Available online at www.sciencedirect.com
SCIENCE DIRECT®
Phytochemistry 65 (2004) 1879–1893
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

The class III peroxidase multigenic family in rice and its evolution in land plants ☆,☆☆

Filippo Passardi [a], David Longet [b], Claude Penel [a], Christophe Dunand [a,*]

[a] Laboratory of Plant Physiology, University of Geneva, Quai Ernest-Ansermet 30, CH-1211 Geneva 4, Switzerland
[b] Department of Zoology and Animal Biology, University of Geneva, Quai Ernest-Ansermet 30, CH-1211 Geneva 4, Switzerland

Received 16 February 2004; received in revised form 11 June 2004
Available online 20 July 2004

Abstract

Plant peroxidases (class III peroxidases, E.C. 1.11.1.7) are secreted glycoproteins known to be involved in the mechanism of cell elongation, in cell wall construction and differentiation, and in the defense against pathogens. They usually form large multigenic families in angiosperms. The recent completion of rice (*Oryza sativa japonica* c.v. Nipponbare) genome sequencing allowed drawing up the full inventory of the genes encoding class III peroxidases in this plant. We found 138 peroxidase genes distributed among the 12 rice chromosomes. In contrast to several other gene families studied so far, peroxidase genes are twice as numerous in rice as in *Arabidopsis*. This large number of genes results from various duplication events that were tentatively traced back using a phylogenetic tree based on the alignment of conserved amino acid sequences. We also searched for peroxidase encoding genes in the major phyla of plant kingdom. In addition to gymnosperms and angiosperms, sequences were found in liverworts, mosses and ferns, but not in unicellular green algae. Two rice and one *Arabidopsis* peroxidase genes appeared to be rather close to the only known sequence from the liverwort *Marchantia polymorpha*. The possible relationship of these peroxidases with the putative ancestor of peroxidase genes is discussed, as well as the connection between the development of the class III peroxidase multigenic family and the emergence of the first land plants.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: *Arabidopsis thaliana*; Rice; Multigenic family; Evolution; Duplication; Laccase; Phytochrome; Peroxidase ancestor

1. Introduction

Plant peroxidases (EC 1.11.1.7), often designated as class III peroxidases (Welinder, 1992a,b), are heme-containing proteins generally encoded by large numbers of paralogous genes. *Arabidopsis* genome, known to harbor a high number of multigenic families (Arabidopsis Genome Initiative, 2000), contains 73 peroxidase genes (Tognolli et al., 2002; Welinder et al., 2002). Almost every living organism contains peroxidases (donor: hydrogen peroxide oxidoreductase) that catalyze the reduction of H_2O_2 by taking electrons to various donor molecules. In the case of class III plant peroxidases, the donor molecules can be phenolics, lignin precursors, or secondary metabolites. Plant peroxidases can also oxidize the growth hormone auxin, as well as other substrates (Gaspar et al., 1982) and produce H_2O_2 (Blee et al., 2001) and hydroxyl radicals (Chen and Schopfer, 1999), two activated oxygen species involved in oxidative burst and in cell elongation (Bolwell et al., 1998; Joo et al., 2001; Rodriguez et al., 2002; Liszkay et al., 2003). The diversity of the reactions catalyzed by plant peroxidases explains the implication of these proteins in a broad range of physiological processes, such as auxin metabolism, lignin and suberin formation, cross linking of cell wall components, defense against pathogens or cell elongation (Penel et al., 1992; Hiraga et al., 2001). The plant peroxidase protein sequence is characterized by the presence of highly conserved amino acids, such as

☆ Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2004.06.023.
☆☆ All rice peroxidase sequences are being submitted to the EMBL database and will be released in case of publication.
* Corresponding author. Tel.: +41-22-379-3012; fax: +41-22-379-3017.
E-mail address: christophe.dunand@bota.unige.ch (C. Dunand).

0031-9422/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2004.06.023

two histidine residues interacting with the heme (distal and proximal histidines) and eight cysteine residues forming disulfide bridges (Fig. 1). The distal histidine is necessary for the catalytic activity. These histidine residues are present in all known heme-containing peroxidase sequences. Plant peroxidase proteins exhibit a constant molecular organization based on the presence of 13 α-helices, 10 of these helices being already present in prokaryotic peroxidases (Welinder, 1992a,b; Gajhede et al., 1997).

Multigenic families originate from gene duplications resulting from different mechanisms (Zhang, 2003): unequal crossing-over, various transposition events, duplication of large chromosome segments or polyploïdization. One of the transposition events, the retrotransposition of cDNA, is characterized by the loss of all introns and regulatory sequences and by a random insertion within the genome. In this case, duplicated genes are usually not transcribed and rapidly degenerate into pseudogenes (Casacuberta and Santiago, 2003). On the other hand, direct transposition of genomic sequences (without an RNA intermediate) creates new genes with higher chances to retain their functionality. Miniature inverted-repeat transposable elements (MITE) belong to this category. These sequences are not able to transpose autonomously. Their replication is probably mediated by a *trans*-acting transposase that recognizes specific sequences present on every MITE called terminal inverted repeats (Casacuberta and Santiago, 2003). MITEs are extremely frequent in rice genome and are generally found in gene-rich regions (Mao et al., 2000). Finally, segmental, chromosomal or whole genome duplications are relatively frequent in plants and are not a source of tandem repeats (Skrabanek and Wolfe, 1998; Blanc et al., 2000; Gebhardt et al., 2003; Zhang, 2003).

It seems difficult to estimate the rate and the conservation of gene duplication without considering the impact of external factors such as the natural or human selections, the chromosomal location and the gene function. Lynch and Conery (2000), have estimated that the rate of duplication is one per gene and per 100 million years (MY) in eukaryotes. A duplicated gene, as every gene, is exposed to nucleotide substitutions at an estimated rate of 0.1 to 0.5 substitution per site and per 100 MY (Graur and Li, 1999). Due to this substitution rate, the usual fate for duplicated genes is to become pseudogenes and to be finally deleted after numerous mutations, insertions and deletions, unless it leads to a selective advantage. In that case, plant will preserve it from degenerative mutations. The retention of a duplicated gene occurs for example when its presence is beneficial for the plant because an extra amount of the encoded protein is profitable for a given biological mechanism. Such cases are known as functional redundancy and are exemplified by the multiple copies of rRNA genes or chloroplast and mitochondria entire genomes (Gillham, 1994). A modification of the duplicated gene expression (subfunctionalization) can also explain its persistence. Lastly, a duplicated gene can acquire a new function (neofunctionalization): this is the most important outcome of gene duplication and can explain the presence of large multigenic families, each paralog becoming specialized for a specific task (Zhang, 2003). The number of retained duplications and of pseudogenes gives information on the gene stability, the gene family evolution, and the importance of the protein function in a given organism.

In September 2003, the Gramene website (Ware et al., 2002), which collects all data concerning grass genomes and particularly rice data from the International Rice Genome Sequencing Project (IRGSP), announced the completion of the whole rice genome sequencing. Consequently, the entire genomic sequences of two Angiosperms, *Arabidopsis* (a Eudicotyledon) and rice (a Monocotyledon) became available for comparison. Both genomes differ in several aspects: *Arabidopsis* genome is composed of five chromosomes, 125 Mbp and more than 26,000 genes, whereas rice genome is nearly four times larger with 12 chromosomes, 420–466 Mbp and probably twice more putative genes (Delseny, 2003; Schoof and Karlowski, 2003). Comparison of the large peroxidase multigenic family in the two plants provided the opportunity to better understand the evolution of these two genomes that diverged from a common ancestor about 150 million years ago, according to one molecular clock approach (Wikstrom et al., 2001). In addition, the completed genome sequencing of rice and

L----Y---**C**P--E-[-V---V-----]------[--A-L-**R**L-F**HDC**]FV-[**GCDAS**VLL]----(--)------[EK]-(-)A-PN--

-S[LRG](--)[F---D--K---E]--(--)**C**P-[**VSCADI**]-[A-AARD-V-]--GGP---V-[**G**R-**DG**-]----

------[LP-P----](-)[--L---F]---GL(-)[---D-V-**LS**GA**H**T-**G**---]-[C--]F--R-----(+1 to 13 a.a.)-----A--

L---C----(+1 to 15 a.a.)----D--TP--[FDN-YY-------GL--S**D**--][-L]---------V---A-----[F---F---]

[MVKMG]-----TG-[-GEI**R**-](-)-[**C**--]-N

Fig. 1. Rice peroxidase consensus amino acid sequence. The 60% consensus sequence has been obtained from the alignment of the 138 OsPrx protein sequences. The boxes represent the residues used for the phylogenetic analysis. The bolded amino acids correspond to a consensus of 100%. Hyphens stand for variable amino acids. Bracketed hyphens represent additional amino acids present exceptionally in a few peroxidases. Bracketed numbers indicate the size range of variable regions.

Arabidopsis and the existence of many EST projects allowed performing an exhaustive analysis of peroxidase genes in *Oryza sativa* ssp. *japonica* and a comparative study of the class III peroxidases among green plants. We found that the rice peroxidase family is composed of 138 genes and 14 pseudogenes, dispersed within the genome. Due to this dispersion, the members of the peroxidase family can be used as milestones to study gene duplications. The intron/exon structure shows different patterns than in *Arabidopsis* with a high intron size variability. The presence of genes encoding class III peroxidases has been observed in every branch of the green plants, except algae, suggesting that ancestor of class III peroxidases appeared around 470 MY ago together with the emergence of the liverworts or other primitive land plants (Kenrick and Crane, 1997).

2. Results

2.1. Rice peroxidase genes and their orthologs in Arabidopsis

The complete genome sequence obtained from the Rice Genome Project (RGP) allowed drawing up the exhaustive list of the peroxidase genes in rice. We identified 138 genes and 14 pseudogenes in the annotated and unannotated BAC sequences (table available online). Rice peroxidases are named hereafter OsPrx, followed by a number. *OsPrx1* is the first peroxidase gene on chromosome I, and so on following the order of the BACs on chromosomes, until *OsPrx138*, the last peroxidase gene on chromosome XII. Automated annotation of rice genome was mainly based on sequence comparisons with the *Arabidopsis* genes (Schoof and Karlowski, 2003). Despite this large and detailed information, numerous BAC sequences were not well annotated by automated programs such as FGenesh, Genscan, RiceHMM, NetGene2, GlimmerR and GeneMark.hmm (Schoof and Karlowski, 2003). In a few cases, these gene-predicting programs did not detect peroxidase sequences. We looked for the presence of peroxidase genes in the unannotated BACs with the programs FGenesh and Genscan, and checked by visual analysis the putative sequences found in order to correct possible prediction mistakes. We also verified every annotated BAC, which allowed correcting predicted peroxidase sequences as well as finding sequences that had been missed in the BAC by automated annotation methods. EST sequences were also searched for in the RGP website (http://rgp.dna.affrc.go.jp). Surprisingly, only about one third of the inventoried peroxidase genes showed registered cDNA sequences in the NCBI EST database. However, when searching in another EST database (Hiraga et al., 2000), we found a few more peroxidases, thus increasing the total expressed peroxidase number to 42%. This expression level remains low compared to *Arabidopsis* (82%; Tognolli et al., 2002). It seems very unlikely that so many functional genes have been preserved but not expressed. The apparent low level of peroxidase gene expression is probably due to the absence of cDNA libraries produced from plants grown in conditions inducing peroxidase gene expression (pathogen interaction, biotic and abiotic stresses). On the other hand, the amount of EST sequences (266,000, http://www.ncbi.nlm.nih.gov) is certainly still too low for the number of genes, considering that many EST sequences are overlapping and hence represent only one gene.

The predicted protein sequences encoded by the peroxidase genes identified in rice genome contain the two histidine and eight cysteine residues characteristic of class III peroxidases (Fig. 1). A 60% consensus protein sequence obtained after the alignment of the 138 predicted OsPrx shows conserved residues and domains also found in *Arabidopsis* peroxidases (AtPrxs), such as FHDC, VSCAD, GAHT or GEIR, the first and the third sequence, respectively, containing the two essential histidines. Similar regions are also typically found with a high conservation in peroxidases from many species, for instance horseradish, barley, tomato, wheat and cucumber (Welinder, 1992a,b). Because our aim was to describe the different peroxidase sequences and families found in rice and also to deduce their possible phylogenetic relationships, only homologous amino acids positions had to be conserved for phylogenetic inference. Therefore, only boxed sequences and positions shown in Fig. 1 were used to construct the OsPrx distance tree (Fig. 2), thus discarding variable regions that were not clearly aligned among all identified peroxidase sequences.

The distance tree of rice peroxidases (Fig. 2) provides a general overview of this large family. We divided it into distinct groups and subgroups based on observed genetic distances and bootstrap supports. To the exclusion of pseudogenes, groups II and VI are strongly supported by bootstrap values. Larger groups, I, IV and V, were divided into generally well-supported subgroups. In the group V, only one subgroup is well defined. We used the peroxidase sequence of *Marchantia polymorpha* to root the tree. Indeed, it provides the maximal sequence length when compared to other outgroups and it is the only known peroxidase sequence in liverworts. Interestingly, OsPrx73 and, to a lesser extent OsPrx116, appeared as the most basal sequences and did not present a phylogenetic association with any other peroxidases in 95% and 51% of bootstrap replicates, respectively. Further analysis with Maximum-Likelihood (ML) tree (data not shown) supported the basal sequence of OsPrx73, Os Prx116 and most of the subgroups.

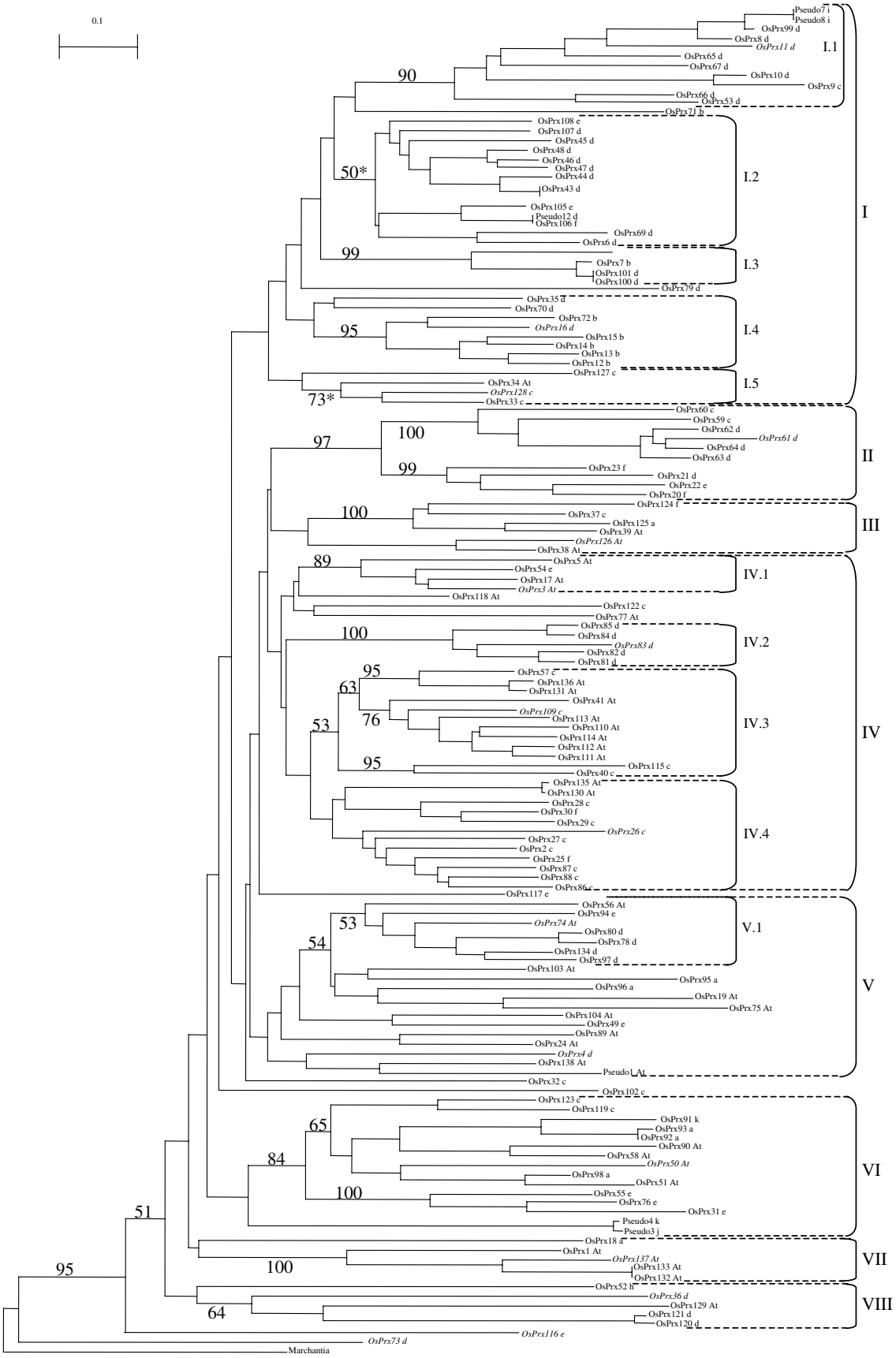
0.1
Pseudo7 i
Pseudo8 i
OsPrx99 d
OsPrx8 d
OsPrx11 d
I.1
OsPrx65 d
OsPrx67 d
OsPrx10 d
OsPrx9 c
OsPrx66 d
OsPrx53 d
OsPrx71 b
90
OsPrx108 e
OsPrx107 d
OsPrx45 d
OsPrx48 d
OsPrx46 d
OsPrx47 d
OsPrx44 d
OsPrx43 d
50*
I.2
OsPrx105 e
Pseudo12 d
OsPrx106 f
OsPrx69 d
OsPrx6 d
I
99
OsPrx7 b
OsPrx101 d
OsPrx100 d
I.3
OsPrx79 d
OsPrx35 d
OsPrx70 d
OsPrx72 b
OsPrx16 d
95
OsPrx15 b
OsPrx14 b
OsPrx13 b
OsPrx12 b
I.4
OsPrx127 c
OsPrx34 At
OsPrx128 c
OsPrx33 c
73*
I.5
OsPrx60 c
OsPrx59 c
OsPrx62 d
OsPrx61 d
100
97
OsPrx64 d
OsPrx63 d
II
99
OsPrx23 f
OsPrx21 d
OsPrx22 e
OsPrx20 f
OsPrx124 f
100
OsPrx37 c
OsPrx125 a
OsPrx39 At
III
OsPrx126 At
OsPrx38 At
89
OsPrx5 At
OsPrx54 e
OsPrx17 At
OsPrx3 At
IV.1
OsPrx118 At
OsPrx122 c
OsPrx77 At
OsPrx85 d
100
OsPrx84 d
OsPrx83 d
OsPrx82 d
OsPrx81 d
IV.2
95
OsPrx57 c
63
OsPrx136 At
OsPrx131 At
OsPrx41 At
OsPrx109 c
76
OsPrx113 At
53
OsPrx110 At
OsPrx114 At
OsPrx112 At
OsPrx111 At
IV.3
95
OsPrx115 c
OsPrx40 c
IV
OsPrx135 At
OsPrx130 At
OsPrx28 c
OsPrx30 f
OsPrx29 c
OsPrx26 c
OsPrx27 c
OsPrx2 c
OsPrx25 f
OsPrx87 c
OsPrx88 c
OsPrx86 c
IV.4
OsPrx117 e
OsPrx56 At
OsPrx94 e
53
OsPrx74 At
OsPrx80 d
OsPrx78 d
54
OsPrx134 d
OsPrx97 d
V.1
OsPrx103 At
OsPrx95 a
OsPrx96 a
OsPrx19 At
OsPrx75 At
V
OsPrx104 At
OsPrx49 e
OsPrx89 At
OsPrx24 At
OsPrx4 d
OsPrx138 At
Pseudo1 At
OsPrx32 c
OsPrx102 c
OsPrx123 c
OsPrx119 c
65
OsPrx91 k
OsPrx93 a
OsPrx92 a
OsPrx90 At
84
OsPrx58 At
OsPrx50 At
OsPrx98 a
OsPrx51 At
100
OsPrx55 e
OsPrx76 e
OsPrx31 e
VI
51
Pseudo4 k
Pseudo3 j
OsPrx18 a
OsPrx1 At
OsPrx137 At
95
100
OsPrx133 At
OsPrx132 At
VII
OsPrx52 h
OsPrx36 d
OsPrx129 At
64
OsPrx121 d
OsPrx120 d
VIII
OsPrx116 e
OsPrx73 d
Marchantia

Fig. 2. Phylogenetic tree of the rice peroxidases based on predicted protein sequences. The tree was constructed by the neighbour-joining method and rooted with the peroxidase sequence of *Marchantia polymorpha*. Values at nodes indicate bootstrap supports greater than 50%. All bootstraps were supported by an ML tree except those marked with an asterisk. All branches are drawn to scale and the scale bar represents 0.1 substitution per site. The letter following each peroxidase stands for its intron pattern (Fig. 7). Italicized peroxidases were used to build the phylogenetic tree shown in Fig. 3.

To provide another evidence of the independent and ancestral origin of OsPrx73 and OsPrx116, a ML tree was inferred from peroxidase cDNA sequences instead of protein sequences. The input data consisted of *OsPrx73*, *OsPrx 116*, and one expressed *OsPrx* (EST RGP database; Hiraga et al., 2000) of each subgroup, as well as expressed peroxidases from the liverwort *M. polymorpha*, the moss *Physcomitrella patens* (PPPrx) and the fern *Ceratopteris richardii* (CrPrx). The ML tree thus obtained (Fig. 3), again suggests that the closest rice homolog to *Marchantia* peroxidase is *OsPrx73*, although the bootstrap support for this relationship is not high. *Ceratopteris* and *Physcomitrella* did not have any close rice homolog, and the branching position of *Osprx116* was not supported enough by a strong bootstrap to be considered as a close rice peroxidase homolog to *Marchantia* peroxidase.

We also checked the codon usage for peroxidases in rice, *Marchantia*, *Physcomitrella* and *Ceratopteris*, by computing their cDNA sequences into the GCUA program (McInerney, 1998). Rice came out with a different codon usage as compared to the other plants (data not shown). Neither *OsPrx73* nor *OsPrx116* showed any significant codon usage deviation from the other *OsPrx*s. A closer look to the cDNA alignment used to build the tree of Fig. 3 showed that variation of sequence between *OsPrx73* and *Marchantia* peroxidase mainly affected the third position of codons.

We further combined OsPrx and AtPrx protein sequences in another phylogenetic tree (data not shown) in order to study the relations between the various peroxidases groups in the two plants. It appeared that rice subgroups I.1, I.2, IV.2 and V.1 had no closely related AtPrx proteins. Therefore, they may correspond to rice or Monocotyledon specific groups. Moreover, when we aligned all OsPrx groups with a peroxidase sequence of *Spirodela polyrrhiza* (NCBI Accession No. Z22920, Chaloupkova and Smart, 1994), a basal Monocotyledon (Judd et al., 2002), group V.1 came out as the one with the highest homology to this protein. In order to look

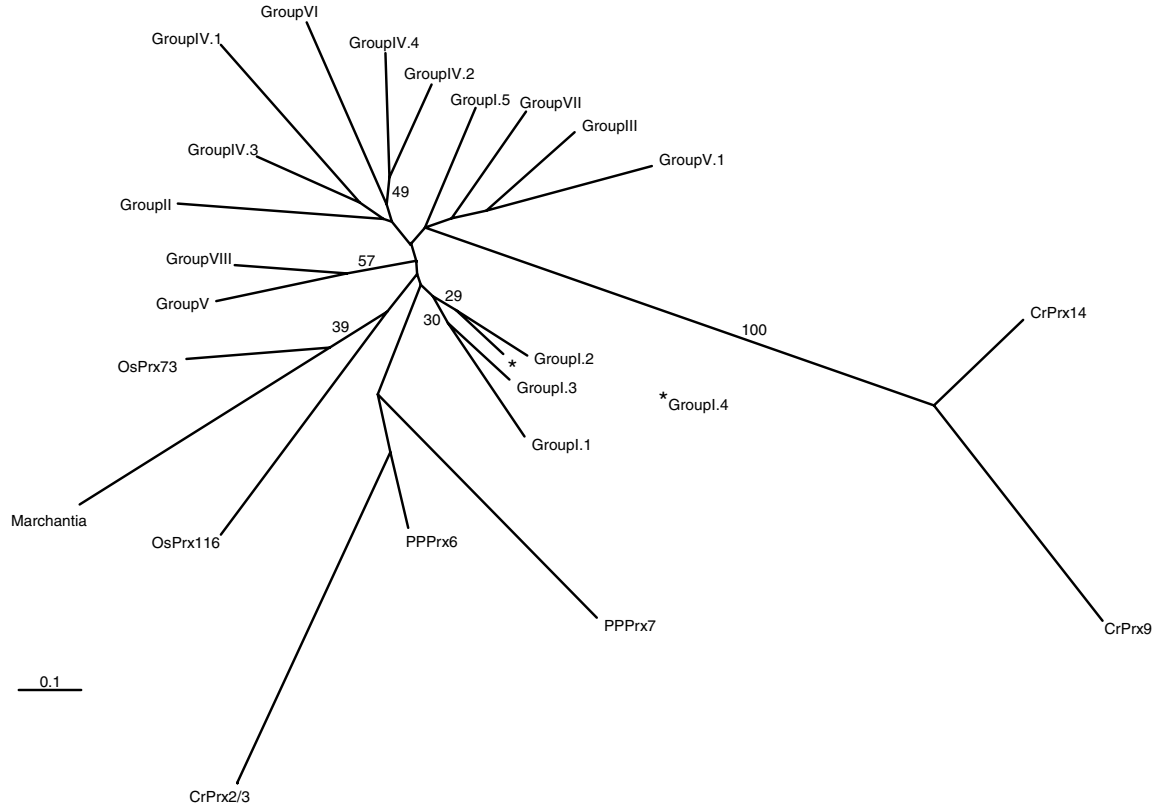

Fig. 3. Unrooted phylogenetic tree of one expressed member of each *Osprx* group and complete peroxidase sequences from the fern *C. richardii*, the liverwort *M. polymorpha* and the moss *P. patens*. The tree is based on cDNA sequences and was constructed by the maximum-likelihood method (see text). Values at nodes indicate bootstrap supports greater than 20%. All branches are drawn to scale and the scale bar represents 0.1 substitution per site.

for a possible rice or Monocotyledon specific motif in peroxidases, we performed a ClustalW alignment including four OsPrx protein sequences from group V.1, four other OsPrx sequences from various groups having AtPrx homologs, as well as four randomly chosen AtPrx and sequences from *Spirodela* and *Hordeum vulgare* (barley). We found a 16-amino acid sequence common to members of group V.I, situated between two highly conserved domains near the signal peptide (Fig. 4). Though less conserved, *Spirodela* sequence obtained a higher homology score with each group V.I peroxidase than with any other OsPrx or AtPrx sequence. The barley peroxidase included in the alignment did not show significant homology with group V.I peroxidases. Barley genome probably contains a large number of peroxidase genes. This means that more peroxidase genes must be identified in this species as well as in other Monocotyledons to confirm the existence of a Monocotyledon-specific motif in class III peroxidases. Finally, it should be mentioned that OsPrx73 and OsPrx116 form a small group of sequences, containing also AtPrx42 known to be constitutively expressed and to have no close paralog in *Arabidopsis* (Tognolli et al., 2002).

2.2. Duplication events among rice peroxidases

Peroxidase genes are not homogeneously distributed among the 12 chromosomes. There are several large clusters of closely homologous genes, most likely resulting from various duplication events (Fig. 5(a)). Six simple tandem duplications can be identified in rice. The most obvious case is *OsPrx100* and *OsPrx101* that are completely identical and probably result from a very recent tandem duplication. Rice genome also shows larger duplications involving 3–6 peroxidase genes. For example, the relationships linking the members of the cluster *OsPrx59* to *OsPrx64* (Fig. 2) strongly suggest that they result from five successive tandem duplications. Furthermore, the presence of introns in all the genes of the cluster excludes retrotransposition events, hence suggesting several rounds of crossing-overs in this region as source of the cluster. Interestingly, the same region of chromosome IV containing this cluster has also been sequenced in *O. sativa* ssp. *indica* c.v. Guangluai 4 (GLA4) in order to contribute to the ab initio annotation of *O. sativa* ssp. *japonica* c.v. Nipponbare (Feng et al., 2002). The BACs corresponding to this region contain *OsPrx59* to *OsPrx64* disposed in a colinear arrangement in GLA4 (respectively, in BACs OSJN00076 and AJ245900), excepted for *OsPrx62*, which was absent. Moreover, the distances between *OsPrx60* and *OsPrx61*, *OsPrx63* and *OsPrx64* as well as *OsPrx61* and *OsPrx63* were much larger in Nipponbare than in GLA4 cultivars. The two first discrepancies are due to large gag-pol polyprotein encoding sequence insertions in Nipponbare cultivar. The OsPrx tree (Fig. 2) reveals that *OsPrx62* is clearly older than *OsPrx61* and *OsPrx64*. These observations show that cluster *OsPrx59* to *OsPrx64* existed before the separation of *indica* and *japonica* subspecies and that, consequently, the difference concerning *OsPrx62* was due to a deletion of this peroxidase in GLA4, rather than to the formation of a new gene in Nipponbare. Another large cluster, *OsPrx42* to *OsPrx48*, is associated with many MITE ele-

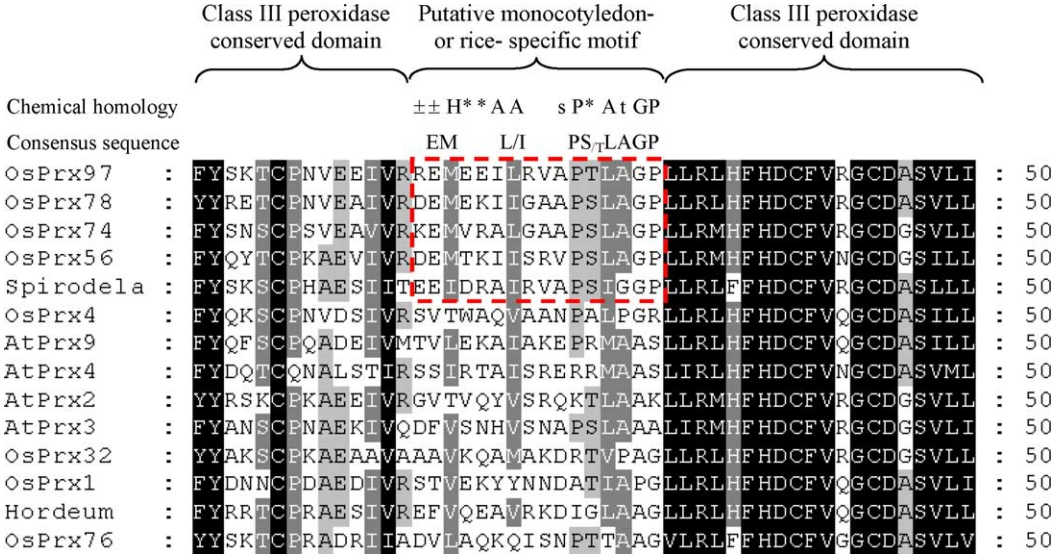

Fig. 4. ClustalW alignment of a peroxidase 50 amino acid sequences located close to the signal peptide sequence. OsPrx 56, 74, 78 and 97 belong to rice-specific subgroup V.1. AtPrx and the remaining OsPrx sequences were chosen randomly. Two typical class III peroxidase conserved domains are shown. The putative rice specific motif is surrounded by a dashed line. Consensus sequences are indicated for this latter motif. One is a chemical homology sequence, whose symbols represent the following categories of amino acids: ±, charged; ∗, polar; H, hydrophobic; A, aliphatic; s, small, t, tiny; P, proline; G, glycine. The second is an amino acid sequence consensus.

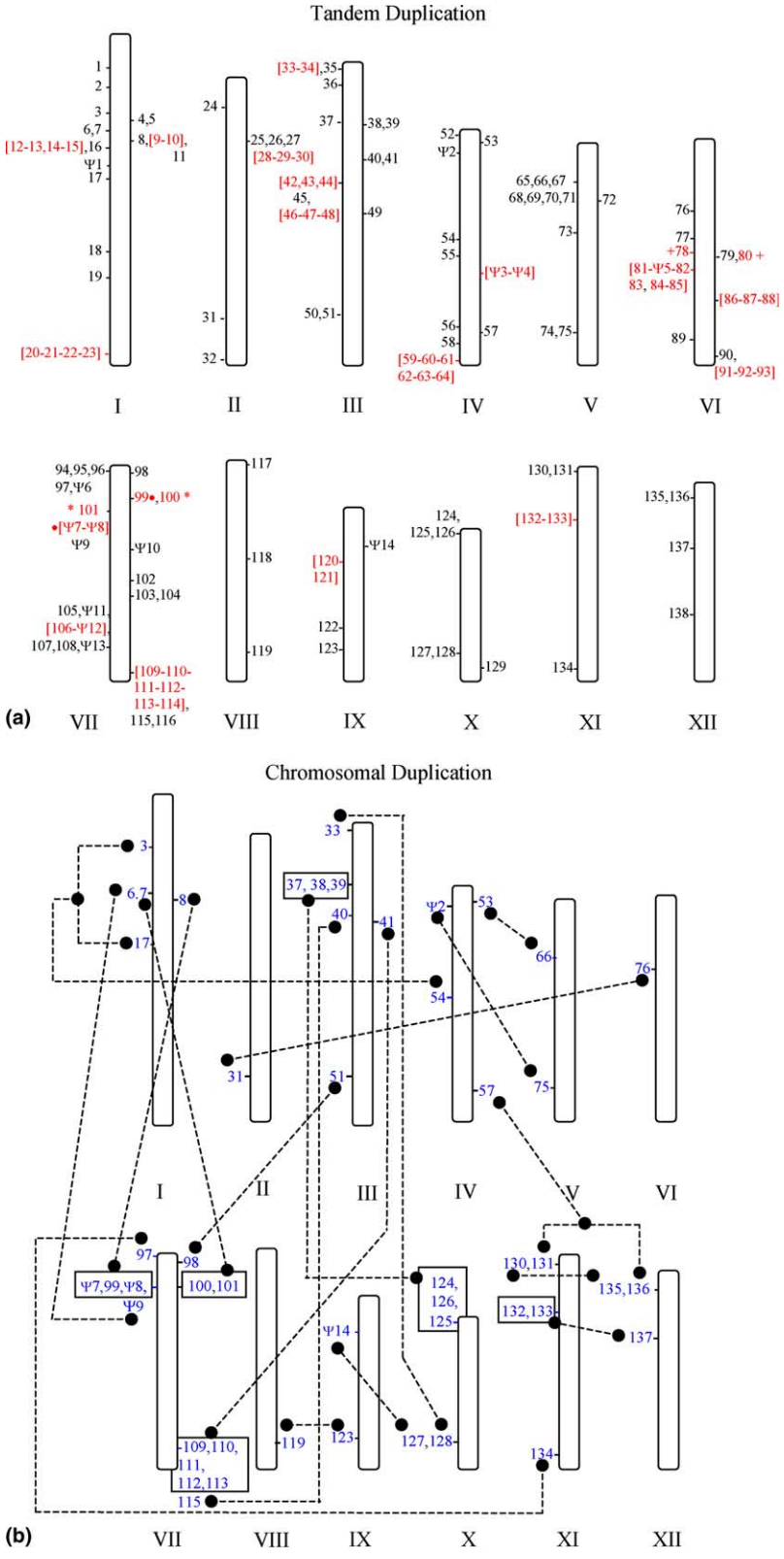

Fig. 5. Distribution of the peroxidase loci on the 12 rice chromosomes. (a) Position of the 138 genes and 14 pseudogenes. Bracketed peroxidase numbers in red, as well as similar symbols (+, ∗ or •) correspond to duplicated genes. (b) Position of the peroxidase genes involved in segmental duplications. Duplicated peroxidases or peroxidase clusters are linked by a dotted line.

ments, located between peroxidase sequences, but none of them was found within a coding region.

Fig. 6 shows a case of double segmental duplication in chromosome VII (Fig. 6). Three large repeated fragments (two of 44,000 bp and one of 32,000 bp) present in BACs AP005409 and AP005454 contain three peroxidase genes (*OsPrx99*, *OsPrx100* and *OsPrx101*) and two pseudogenes (*Ψ7* and *Ψ8*). Other undefined homologous sequences have been used to position correctly the repeated regions. These three segments are surrounded by several *gag-pol* sequences and *RIRE* orf elements.

Gene duplications often result in plants from the duplication of chromosomal segments. Seven chromosomal duplications involving peroxidase genes were found in *Arabidopsis* (Tognolli et al., 2002). The homologies observed in the OsPrx phylogenetic tree (Fig. 2) suggest that more than twenty inter-chromosomal duplications occurred in rice (Fig. 5(b)). Some chromosomes, such as II, VI and VIII, have only one gene exhibiting a close homology with a gene on another chromosome. This may indicate that they are more recent, and therefore have had fewer possibilities to share duplications with older chromosomes. Chromosomes XI and XII look closely related due to two duplication events involving *OsPrx130–OsPrx131* and *OsPrx135–OsPrx136*, as well as *OsPrx132–OsPrx133* and *OsPrx137*. Homologies were also found between chromosomes III and X. The comparison of entire BAC sequences in a "BLAST2 sequences" alignment (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html) revealed that the regions encompassing *Osprx33* to *OsPrx34* and *OsPrx37* to *OsPrx39* exhibited a very strong homology with two regions on chromosome X, respectively, around *OsPrx127* to *OsPrx128* and *OsPrx124* to *OsPrx126*. The phylogenetic tree shown in Fig. 2 does not allow to determine if chromosome X is younger than chromosome III. Similarities in segmental regions between chromosomes III and VII may further support the idea that chromosomes VII and X originated from chromosome III. Confirming this assertion of an ancestral chromosome containing the initial information, phytochrome family genes, which play a key role in plant photomorphogenesis (Fankhauser, 2001), were uniquely found on chromosome III.

2.3. Intronic structure of rice and Arabidopsis peroxidase genes

Forty eight of the 72 peroxidase encoding genes in *Arabidopsis* consist of 3 introns and 4 exons (Tognolli et al., 2002). This pattern, named hereafter *At*, is also well represented in rice (Fig. 7). *O. sativa* ssp. *japonica* has three main intronic patterns: *At* (27%), *c* (17%) and *d* (31%). Intron models *a* to *f* either lack one, two, or all three introns present in *At* model. The last five intron

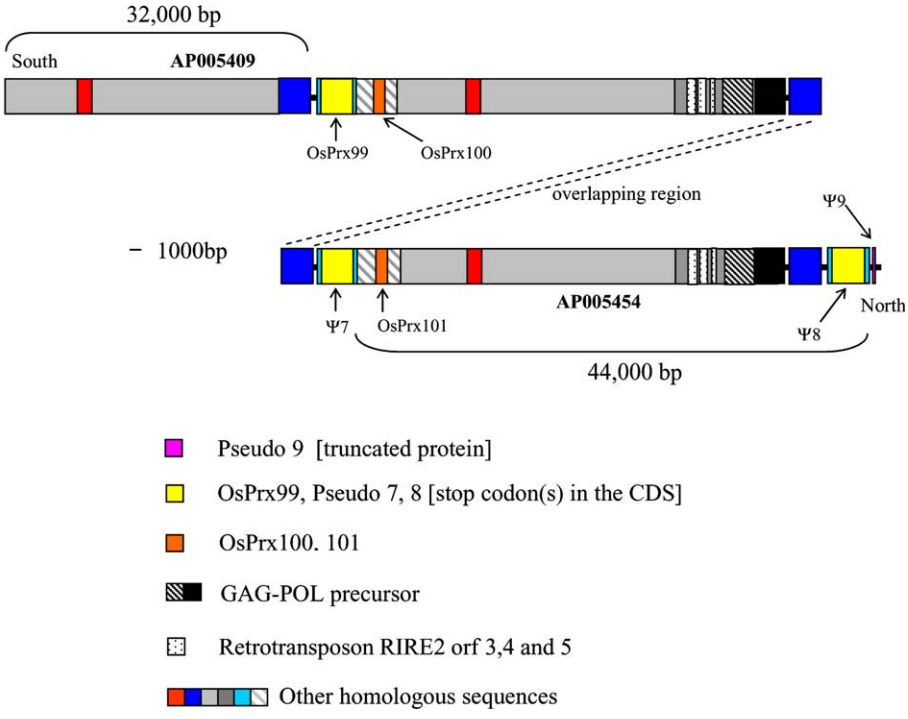

Fig. 6. Structure of AP005454 and AP005409, two contiguous BACs on chromosome VII with an overlapping region of about 2800 bp. Similar patterns denote homologous sequences. Lines represent unique sequences.

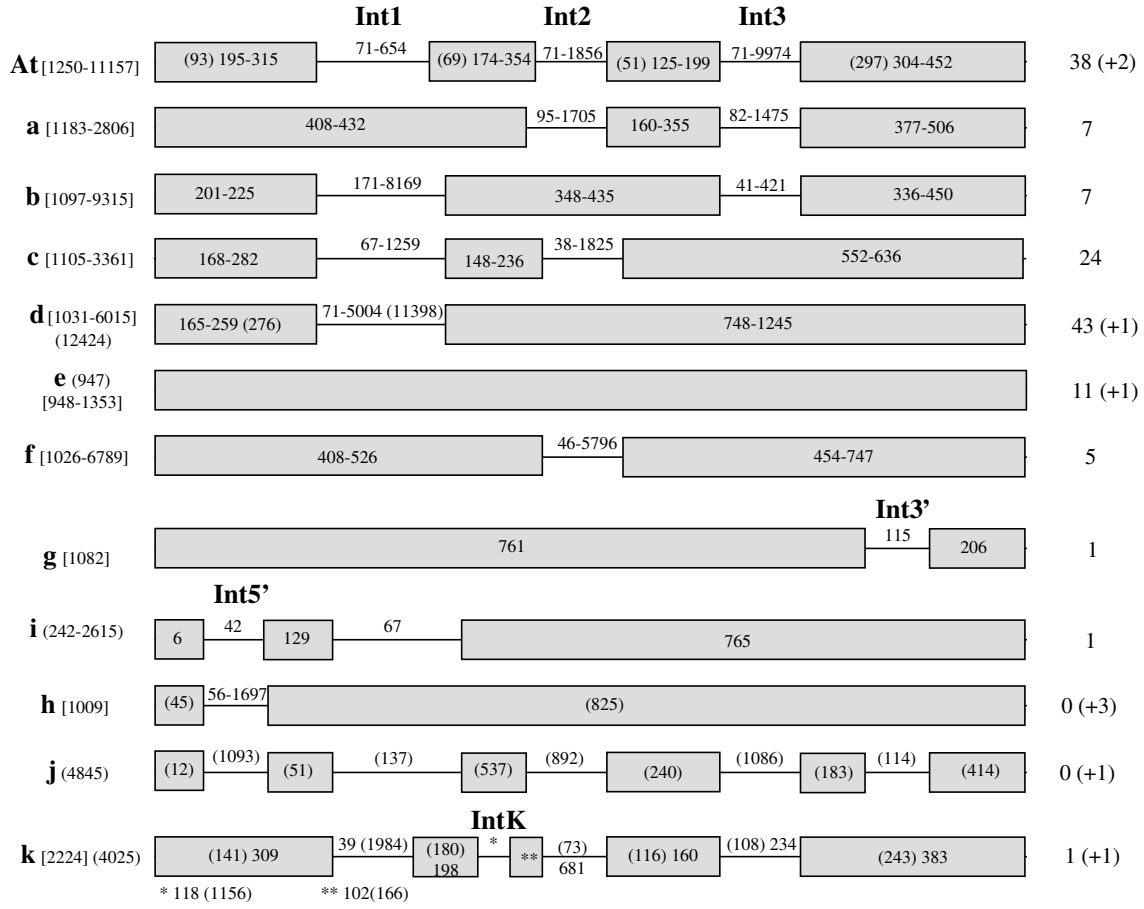

Fig. 7. Structure of rice peroxidase genes. Intron patterns have been identified according to the nomenclature used for *Arabidopsis* peroxidase genes (Tognolli et al., 2002). Grey boxes represent exons, and lines introns. Numbers above each intron/exon region or in front of each gene type indicate the range of the base pairs length. When out of range, values belonging to pseudogenes are written into brackets. Numbers of peroxidase genes in each family are indicated on the right, as well as numbers of pseudogenes in brackets.

patterns (*g* to *k*) are found only in few genes or in pseudogene sequences. They are characterized by different intron positions, that were named *Int5'*, *Int3'*s and *IntK*. In contrast to intron patterns *At* and *a* to *e*, patterns *f* to *k* are not present in *Arabidopsis thaliana*. Rice introns (38–9974 bp) exhibited a broader size range than *Arabidopsis* (71–2850 bp). Surprisingly, several genes were expressed, despite their long intronic sequences or their unusual intronic pattern (Table 1, RGP EST database). *OsPrx89*, *OsPrx124* genes and *Ψ12* were characterized by very large introns containing retrovirus-related *gag-pol* sequences. An EST was detected for *OsPrx89*, but not for *OsPrx124*.

The intron patterns are often conserved within duplicated genes, without conservation of the intronic sequence. The distribution of the intron patterns in the phylogenetic tree (Fig. 2) shows that gene type *e* (no introns) can generate only other genes without intron (type *e*), in line with the hypothesis of a natural intron deletion process described for catalase genes evolution (Frugoli et al., 1998; Iwamoto et al., 1998). Moreover, model *At* and *c* are frequently conserved after duplication, for example in group IV.

2.4. Origin and expansion of class III peroxidases

Class III peroxidases are present in all plants and form large multigenic families (Gaspar et al., 1982; Penel et al., 1992; Welinder, 1992a,b). The development of several genome and EST projects is continuously bringing new elements to confirm this fact. Typical peroxidase sequences were found in numerous Mono- and Eudicotyledonous genomes and in Gymnosperms (Fig. 8), with at least 60% of homology to *Arabidopsis* peroxidases. For example, an EST search in NCBI shows that many peroxidases are present, among others, in tomato (*Lycopersicon esculentum*), spinach (*Spinacia oleracea*), tobacco (*Nicotiana tabacum*) and two species of pine trees (*Pinus sylvestris*, *Pinus pinaster*). But the presence of peroxidases is not restricted to Gymno- and Angiosperms. They have been found as EST sequences in other groups. We identified eight peroxidase

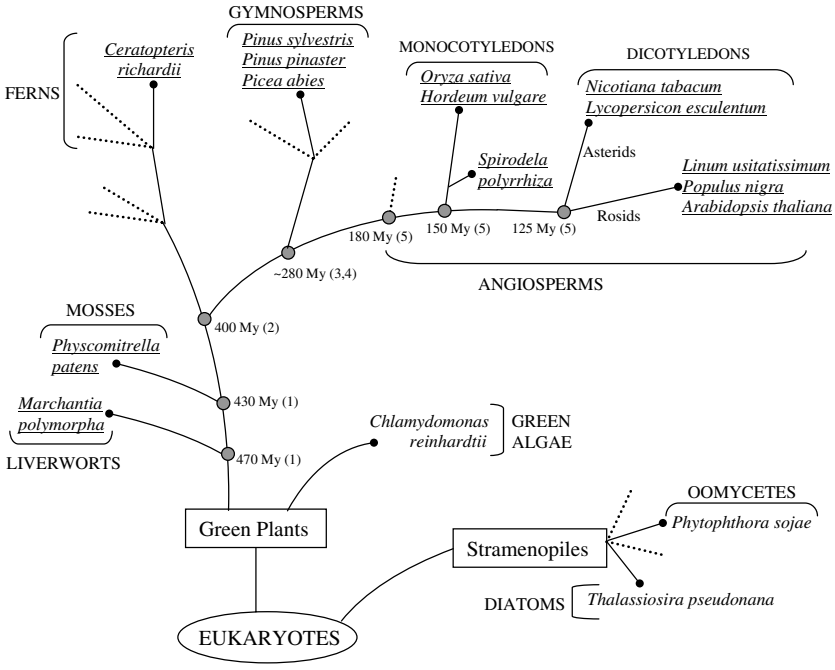

Fig. 8. Plant taxonomic tree. The branches are not drawn to scale. The underlined species contain at least one sequence coding for a class III peroxidase. Many more species were found in databases within the groups listed on the taxonomic tree, excepted for groups with one species shown. Dotted lines are used for groups in which no peroxidase was found, probably due to lack of data (at least in terrestrial green plants). (1) (Kenrick and Crane, 1997); (2) (Pryer et al., 2001); (3) (Savard et al., 1994); (4) (Schmidt and Schneider-Poetsch, 2002); (5) (Wikstrom et al., 2001).

sequences in *C. richardii* (fern) and one in *M. polymorpha* (liverwort) within small EST databases (3600 and 1400 EST, respectively, http://www.plantgdb.org). In *P. patens* (moss), we found 11 to 14 different putative non-overlapping peroxidase sequences in the PEP EST database (www.moss.leeds.ac.uk). Importantly, this EST database has been obtained from two different culture conditions, one of them being an auxin-rich environment. It is known that auxin increases the expression of peroxidase genes (Lee et al., 2002). Since several *Arabidopsis* peroxidase promoters contain auxin-inducible elements, as found through a PLACE search (http://www.dna.affrc.go.jp/htdocs/PLACE/; Penel et al., 2002), it can be hypothesized that the majority of *Physcomitrella* peroxidase transcripts were present in the cDNA library. On the other hand, the entirely sequenced genome of the green alga *Chlamydomonas reinhardtii* and the draft genome sequence of the stramenopiles *Thalassiosira pseudonana* and *Phytophthora sojae* do not contain class III peroxidase encoding genes nor any laccase homolog (secreted proteins), though both algae and stramenopiles genomes contain catalase (cytosolic localization) and cytochrome *C* peroxidase (mitochondrial) sequences. A further search within a cyanobacterium genome database containing sequence information on seven different cyanobacterium species (www.kazusa.or.jp/cyano/) did not retrieve any class III peroxidase sequence. None were found either in other prokaryotic organism through a TBLASTN search in the NCBI website (http://www.ncbi.nlm.nih.gov). Guaiacol oxidation in the presence of H_2O_2, which is a common method used to detect class III peroxidase activity (Greppin et al., 1986), was tested on *C. reinhardtii* and four other algae: *Zygnema* sp., *Staurastrum lacustris*, *Pediastrum boryanum* and *Chloromonas* sp. *Zygnema* and *S. lacustris* belong to *Streptophyta*, which is the Division that contains all the terrestrial plants. The two last ones are members of the *Chlorophyceae* class, like *Chlamydomonas*, which is issued from the Division of *Chlorophyta*. None of these algae showed any guaiacol oxidation in the presence of H_2O_2 (data not shown).

3. Discussion

3.1. Evolution and duplication events

The main goal of this work was to compare the peroxidase multigenic family in rice and in *Arabidopsis*. It appeared that rice genome contains twice as many peroxidase genes as *Arabidopsis*. Other reports demonstrated so far a one to one ratio for several multigenic families (Iwamoto et al., 1998; Baumberger et al., 2003; Baxter et al., 2003; Lijavetzky et al., 2003). In parallel to our search for peroxidases, we also looked for the phytochrome kinase substrate (PKS) family (data not shown). It was found to be composed of 4 genes in

Arabidopsis and only 3 paralogs in *O. sativa* ssp. *japonica*. However, a recent report revealed that expansins are also encoded by twice the number of genes in rice than in *Arabidopsis* (Li et al., 2003). The total number of genes in rice is still hypothetical. Predictions range from 32,000 to 62,000, that is from a 1:1 to a 2:1 ratio compared to *Arabidopsis* (Delseny, 2003; Schoof and Karlowski, 2003). However, the current opinion tends to bet on a number of genes closer to 32,000 than to 62,000 (Delseny, 2003), because the latter number was based on chromosome 1 gene density (Sasaki et al., 2002), that turned out to be higher than the average of the other chromosomes (Chen et al., 2002). A 1:1 ratio in the number of genes within a family would then be more likely expected between rice and *Arabidopsis*. Interestingly, expansin and peroxidase gene families that show both a 2:1 ratio encode cell wall associated proteins. They may have followed a similar gene expansion during evolution in connection with their role in cell elongation and cell wall mechanical properties.

Monocotyledons (rice) branched off from Eudicotyledons (*Arabidopsis*) some 150 MY ago, based on a NPRS molecular clock approach (Wikstrom et al., 2001), although fossil-based methods tend to place this divergence more recently, around 120 MY ago (Magallon et al., 1999). It can be assumed that, at their origin, the first Monocotyledons and Eudicotyledons had a similar number of peroxidase genes. The evolution leading to today's *Arabidopsis* and rice yielded a quite different number of peroxidase genes in the two plants. This difference could result from a better conservation of duplicated genes in rice, maybe because they met some needs in herbaceous Monocotyledonous plants, while they were less useful in Eudicotyledons. Additionally, a polyploidy event in rice may explain the twofold increase in the number of peroxidase genes. This difference also indicates that the estimation of a duplication rate of one gene per 100 million years in eukaryotes proposed by Lynch and Conery (2000) seems a little low in view of our findings.

Rice is under mankind selection since 9000 years (Khush, 1997). This could have favored the conservation of some peroxidase encoding genes. Human influence may also be responsible for differences in peroxidases numbers between the two subspecies of rice, *O. sativa indica* (GLA4 c.v.) and *japonica* (Nipponbare c.v.), that diverged from a common ancestor only one MY ago (Khush, 1997). In contrast to Angiosperms, mosses, exemplified by *P. patens*, contain apparently a rather small number of peroxidase genes. This difference could be explained by the greater level of organ and tissue differentiation in Angiosperms. The presence in these plants of molecules polymerized by peroxidases such as lignin and suberin and the fact that mosses generally live near the soil, in a rather humid environment, while Angiosperms became adapted to a wider range of conditions and reached larger sizes could also play a role. All these considerations are based on the assumption that the rate of gene duplication was the same in the various phyla. Other events such as MITE-mediated duplication or still undefined mechanisms could have also played some role in the differential enlargement of peroxidase multigenic family among the various phyla.

3.2. Implication of retrotransposons in peroxidase gene duplication

MITE elements are supposed to have been generated from autonomous MITE precursors through a two-step process (Feschotte et al., 2002): every time that a precursor excises itself, it creates double-stranded breaks that are repaired by an internally-deleted copy of the transposon, which becomes non-autonomous due to loss of its transposase-encoding region. This deleterious mechanism is named abortive repair. This non-autonomous transposon generates a large number of the so-called MITE elements, probably through a "cut and paste" mechanism, using the same transposase that recognizes specific repeats on the MITE sequence. The "cut and paste" process generates double-stranded breaks, which are believed to create recombination regions on chromosomes (Shalev and Levy, 1997). This series of events may be responsible for the formation of MITE-containing gene clusters such as *OsPrx42* to *OsPrx48* through numerous recombination rounds. Such multiple gene duplications may then confer selective advantages to the plant and the resulting clusters would hence be retained. For instance, peroxidases play a role in plant resistance against pathogens (Moerschbacher, 1992). This could be a possible function for the cluster *OsPrx42* to *OsPrx48*, in which recombinations helped plant to widen its pathogen resistance range. Interestingly, the members of this cluster do not have any close homologs in *Arabidopsis*. They may have evolved to adapt the defense of rice against new pathogens different from the *Arabidopsis* pathogens. On the other hand, the presence of retrotransposable elements such as *RIRE* and *gag-pol* sequences in a cluster of peroxidases could argue in favor of duplication processes correlated with retrotransposition events.

3.3. OsPrx introns

Introns in rice peroxidase genes show large size variability (Fig. 7) and, in a few cases, they can reach a considerable length. Rice splicing machinery has the capacity to deal with these large introns, some of them containing even internal retroviral-related polyprotein coding sequences. We may then assume that rice introns can play a role of viral genomic insertions "buffers", protecting the plant against gene inactivation pheno

mena. However, this defense system against insertions is apparently not always efficient, since truncated proteins may be produced or tissue-specific expression may be altered by the presence of very large introns (Casacuberta and Santiago, 2003).

Intron distribution within rice peroxidase genes is quite variable. The evolution does not seem to have preserved a major intron organization pattern, even between paralogous genes. As the *At* intronic model is found in both rice and *Arabidopsis* in a significant proportion, it may well reflect the ancestral intronic model. Loss of introns would then have occurred in rice to create most of the other intronic models observed, excepted in rare cases such as models *g* to *k* (Fig. 7), in which new introns appeared at unusual positions. The insertion of new introns in models *i* and *j* seems to have been deleterious, since these models are found only in pseudogenes. This hypothesis of the loss of intron during the evolution would be in line with previous studies on catalase genes in *Poaceae* (Iwamoto et al., 1998), but definitely needs more sequencing data of primitive organisms to be confirmed. The putative ancient origin of model *At* and the hypothesis of an evolutionary intron loss converge to the conclusion that *At* would be the ancestor model in Angiosperms. In this respect, a particularly interesting plant would be *Amborella trichopoda*, supposedly the sister group of all Angiosperms (Qiu et al., 1999), whose lineage appeared around 180 MY ago (Wikstrom et al., 2001) and whose EST sequences listed in NCBI do not contain any peroxidase so far. Several key questions arise from these observations, such as why not only one intronic pattern has been preserved in rice during evolution and how organisms managed to reorganize their gene sequences. One justification could be the transcriptional control and the RNA stabilization by intronic sequences (Gonzalez et al., 2002).

3.4. Searching for the ancestral peroxidase

Despite a high duplication rate of the encoding genes, the crucial peroxidase specific motifs and residues were conserved throughout the evolution from liverworts and mosses to rice and *Arabidopsis*. This is the case of two conserved domains, $LxRLxxHDC_2xxxGC_3DxS$ and $LxxxHxxGxxxC_6$ (Fig. 1), necessary for the interaction with the heme moiety through histidine residues. Disulfide bridges involving cysteines C_2, C_3 and C_6 have also important functions in catalytic cleavage of hydrogen peroxide (Poulos and Kraut, 1980). The two histidine residues are the signature of all the hemo-peroxidases and are conserved in every organism (Zamocky et al., 2000). A third conserved domain, C_4xxVSC_5xD, has an unknown function. The other cysteine residues involved in disulfide bridges (C_1, C_7 and C_8) are also preserved. These conserved motifs were very useful not only to search for class III peroxidases, but also to correct rice predicted OsPrx sequences. Variable regions of the protein can also be very informative. A 16-amino acid domain, found in rice subgroup V.1, showed significant homology with a domain of *Spirodela* peroxidase. As *Spirodela* is considered as a basal Monocotyledon, it may have preserved ancient sequences that should be found in a major number of Monocotyledons. Therefore, this motif might also be considered as Monocotyledon-specific. The importance of this putative sequence would be either structural or functional.

The fact that class III peroxidases have been found in all the major divisions of land plants but not in green algae denotes the functional importance of these proteins and points to the existence of a common peroxidase ancestor 470 million years ago, before or during the emergence of liverworts (Kenrick and Crane, 1997), the very first land plants (Wellman et al., 2003). Nevertheless, it will be necessary to wait for sequencing of other green algae and stramenopiles species to definitively confirm the exclusive presence of peroxidases in terrestrial green plants.

Distance (Fig. 2) and maximum-likelihood (Fig. 3) phylogenetic analyses converge towards an interesting point: OsPrx73 is the closest peroxidase rice homolog to *Marchantia* peroxidase. Although this conclusion is supported by a relatively low bootstrap value in Fig. 3, the congruence observed between both analyses is strengthened by the fact that two very different data sources, such as protein (Fig. 2) and cDNA (Fig. 3), tend to the same conclusion. Moreover, codon usage analysis showed that *OsPrx73* has not preserved the codon usage of *Marchantia* peroxidase, but has instead followed the general OsPrx usage. Therefore, evolution has successfully maintained OsPrx73 amino acid sequence despite the modification of its corresponding DNA sequence, thanks to codon third position flexibility. The role of OsPrx73 in plant is unknown, but it could be assumed that it has been preserved because of an important function possibly existing also in *Marchantia*. Why is OsPrx73 so different from the other OsPrx, and why is it so similar to a *Marchantia* peroxidase? Convergent evolution may be an explanation, though very improbable, as rice and liverworts do not live in the same biotopes, and are morphologically very dissimilar. A second possibility is that OsPrx73 represents the closest rice sequence to the class III peroxidase ancestor, which probably appeared around 470 MY ago. Interestingly, *AtPrx42* is also not a member of any *Arabidopsis* peroxidase subgroup (Tognolli et al., 2002) and is homolog to *OsPrx73* and *OsPrx116*. Moreover, it is strongly expressed in all organs of *Arabidopsis*, which may indicate an ancient regulation system.

The apparition and the diversification of class III peroxidases may be related to the adaptation of plant to the life out of water. Two major challenges may have

played a role in the birth of the first class III peroxidase. A radical change in oxygen environment has probably had a strong impact in chemical reaction rates involving oxygen, and thereby in the various concentrations of oxygen containing molecules, such as the peroxidase substrate H_2O_2. As well, there may be a correlation between the appearance of class III peroxidases and the emergence of cell wall structures adapted to terrestrial life. In line with our results on peroxidases, we also searched for laccase homologs in plants and other kingdoms, and only found these enzymes in land plants. Interestingly, laccases are also involved in cell wall construction (O'Malley et al., 1993; Ranocha et al., 1999). Their absence in other organisms than terrestrial plants suggests that both proteins may have appeared at the same time during evolution in order to build novel walls of terrestrial plants.

Class III peroxidases are probably absent from prokaryotes, which contain class I peroxidases. As both classes have strikingly conserved a general similar structure as well as two critical histidine residues involved in the interaction with heme (Welinder, 1992a,b; Zamocky et al., 2000), class III peroxidases may well have originated from class I peroxidases. The first land plants in possession of modified class I peroxidases able to deal with changes such as increased H_2O_2 availability and cell wall structure constraints were then probably favored in their progress on land colonization.

Present-day Angiosperm peroxidases catalyze many essential reactions in cell wall, including cross-linking between structural polysaccharides (Fry, 1986), oxidative lignin or suberin polymerization (Ros Barceló, 1997), or formation of activated oxygen species that are involved in the modulation of the mechanical properties of cell wall (Rodriguez et al., 2002; Liszkay et al., 2003). The appearance of these diversified functions during land plant evolution most likely came from neofunctionalization of duplicated peroxidase genes. The formation of new organs, such as stems, leaves and flowers, during green plants evolution also favored the diversification of new peroxidases necessary for the specific needs of these new organs. Many plant peroxidases have been shown to be involved in the responses of plants to a wide range of biotic and abiotic stresses (Castillo, 1992). In this respect, stress factors also have varied during evolution with the colonization of new habitats and the constant emergence of novel plant pathogens. The appearance of insects about 420 MY ago, for instance, has been a major threat to plants, which, as suggested by Gaunt and Miles (Gaunt and Miles, 2002), induced ferns to build higher stems to escape herbivory. Ultimately, the development of agriculture subjected plants, like rice, to a strong selective pressure. In order to face these challenges, subfunctionalization and neofunctionalization were critical survival events that allowed plants to evolve.

4. Experimental

4.1. Rice data mining

Arabidopsis class III peroxidases protein sequences have been used as a starting point for the rice data mining. Each of the 73 different amino acid sequences of *Arabidopsis* class III peroxidases was submitted against the whole genomic rice database with a tblastn search on the Rice Genome Project (RGP) website (http://rgp.dna.affrc.go.jp/), that reports genomic sequence information on the Nipponbare cultivar of *O. sativa japonica*. When the BAC sequence was not annotated or contained putative peroxidase sequences, BACs and sequences were analyzed for gene presence with different programs such as FGenesh (http://www.softberry.com/berry.phtml) and GenScan (http://genes.mit.edu/GENSCAN.html). The corresponding coding sequence (CDS) was translated with the "translate" tool on Expasy (http://us.expasy.org/tools/dna.html) and controlled for specific peroxidases motifs. BACs were positioned on the chromosomes following the indications of the Gramene website (http://www.gramene.org/).

In order to find peroxidases in other organisms, various OsPrx protein sequences were used as input sequences in TBLASTN searches within different databases. Peroxidases were sought in the NCBI website (http://www.ncbi.nih.gov/BLAST) for *M. polymorpha*, in the PEP EST database (www.moss.leeds.ac.uk) for *P. patens*, in the PlantGDB database (http://www.plantgdb.org/) for *C. richardii* and in the DOE Joint Genome Institute (JGI) website for *C. reinhardtii*, *T. pseudonana* and *P. sojae* (http://genome.jgi-psf.org/). A similar search was performed for cytochrome *C* peroxidases laccases, catalases, phytochromes, and phytochrome kinase substrates (PKS). *Arabidopsis* sequences drawn from NCBI were used as input sequences.

4.2. Comprehensive phylogenetic analysis of peroxidase sequences identified in O. sativa

All 138 peroxidase sequences present in rice and the single peroxidase sequence of *M. polymorpha*, used as an outgroup, were aligned using Clustal W (Thompson et al., 1994). The alignment was further inspected and visually adjusted. Only unambiguously aligned amino acid positions were retained for analyses, leading to an alignment consisting of 145 sequences and 156 sites. Distance and maximum likelihood analyses were performed with the help of the PHYLIP 3.6a3 package (Felsenstein, 1993). The distance tree was constructed with the NEIGHBOR option under the JTT substitution frequency matrix, and 1000 bootstrap replicates were carried out for this method with the SEQBOOT option. The maximum likelihood tree was inferred with the ProML option, under the JTT substitution

frequency matrix, by using the global rearrangement option. Maximum likelihood bootstrap analysis was not performed due to computational limitations. The Njplot software was used to visualize phylogenetic trees.

Putative Monocotyledon-specific domains were located visually and then realigned with ClustalW. One 16 aa putative sequence gave a significant similarity score (computed with GeneDoc software) for OsPrx group V.I and *Spirodela*. On a 23 aa sequence encompassing the putative 16 amino acids (between highly conserved amino acid sequences IVR and LLRL), all comparisons within group V.I and *Spirodela* were scored below 290 (sum of pairs method) and with a minimal homology of 69%, whereas other OsPrx groups and AtPrx compared with group V.1 and *Spirodela* were scored above and with lower homologies.

4.3. Phylogenetic analysis of expressed peroxidase sequences present in different plant lineages

We also examined the phylogenetic relationships among expressed peroxidases present in rice and in several distantly related plants. We analyzed a restricted data set consisting of 18 expressed (RGP website EST search) *O. sativa* peroxidase sequences representing the global phylogeny, along with 2 peroxidase sequences of the moss *P. patens*, 3 of the fern *C. richardii* and the single sequence of *M. polymorpha*. OsPrx peroxidases chosen for this alignment are shown in italic in the phylogenetic tree of Fig. 2. All 24 amino acid sequences were aligned as explained above, resulting in 80 unambiguously aligned amino acids positions. Because of this limited sequence length and consequently limited phylogenetic signal, we used nucleotide sequences for phylogenetic reconstruction. All 24 cDNA sequences were aligned manually against the 80 unambiguously aligned amino acids, resulting in 240 aligned nucleotides. The limited number of analyzed sequences allowed us to perform maximum likelihood (ML) analyses using PAUP* (Swofford, 1998), and to carry out a bootstrap analysis using 100 replicates. ML trees were inferred from all codon positions. Because preliminary distance analyses suggested that some expressed peroxidases (e.g., CrPrx 14 and 9) are fast evolving sequences, we chose to minimize possible long branch attraction artifacts by taking into account rate variations among sites. A gamma distribution with 8 categories plus invariable sites was therefore added to the model of sequence evolution. All necessary parameters (substitution matrix, nucleotide frequencies, among-site rate variations) were estimated via ML from a starting distance tree. Heuristic ML tree searches were then performed starting from this distance tree under the General-Time-Reversible model of sequence evolution since it allows all six pairs of substitutions to have different rates.

4.4. Codon usage analysis

Codon usage was studied by inserting cDNA sequences of *M. polymorpha*, *P. patens* and *C. richardii* peroxidases, as well as all rice peroxidases into the GCUA program (McInerney, 1998). Codon usage was then calculated for each amino acid and visually analysed.

Acknowledgements

We thank Dr. P. Cuenoud for helpful discussion and critical reading of the manuscript. F.P. is supported by the Swiss National Science Foundation (31-068003.02).

References

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Baumberger, N., Doesseger, B., et al., 2003. Whole-genome comparison of leucine-rich repeat extensins in Arabidopsis and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade. Plant Physiol. 131, 1313–1326.

Baxter, I., Tchieu, J., et al., 2003. Genomic comparison of P-type ATPase ion pumps in Arabidopsis and rice. Plant Physiol. 132, 618–628.

Blanc, G., Barakat, A., et al., 2000. Extensive duplication and reshuffling in the Arabidopsis genome. Plant Cell 12, 1093–1101.

Blee, K.A., Jupe, S.C., et al., 2001. Molecular identification and expression of the peroxidase responsible for the oxidative burst in French bean (*Phaseolus vulgaris* L.) and related members of the gene family. Plant Mol. Biol. 47, 607–620.

Bolwell, G.P., Davies, D.R., et al., 1998. Comparative biochemistry of the oxidative burst produced by rose and french bean cells reveals two distinct mechanisms. Plant Physiol. 116, 1379–1385.

Casacuberta, J.M., Santiago, N., 2003. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311, 1–11.

Castillo, F.J., 1992. Peroxidases and stress. In: Penel, C., Gaspar, T., Greppin, H. (Eds.), Plant Peroxidases. University of Geneva, Switzerland.

Chaloupkova, K., Smart, C.C., 1994. The abscisic acid induction of a novel peroxidase is antagonized by cytokinin in *Spirodela polyrrhiza* L. Plant Physiol. 105, 497–507.

Chen, M., Presting, G., et al., 2002. An integrated physical and genetic map of the rice genome. Plant Cell 14, 537–545.

Chen, S.X., Schopfer, P., 1999. Hydroxyl-radical production in physiological reactions. A novel function of peroxidase. Eur. J. Biochem. 260, 726–735.

Delseny, M., 2003. Towards an accurate sequence of the rice genome. Curr. Opin. Plant Biol. 6, 101–105.

Fankhauser, C., 2001. The phytochromes, a family of red/far-red absorbing photoreceptors. J. Biol. Chem. 276, 11453–11456.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package). Seattle, University of Washington.

Feng, Q., Zhang, Y., et al., 2002. Sequence and analysis of rice chromosome 4. Nature 420, 316–320.

Feschotte, C., Jiang, N., et al., 2002. Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. 3, 329–341.

Frugoli, J.A., McPeek, M.A., et al., 1998. Intron loss and gain during evolution of the catalase gene family in angiosperms. Genetics 149, 355–365.

Fry, S.C., 1986. Polymer-bound phenols as natural substrates of peroxidases. In: Greppin, H., Penel, C., Gaspar, T. (Eds.), Molecular and Physiological Aspects of Plant Peroxidases. University of Geneva, Switzerland, pp. 169–182.

Gajhede, M., Schuller, D.J., et al., 1997. Crystal structure of horseradish peroxidase C at 2.15 Å resolution. Nat. Struct. Biol. 4, 1032–1038.

Gaspar, T., Penel, C., et al., 1982. Peroxidases 1970–1980. University of Geneva, Switzerland.

Gaunt, M.W., Miles, M.A., 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Mol. Biol. Evol. 19, 748–761.

Gebhardt, C., Walkemeier, B., et al., 2003. Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* reveals structurally conserved domains and ancient duplications in the potato genome. Plant J. 34, 529–541.

Gillham, N.W., 1994. Organelle Genes and Genomes. Oxford University Press, Oxford, GB.

Gonzalez, J.M., Penzes, Z., et al., 2002. Stabilization of a full-length infectious cDNA clone of transmissible gastroenteritis coronavirus by insertion of an intron. J. Virol. 76, 4655–4661.

Graur, D., Li, W.H., 1999. Fundamentals of Molecular Evolution. Sinauer Associates, Sunderland, MA.

Greppin, H., Penel, C., et al., 1986. Molecular and Physiological Aspects of Plant Peroxidases. University of Geneva, Switzerland.

Hiraga, S., Sasaki, K., et al., 2001. A large family of class III plant peroxidases. Plant Cell. Physiol. 42, 462–468.

Hiraga, S., Yamamoto, K., et al., 2000. Diverse expression profiles of 21 rice peroxidase genes. FEBS Lett. 471, 245–250.

Iwamoto, M., Maekawa, M., et al., 1998. Evolutionary relationship of plant catalase genes inferred from exon-intron structures: isozyme divergence after the separation of monocots and dicots. Theor. Appl. Genet. 97, 9–19.

Joo, J.H., Bae, Y.S., et al., 2001. Role of auxin-induced reactive oxygen species in root gravitropism. Plant Physiol. 126, 1055–1060.

Judd, W.S., Campbell, C.S., et al., 2002. Plant Systematics: A Phylogenetic Approach. Sinauer Associates, Sunderland, MA.

Kenrick, P., Crane, P.R., 1997. The origin and early evolution of plants on land. Nature 389, 33–39.

Khush, G.S., 1997. Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. 35, 25–34.

Lee, D.J., Kim, S.S., et al., 2002. The regulation of Korean radish cationic peroxidase promoter by a low ratio of cytokinin to auxin. Plant Sci. 162, 345–353.

Li, Y., Jones, L., et al., 2003. Expansins and cell growth. Curr. Opin. Plant Biol. 6, 603–610.

Lijavetzky, D., Carbonero, P., et al., 2003. Genome-wide comparative phylogenetic analysis of the rice and *Arabidopsis* Dof gene families. BMC Evol. Biol. 3, 17.

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

Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.

Magallon, S., Crane, P.R., et al., 1999. Phylogenetic pattern, diversity, and diversification of eudicots. Ann. Missouri Bot. Garden 86, 297–372.

Mao, L., Wood, T.C., et al., 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res. 10, 982–990.

McInerney, J.O., 1998. GCUA: general codon usage analysis. Bioinformatics 14, 372–373.

Moerschbacher, B.M., 1992. Plant peroxidases: involvement in response to pathogens. In: Penel, C., Gaspar, T., Greppin, H. (Eds.), Plant Peroxidases. University of Geneva, Switzerland.

O'Malley, D.M., Whetten, R., et al., 1993. The role of laccase in lignification. Plant J. 4, 751–757.

Penel, C., Dunand, C., et al., 2002. In search of function for 73 *Arabidopsis* peroxidases genes. In: Acosta, M., Rodriguez-Lopez, J.N., Pedreno, M.A. (Eds.), Plant Peroxidase Biochemistry and Physiology. University of Murcia, pp. 135–141.

Penel, C., Gaspar, T., et al., 1992. Plant Peroxidases. University of Geneva, Switzerland.

Poulos, T.L., Kraut, J., 1980. The stereochemistry of peroxidase catalysis. J. Biol. Chem. 255, 8199–8205.

Pryer, K.M., Schneider, H., et al., 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409, 618–622.

Qiu, Y.L., Lee, J., et al., 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402, 404–407.

Ranocha, P., McDougall, G., et al., 1999. Biochemical characterization, molecular cloning and expression of laccases – a divergent gene family – in poplar. Eur. J. Biochem. 259, 485–495.

Rodriguez, A.A., Grunberg, K.A., et al., 2002. Reactive oxygen species in the elongation zone of maize leaves are necessary for leaf extension. Plant Physiol. 129, 1627–1632.

Ros Barceló, A., 1997. Lignification in plant cell walls. Intern. Rev. Cytol. 176, 87–132.

Sasaki, T., Matsumoto, T., et al., 2002. The genome sequence and structure of rice chromosome 1. Nature 420, 312–316.

Savard, L., Li, P., et al., 1994. Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extant seed plants. Proc. Natl. Acad. Sci. USA 91, 5163–5167.

Schmidt, M., Schneider-Poetsch, H.A., 2002. The evolution of gymnosperms redrawn by phytochrome genes: the Gnetatae appear at the base of the gymnosperms. J. Mol. Evol. 54, 715–724.

Schoof, H., Karlowski, W.M., 2003. Comparison of rice and *Arabidopsis* annotation. Curr. Opin. Plant Biol. 6, 106–112.

Shalev, G., Levy, A.A., 1997. The maize transposable element Ac induces recombination between the donor site and an homologous ectopic sequence. Genetics 146, 1143–1151.

Skrabanek, L., Wolfe, K.H., 1998. Eukaryote genome duplication – where's the evidence? Curr. Opin. Genet. Dev. 8, 694–700.

Swofford, D.L., 1998. PAUP*: phylogenetic analyses under parsimony (*and other methods). Sinauer Associates, Sunderland, MA.

Thompson, J.D., Higgins, D.G., et al., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tognolli, M., Penel, C., et al., 2002. Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana*. Gene 288, 129–138.

Ware, D., Jaiswal, P., et al., 2002. Gramene: a resource for comparative grass genomics. Nucleic Acids Res. 30, 103–105.

Welinder, K.G., 1992a. Plant peroxidases: structure–function relationships. In: Penel, C., Gaspar, T., Greppin, H. (Eds.), Plant Peroxidases. University of Geneva, Switzerland.

Welinder, K.G., 1992b. Superfamily of plant, fungal and bacterial peroxidases. Curr. Opin. Struct. Biol. 2, 388–393.

Welinder, K.G., Justesen, A.F., et al., 2002. Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*. Eur. J. Biochem. 269, 6063–6081.

Wellman, C.H., Osterloff, P.L., et al., 2003. Fragments of the earliest land plants. Nature 425, 282–285.

Wikstrom, N., Savolainen, V., et al., 2001. Evolution of the angiosperms: calibrating the family tree. Proc. R. Soc. Lond. B: Biol. Sci. 268, 2211–2220.

Zamocky, M., Janecek, S., et al., 2000. Common phylogeny of catalase-peroxidases and ascorbate peroxidases. Gene 256, 169–182.

Zhang, J., 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18, 292–298.