is possible that their retention is related to the key role they have in removing potentially dangerous hydrogen peroxide in very stressing habitats and in controling H_2O_2 level during signaling (Zamocky *et al.*, 2008).

Global and precise phylogenetic analysis could be performed with other peroxidase classes detected in the various kingdoms. This analysis will bring interesting new insights regarding the evolution of the different peroxidase classes and will orient the direction of future data mining.

Improving PeroxiBase: new tools and new sequences

Since the last paper on PeroxiBase updates (Passardi *et al.*, 2007*b*), new developments have been carried out, at least in three major fields: user interface, search of new sequences (multi-criteria search), tools (Blast, PeroxiScan…).

As concerns the first issue, the stored data can now be directly browsed through one of the six sections available in the PeroxiBase toolbar, namely 'Classes', 'Organisms', 'Cellular localizations', 'Inducers', 'Repressors', and 'Tissue types' (Fig. 3). All the fields are periodically updated, and,

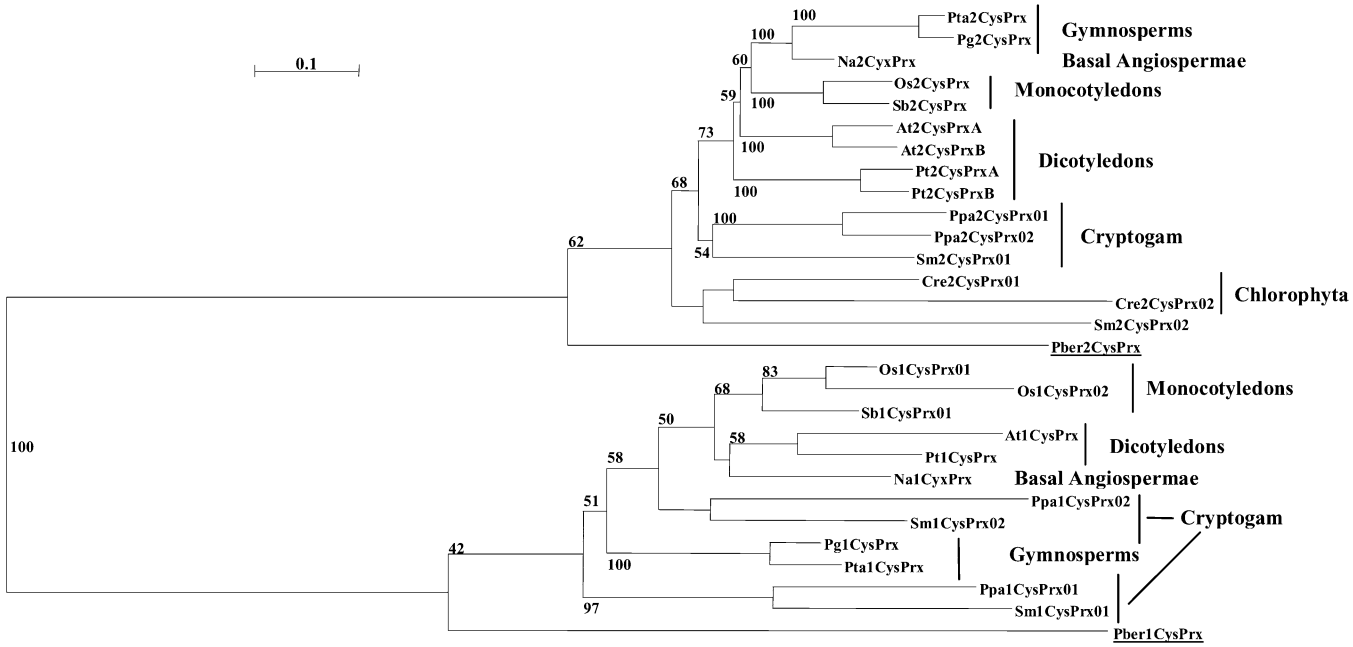

Fig. 2. Phylogenetic tree of protein sequences of 1CysPrx and 2CysPrx from Viridiplantae. The tree was constructed by using Neighbor–Joining method. Values at nodes indicate bootstrap supports greater than 50%. All branches are drawn to scale and the scale bar represents 0.1 substitution per site. Underlined sequences coming from the apicomplexa *Plasmodium berghei* have been used as outgroup sequences.

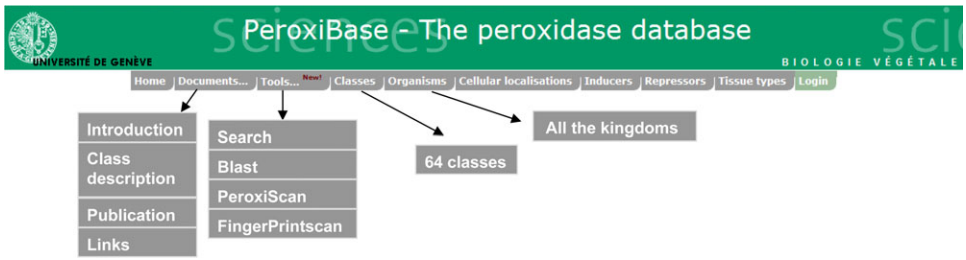

Fig. 3. Screeshot of PeroxiBase toolbar with detailed options. The toolbar includes various sections. Documents contains Introduction, Class description, Publications (related to PeroxiBase) and Links (specific and general databases). Tools menu contains the following rubrics: Search (multi criteria), BLAST, PeroxiScan and FingerPrintscan.

in particular for 'Cellular localizations', 'Inducers' and 'Repressors', data from published papers have been complemented with new insights coming from microarrays. As a fundamental requirement for improving the database, users can enter new sequences, after asking the administrator for a password. A completely new sheet has been created to make the insertion of the sequences easier for the novel reviewer: menus and cross-links permit a fast and complete description of the new sequence.

Another recent goal has been to retrieve new sequences, after the public opening of new sequences sources. In particular, releases of new genome sequences mainly from the DOE Joint Genome Institute (Tuskan *et al.*, 2006; Merchant *et al.*, 2007; Rensing *et al.*, 2008) and large library of ESTs allowed the coverage of the database to be greatly increased. Currently, PeroxiBase possesses more than 6000 peroxidases sequences coming from 900 organisms and distributed among 64 protein classes. The classification we have followed is in agreement with the phylogenetic tree proposed by Baldauf (2003). Peroxidase-encoding sequences are now well represented in each of the five groups described [Prokaryotes, Viridiplantae, Fungus, Animals (opisthokonts), and other eukaryotes (such as Alveolates, Amoebozoa, Excavate, Rhizaria, Stramenopiles (heterokonts)] (Table 1). As the database was initially devoted to the class III peroxidases from plants, of course the major part of sequences (about 3500) is still associated with the 'plants' group. However, recent efforts have been strongly focused on collecting data concerning poorly represented organisms within Viridiplantae but also in other taxonomic groups. A relevant benefit of this update is that multiple alignments may be built up from more closely related sequences, thus making broader phylogenetic studies possible. In addition, the continuous recruitment of sequences in the same group (e.g. Brassicacea: *Brassica napus*, *B. oleracea*, and *B. rapa*) will

allow reliable prediction of the duplication processes occurring within one class in a single organism.

In addition to sequences and annotations, several tools are available to classify and analyse peroxidases: The new button 'Search' with multicriteria, permits specific sequences to be found in the database by using known information such as the cellular localization and possible repressors. On the other hand, a comparison between a query sequence and the peroxidases stored in PeroxiBase is possible by performing a BLASTP and/or a BLASTX search, available in the 'BLAST' section. Search results as well as Blast hits can be retrieved in Fasta format, easy to export for evolution analysis via external bioinformatics software (Fig. 1). Eventually, 'FingerPrintscan' and 'PeroxiScan' help to classify a query sequence in the right group. The former associates the sequence to the corresponding family (Scordis *et al.*, 1999), whilst the latter uses an innovative method of profile design to identify the peroxidase class of an unknown sequence.

PeroxiBase: accuracy and completeness of data

Since its birth, PeroxiBase distinguished itself for the accuracy of data: new entries are manually annotated and always undergo a double check by two researchers. Until entries are checked, they are put in 'limbo' (the 'Pending peroxidase' section), before being uploaded in the database and thus being available for the users. The need to put each new sequence in the right set of peroxidases necessitated the creation of a 'technical sheet' describing each single class. These pages are arranged in the following way: (i) the name of the class (or group) followed by the abbreviation used in the database, (ii) a series of links to entries of other databases describing the class features in terms of patterns and motifs, and (iii) a brief summary of the class-specific features and their detection in the various organisms. Eventually (iv), a list of publications is reported. The 'technical sheets' are carefully verified by highly competent researchers in the field and are updated when there are new relevant findings.

The completeness of the data is continuously verified at three distinct levels: single sequence, organism, and taxonomic group. The sequences stored in the database are complete or partial. The partial ones are fragments, often deriving from ESTs, but recognizable as peroxidases thanks to characteristic motives displayed in the known part(s) of sequence. Periodic data mining is performed for searching complete sequences in the available databases in order to replace the fragments stored in PeroxiBase. An additional check point for the single sequence consists in looking for sequencing errors and frameshifts. Overlaps between sequences, relative to the same gene but coming from different sources, contribute to being able to improve the quality of the sequences.

The second level of completeness concerns the expansion of the number of sequences for each key organism. This occurs by undertaking two directions: the increase of sequences number within a given class and the expansion of the number of classes. Both objectives are achieved by extensive Blasting in a specific database (EST library or genomic project).

A wider analysis consists of collecting data about an additional species member for a specific subclass, order or family. In particular, relevant attention is given to plants whose genome has been recently sequenced or whose EST library is large. Other than providing new interesting facets about the evolution of peroxidases, a rapid annotation of enzymes from newly sequenced organisms offers an unambiguous annotation, able to prevent possible misunderstandings for researchers working in the field. This aspect is important, especially if the number of members within the same class is amazingly high, such as for class III peroxidases and glutathione peroxidases.

Conclusions

PeroxiBase was born with the ambitious aim of collecting data about peroxidases and to attract different groups working on this subject. The effort of researchers from several universities and institutes for the improvement of the database, seems to confirm the importance and the usefulness of this website. Nevertheless, new updates have to be done to make PeroxiBase more and more competitive and to increase the organism coverage for a better and more comprehensive analysis of the peroxidase evolution. Indeed, the high conservation of their sequence and the high rate of duplication allow peroxidases to be used as a powerful evolutionary marker. The next challenge will be the insertion of new useful tools to analyse and to assemble the collected sequences. In addition, the 3D visualization of the known resolved structures of peroxidases could integrate comparisons between sequences and among models. It is hoped that other research teams will join our work and offer their competence and efforts to complete and expand PeroxiBase.

Acknowledgements

We thank Nenad Bakalovic, Laurent Falquet, and Vassilios Ioannidis for their efforts in the initial development of the PeroxiBase database, as well as the Swiss Institute of Bioinformatics for web hosting. We are also indebted to Amos Bairoch and his team for cross-referencing PeroxiBase entries in UniProt Knowledgebase, and Alessandro Greco for critical reading of the manuscript. The financial support of the Swiss National Science Foundation (grant 31-068003.02) to CD is gratefully acknowledged.

References

Bakalovic N, Passardi F, Ioannidis V, Cosio C, Penel C, Falquet L, Dunand C. 2006. PeroxiBase: a class III plant peroxidase database. *Phytochemistry* **67,** 534–539.

Baldauf SL. 2003. The deep roots of eukaryotes. *Science* **300,** 1703–1706.

Liszkay A, Kenk B, Schopfer P. 2003. Evidence for the involvement of cell wall peroxidase in the generation of hydroxyl radicals mediating extension growth. *Planta* **217,** 658–667.

Margis R, Dunand C, Teixeira FK, Margis-Pinheiro M. 2008. Glutathione peroxidase family: an evolutionary overview. *FEBS Journal* **275,** 3959–3970.

Merchant SS, Prochnik SE, Vallon O, *et al*. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318,** 245–250.

Passardi F, Bakalovic N, Teixeira FK, Pinheiro-Margis M, Penel C, Dunand C. 2007*a*. Prokaryotic origins of the peroxidase superfamily and organellar-mediated transmission to eukaryotes. *Genomic* **89,** 567–579.

Passardi F, Cosio C, Penel C, Dunand C. 2005. Peroxidases have more functions than a Swiss army knife. *Plant Cell Reports* **24,** 255–265.

Passardi F, Longet D, Penel C, Dunand C. 2004. The class III peroxidase multigenic family in rice and its evolution in land plants. *Phytochemistry* **65,** 1879–1893.

Passardi F, Theiler G, Zamocky M, *et al*. 2007*b*. PeroxiBase: the peroxidase database. *Phytochemistry* **68,** 1605–1611.

Rensing SA, Lang D, Zimmer AD, *et al*. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319,** 64–69.

Scordis P, Flower DR, Attwood TK. 1999. FingerPRINTScan: intelligent searching of the PRINTS motif database. *Bioinformatics* **15,** 799–806.

Tuskan GA, Difazio S, Jansson S, *et al*. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313,** 1596–1604.

Zamocky M, Furtmuller PG, Obinger C. 2008. Evolution of catalases from bacteria to humans. *Antioxidants and Redox Signaling* **10,** 1527–1548.