The Plant Cell, Vol. 26: 2404–2419, June 2014, www.plantcell.org

RESEARCH ARTICLES

Subcellular Relocalization and Positive Selection Play Key Roles in the Retention of Duplicate Genes of *Populus* Class III Peroxidase Family[W][OPEN]

Lin-Ling Ren,[a,b,1] Yan-Jing Liu,[a,1] Hai-Jing Liu,[a] Ting-Ting Qian,[a] Li-Wang Qi,[c] Xiao-Ru Wang,[d] and Qing-Yin Zeng[a,2]

[a] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Laboratory of Cell Biology, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China
[d] Department of Ecology and Environmental Science, UPSC, Umeå University, SE-90187 Umeå, Sweden

Gene duplication is the primary source of new genes and novel functions. Over the course of evolution, many duplicate genes lose their function and are eventually removed by deletion. However, some duplicates have persisted and evolved diverse functions. A particular challenge is to understand how this diversity arises and whether positive selection plays a role. In this study, we reconstructed the evolutionary history of the class III peroxidase (PRX) genes from the *Populus trichocarpa* genome. PRXs are plant-specific enzymes that play important roles in cell wall metabolism and in response to biotic and abiotic stresses. We found that two large tandem-arrayed clusters of PRXs evolved from an ancestral cell wall type PRX to vacuole type, followed by tandem duplications and subsequent functional specification. Substitution models identified seven positively selected sites in the vacuole PRXs. These positively selected sites showed significant effects on the biochemical functions of the enzymes. We also found that positive selection acts more frequently on residues adjacent to, rather than directly at, a critical active site of the enzyme, and on flexible regions rather than on rigid structural elements of the protein. Our study provides new insights into the adaptive molecular evolution of plant enzyme families.

INTRODUCTION

Gene duplication is an important mechanism for the evolution of novel gene function. The accumulation of whole-genome sequences provides substantial evidence for the abundance of duplicate genes in plant species (Tuskan et al., 2006). Over the course of evolution, many duplicate genes may lose their functions and be ultimately removed by deletion. However, some duplicates have persisted and evolved diverse functions. A particular challenge is to understand the evolutionary mechanisms responsible for the initial retention and subsequent divergence of newly created duplicate genes. The neofunctionalization model has been proposed as an important process driving the retention of duplicated genes by evolving novel functions (Byrne and Wolfe, 2007; Sakuma et al., 2013). However, this model cannot satisfactorily explain how a duplicate gene can escape the load of deleterious mutations that would probably accumulate before enough beneficial mutations could confer a new function (Byun-McKay and Geeta, 2007). Because deleterious mutations are more probable than advantageous mutations after gene duplication, it has been assumed that most duplicate copies are lost and a limited number of duplicates that gain new functions are maintained by positive selection. Positive selection is now recognized as playing an important role in the initial retention and subsequent divergence of duplicate genes (Shiu et al., 2006; Barkman et al., 2007; Beisswanger and Stephan, 2008; Lan et al., 2013). However, the lack of appropriate functional assays has generally hindered attempts to elucidate the molecular mechanisms and effects of positive selection on the retention and divergence of duplicate genes.

Protein subcellular relocalization has recently been considered as another form of functional divergence of duplicate genes (Byun-McKay and Geeta, 2007; Marques et al., 2008; Qian and Zhang, 2009; Wang et al., 2009) based on the following reasoning: (1) different subcellular structures play different biological functions in the eukaryotic cell; (2) most eukaryotic proteins are localized to specific subcellular structures to perform their functions; and (3) proteins could alter their functions when relocalized to a new subcellular compartment due to altered microenvironments. Even the same protein might have a different biological function when localized to a new subcellular structure (Leissring et al., 2004). Thus, changes in protein subcellular localization may lead to the origin of new functions.

Class III peroxidases (PRXs; EC 1.11.1.7) are plant-specific enzymes that can catalyze the reduction of H_2O_2 by moving electrons to various donor molecules such as phenolic compounds, lignin precursors, or secondary metabolites (Cosio and

[1] These authors contributed equally to this work.
[2] Address correspondence to qingyin.zeng@ibcas.ac.cn.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Qing-Yin Zeng (qingyin.zeng@ibcas.ac.cn).
[W] Online version contains Web-only data.
[OPEN] Articles can be viewed online without a subscription.
www.plantcell.org/cgi/doi/10.1105/tpc.114.124750

Dunand, 2009). Plant PRX proteins are involved in several important physiological and developmental processes, including lignin and suberin formation, the cross-linking of cell wall components, wound healing, the removal of H_2O_2, the oxidation of toxic reductants, and defense against pathogen or insect attack (Gabaldón et al., 2005; Bindschedler et al., 2006; Cosio and Dunand, 2009; Daudi et al., 2012). PRXs are encoded by a large gene family. The *Arabidopsis thaliana* and rice (*Oryza sativa*) genomes contain 73 and 138 PRX genes, respectively (Tognolli et al., 2002; Passardi et al., 2004). Based on the absence or presence of the C-terminal signal peptide, plant PRXs are divided into cell wall and vacuole types (Matsui et al., 2011). The wall-type PRXs are involved in cell wall metabolism including lignin polymerization, whereas the vacuole-type PRXs are involved in defense responses against abiotic and biotic stresses (Ostergaard et al., 2000; Li et al., 2003; Gabaldón et al., 2005; Bindschedler et al., 2006; Costa et al., 2008). This functional divergence between cell wall and vacuole PRXs makes this gene family a model system for exploring the contribution of protein subcellular relocalization to the expansion of the gene family.

Populus trichocarpa is a model woody perennial plant with a recently sequenced genome (Tuskan et al., 2006). Unlike herbaceous plants such as *Arabidopsis* and rice, *Populus* shows tree-specific traits such as wood formation and perennial growth in the field. Some of these traits might be related to the functional evolution of the PRX gene family. Analysis of the *Populus* genome suggests a recent occurrence of a whole-genome duplication event in the Salicaceae 60 to 65 million years ago that affected ~92% of the genome (Tuskan et al., 2006). In addition, *Populus* and *Arabidopsis* share more ancient large-scale duplication events (Tuskan et al., 2006). The complex history of genome duplications and chromosomal rearrangements in *Populus* provides an opportunity to study the molecular mechanisms of duplicate gene retention over the course of genome evolution. In this study, we conducted genome-wide annotation of PRX genes in the *Populus* genome and identified the genetic events responsible for their expansion and organization. Using ancestral protein reconstruction, protein subcellular localization, statistical tests for positive selection, site-directed mutagenesis, and enzyme functional assays of ancestral, current, and mutant proteins, this study revealed that protein subcellular relocalization and positive selection likely have played important roles in the retention of duplicate genes.

RESULTS

Identification of the PRX Gene Family in the *Populus* Genome

Ninety-three full-length genes encoding PRX proteins were identified in the *Populus* genome (Supplemental Data Set 1). Among these, 10 (*PRX5*, *8*, *18*, *32*, *38*, *47*, *60*, *66*, *69*, and *76*) were considered putative pseudogenes based on the presence of a frame shift disrupting the coding region or a stop codon occurring prematurely, resulting in a truncated protein. After removing these stop codons or revising the frame shifts by deleting one or two nucleotides, the 10 full-length sequences were included in the phylogenetic and gene expression analyses. In addition to full-length PRX genes, 43 partial PRX fragments were identified (Supplemental Table 1). The length of these fragments ranged from 24 to 279 amino acids. The conserved domain analysis on these fragments showed that they contained a partial version of the peroxidase domain characteristic of PRX; thus, they were considered to be pseudogenes. We were unable to reliably analyze the phylogenetic relationships of these fragments because of their small size.

Expression of *Populus* PRX Genes under Normal Growth Conditions and Abiotic Stress

We investigated the expression patterns of *Populus* PRX genes under normal growth conditions and in response to stress treatments (H_2O_2, salicylic acid, salt, and drought applications). We examined the expression of all 93 *Populus* PRX genes in five tissues, including roots, shoots, leaves, buds, and phloem (Figure 1B). Among the 93 PRX genes, 41 were expressed under all growth conditions and 44 were selectively expressed either in a specific tissue and/or in response to a specific treatment, whereas eight (*PRX5*, *8*, *14*, *32*, *60*, *68*, *69*, and *76*) were neither detected in any tissue nor in response to any treatment. To validate whether the lack of expression was due to PCR failure, we used the same sets of primers to amplify these eight genes from genomic DNA. We found that all primers produced the expected results. We further designed additional eight pairs of primers, each specific to one of the eight PRX genes (Supplemental Table 2), and performed amplification detection on both mRNA and genomic DNA. Again, we did not detect expression of these PRX genes in any of the tissues examined, but we did observe but positive amplifications from genomic DNA. Thus, we consider that these PRX genes are likely expressed at subdetectable levels, or only induced in response to treatments and/or in tissues not examined in our study, or pseudogenes.

Among the five types of tissue examined, PRX genes showed preferential expression in roots (Supplemental Figure 1). For example, 81 of the 93 PRX genes were expressed in roots under normal growth conditions compared with 63, 52, 61, and 68 genes in shoots, leaves, buds, and phloem, respectively. Under both normal growth and the four stress treatment conditions, 77 genes were expressed in roots, while 52, 46, 56, and 45 genes were expressed in shoots, leaves, buds, and phloem, respectively. This preferential expression of PRX genes in roots was also observed in *Arabidopsis* and rice (Hiraga et al., 2000; Tognolli et al., 2002) and is thought to be associated with their role in defense responses against soil pathogens (Passardi et al., 2005; Hatsugai et al., 2009). In addition, we found 17, 10, 8, and 5 *Populus* PRX genes that were not expressed in leaves, buds, phloem, and shoots under normal growth conditions were expressed in response to one or several stress treatments, which indicates that they may play a role in the response to abiotic stresses.

Chromosomal Distribution of the PRX Gene Family

To understand the duplication and expansion history of the *Populus* PRX family, we examined the distribution of the 93 PRX genes on chromosomes. We found that although each of the 19 *Populus* chromosomes harbored PRX genes, the distribution was obviously heterogeneous (Supplemental Figure 2). Five clusters (clusters I to V) with high densities of PRX genes (more

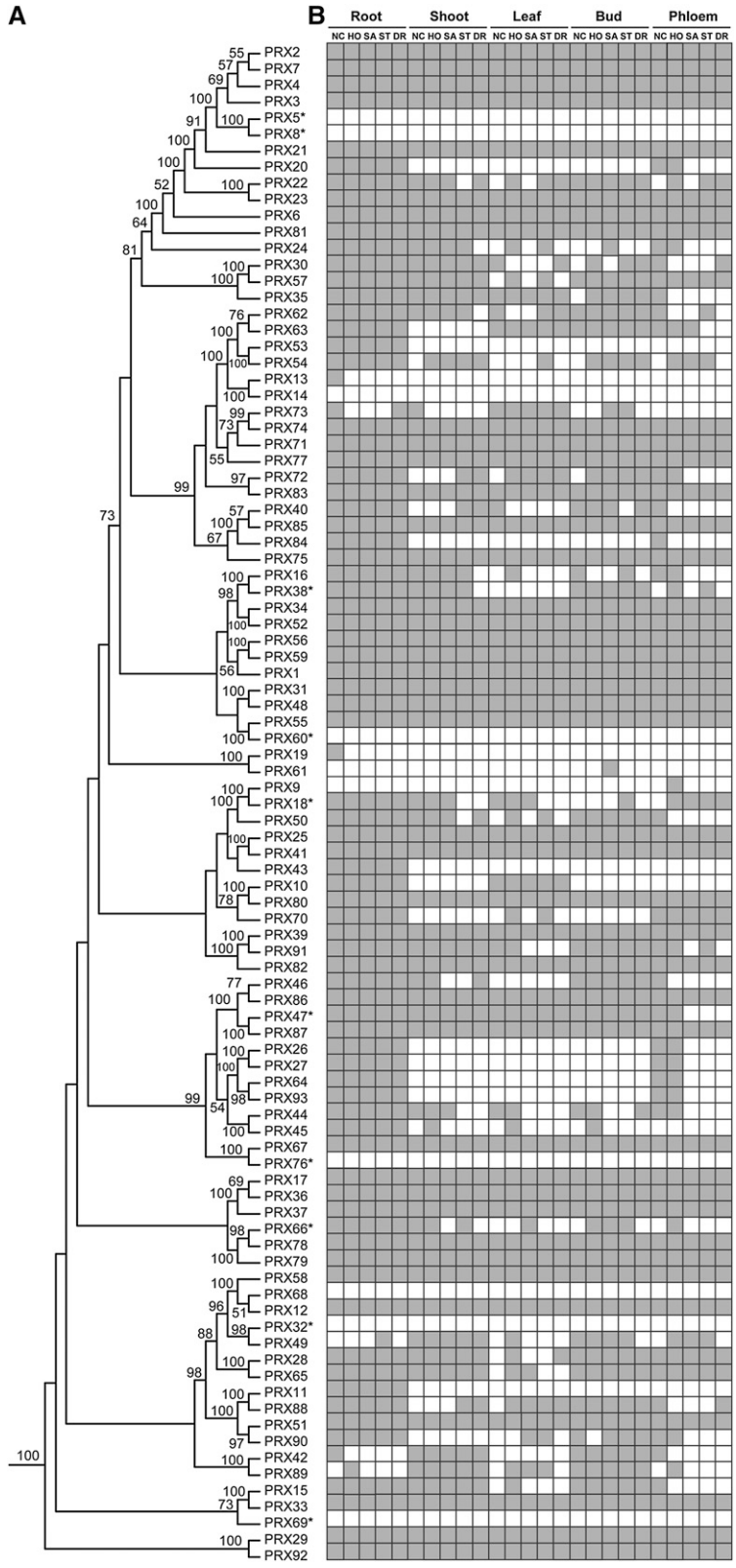


Figure 1. Phylogenetic Relationships and Expression Patterns of *Populus* PRXs.

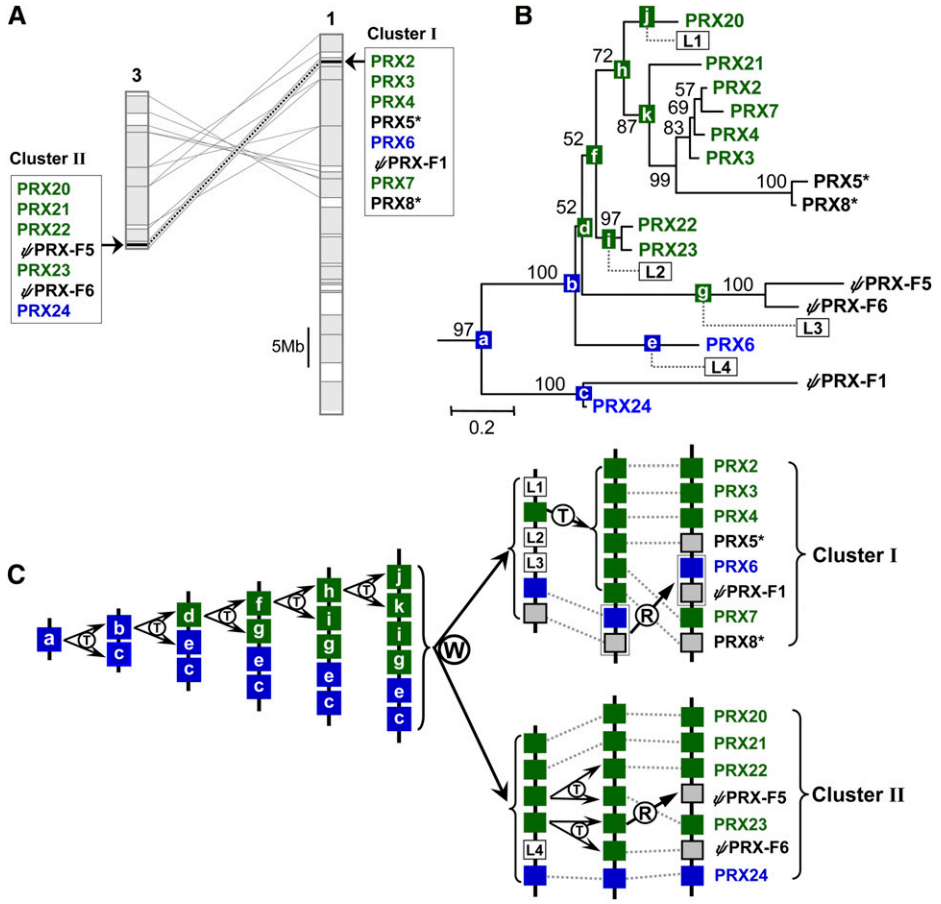


Figure 2. Genomic Localization, Phylogenetic Relationships, and Hypothetical Evolutionary Histories of the *Populus* PRX Genes in Clusters I and II.

In **(A)**, regions that are assumed to correspond to homologous genome blocks are shaded in gray and connected with lines. Paralogous clusters I and II are indicated by dashed lines within the gray-shaded trapezoids. In **(B)**, phylogenetic relationships were reconstructed using the WAG+I+G model. Numbers on branches indicate the bootstrap percentage values calculated from 1000 bootstrap replicates. *Populus* PRX71, 72, and 75 were used as outgroups. In **(C)**, the letters T, W, and R in the schematic diagram showing the hypothetical origins of PRX genes indicate putative tandem duplication, whole-genome duplication, and rearrangements, respectively. Gray and white boxes represent pseudogenes and lost genes, respectively. Blue and green boxes represent PRX genes located in the cell wall and vacuole, respectively. The letters in the blue and green boxes represent ancestral PRX genes. PRX fragments are indicated with ψ. Putative full-length pseudogenes are indicated with asterisks.

than two copies) were observed on chromosomes 1, 3, 7, 13, and 16, in addition to seven pairs (*PRX13/14*, *PRX26/27*, *PRX36/37*, *PRX53/54*, *PRX62/63*, *PRX78/79*, and *PRX86/87*) arranged in tandem on chromosomes 1, 4, 5, 8, 10, 15, and 17. A total of 37 PRX genes were arranged in tandem repeats, indicating that tandem duplications contributed significantly to the expansion of this gene family. The remaining PRX genes were dispersed on all 19 chromosomes. The poplar genome underwent a recent whole-genome duplication ~60 to 65 million years ago (Tuskan et al., 2006). Paralogous segments created by this genome duplication were identified in a previous genome annotation (Tuskan et al., 2006). Based on this information, we identified 14 duplicate gene pairs, including *PRX9/18*, *PRX11/88*, *PRX15/33*, *PRX16/38*, *PRX28/65*, *PRX29/92*, *PRX30/57*, *PRX31/48*, *PRX32/49*, *PRX34/52*, *PRX39/91*, *PRX42/89*, *PRX55/60*, and *PRX56/59*, which were each located in a pair of paralogous blocks and can thus be considered as direct results of this whole-genome duplication event (Supplemental Figure 2).

Figure 1. (continued).

The maximum-likelihood procedure using the WAG+I+G model with 100 bootstrap replicates was used for phylogeny reconstruction. Numbers at each node in the phylogenetic tree are bootstrap values, and only bootstrap values >50% are shown. Putative pseudogenes are indicated with asterisks. In **(B)**, the gray box indicates positive detection of gene expression in the corresponding tissue under normal growth conditions (NC) and following H_2O_2 (HO), salicylic acid (SA), salt (ST), and drought (DR) stress treatments.

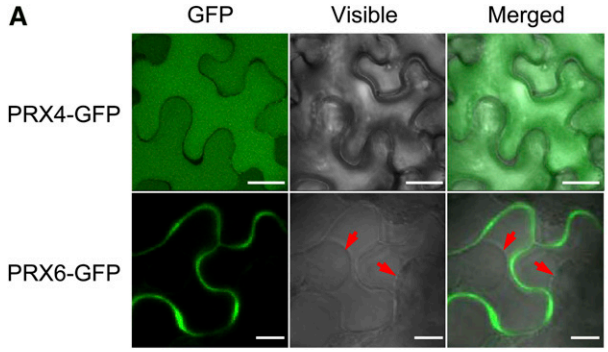
Clusters I and II were the largest of the five tandem duplication clusters, containing seven and five full-length copies, respectively, and three PRX fragments (Figure 2A; Supplemental Figure 2). Among these 12 copies, *PRX5* and *PRX8* were considered putative pseudogenes based on the presence of a frame shift disrupting the coding region and their lack of expression. The other 10 PRX genes in these two clusters were expressed under all growth conditions or in a specific tissue (Figure 1B), suggesting that they were functional. In the following sections, we focused on the retention mechanisms of the PRX genes in these two clusters.

Duplication History of PRX Genes in Clusters I and II

Clusters I and II were each located in a pair of paralogous blocks (Figure 2A); we thus considered them as created by the recent whole-genome duplication event. On the phylogenetic tree of all 93 *Populus* PRX genes, the 12 PRX genes in these two clusters were grouped together (Figure 1A), indicating that they descended from a single common ancestor. Based on the relationships among these copies (Figure 2B), we reconstructed the most parsimonious history of gene duplication, loss, and rearrangement as illustrated in Figure 2C. It seems likely that five rounds of tandem duplications created six ancestral genes (j, k, i, g, e, and c, Figure 2C) before the recent whole-genome duplication, after which this set of genes evolved independently into the organization in clusters I and II, by complex tandem duplication, gene loss, and rearrangement events. If this scenario is valid, we should expect the levels of divergence between paralogs to comport with their hypothetical divergence history. We calculated divergence at synonymous sites (K_S) for all pairwise comparisons of the 10 functional PRX genes in clusters I and II (Supplemental Table 3). We found the lowest average K_S (0.0315) between copies created most recently after the recent whole-genome duplication, 0.4359 for those due to the recent whole genome duplication, and 0.4648 to 1.5264 for duplications more ancient, with the highest average K_S (1.5264) between copies derived from the first and second round of ancestral duplication. These results support the hypothetical history of the PRX genes in these two clusters.

Subcellular Localization of the PRX Proteins in Clusters I and II

Among the 12 PRX genes in clusters I and II, two (*PRX5* and *8*) were considered putative pseudogenes based on the presence of frame shifts disrupting the coding region and their lack of expression. Thus, we only analyzed subcellular localization of the remaining 10 PRX genes. We generated a C-terminal green fluorescent protein (GFP) fusion for each protein and visualized their subcellular location by confocal microscopy after transient expression of the fusions in *Nicotiana benthamiana*. Confocal microscopy analysis showed that eight fusion proteins (PRX2, 3, 4, 7, 20, 21, 22, and 23-GFP) resulted in the accumulation of GFP fluorescence in the vacuole, as illustrated by PRX4-GFP (Figure 3A), indicating that these eight fusion proteins localize to the vacuole. The fluorescent signal of the other two fusion proteins (PRX6 and 24-GFP) were only detected in the peripheral regions of the *N. benthamiana* epidermal cells, suggesting



B

HRP C1a	TGTQGQIRLNCRVVNSNSLLHDMVEVVDFVSSM	353
At3g32980	TGTQGQIRQNCRVVNSNSLLHDVVEIVDFVSSM	352
At3g49110	TGTQGQIRLNCRVVNSNSLLHDVVDIVDFVSSM	354
At3g49120	TGTQGQIRLNCRVVNSNSLLHDVVDIVDFVSSM	353
PRX2	TGTEGEIRLNCRVVNANLAGPDS----MLVSSI	343
PRX3	TGTEGEIRLNCRVVNANLAGPDS----MLVSSI	343
PRX4	TGTEGEIRLNCSVVNANLAGPDS----MLVSSI	343
PRX7	TGTEGEIRLNCRVVNANLAGPDS----MLVSSI	343
PRX20	TGTQGEIRLNCRVVNGNSTGQNI----MLVSSI	348
PRX21	TGTEGEIRLNCRVVNANLAGKDS----VLVSSV	343
PRX22	TGTDGEIRLNCRRVNDNSTGSNA----LLVSSI	349
PRX23	TGTDGEIRLNCRIVN-NSTGSNA----LLVSSI	346
PRX6	TGSNGEIRADCKRVN	318
PRX24	TGSSGEIRKKCSVVNS	326

Figure 3. Subcellular Localization and C-Terminal Amino Acid Sequence Alignment of the Plant PRXs.

In **(A)**, the plasma membrane separated from the cell wall after plasmolysis is indicated by red arrows. Bars = 10 μm. In **(B)**, cell wall and vacuole PRXs are shaded in green and yellow, respectively. Vacuolar sorting C-terminal signal peptides are underlined. The C-terminal hexapeptides conserved among vacuole PRX proteins are shaded in black.

plasma membrane and/or cell wall localization. To distinguish between the plasma membrane and/or cell wall localizations, the epidermal cells were plasmolyzed by treatment with 0.3 g/mL sucrose. After plasmolysis, the fluorescent signal was retained in the cell wall, as illustrated by PRX6-GFP (Figure 3A), indicating that these two PRX-GFP fusion proteins localize to the cell wall. We thus divided these 10 PRX proteins into two types: cell wall PRXs (PRX6 and 24) and vacuole PRXs (PRX2, 3, 4, 7, 20, 21, 22, and 23).

Divergence in Selective Pressure between Clusters I and II PRX Genes

Phylogenetic relationships among PRX genes showed that the eight vacuole *PRX*s (*PRX2*, *3*, *4*, *7*, *20*, *21*, *22*, and *23*) in clusters I and II grouped together with 100% support and the two cell wall–type *PRX6* and *PRX24* diverged earlier (Figure 1A). To determine whether there was a significant change in the selective pressure between the two groups of *PRX*s (clades A and B in Figure 4), we performed maximum likelihood codon models analysis using phylogenetic analysis by maximum likelihood (PAML) package version 4.4 (Yang, 2007). Two assumptions were tested: a one-ratio model that assumed the same ω (= d_N/d_S) ratio for both groups of *PRX*s and a two-ratio model in which

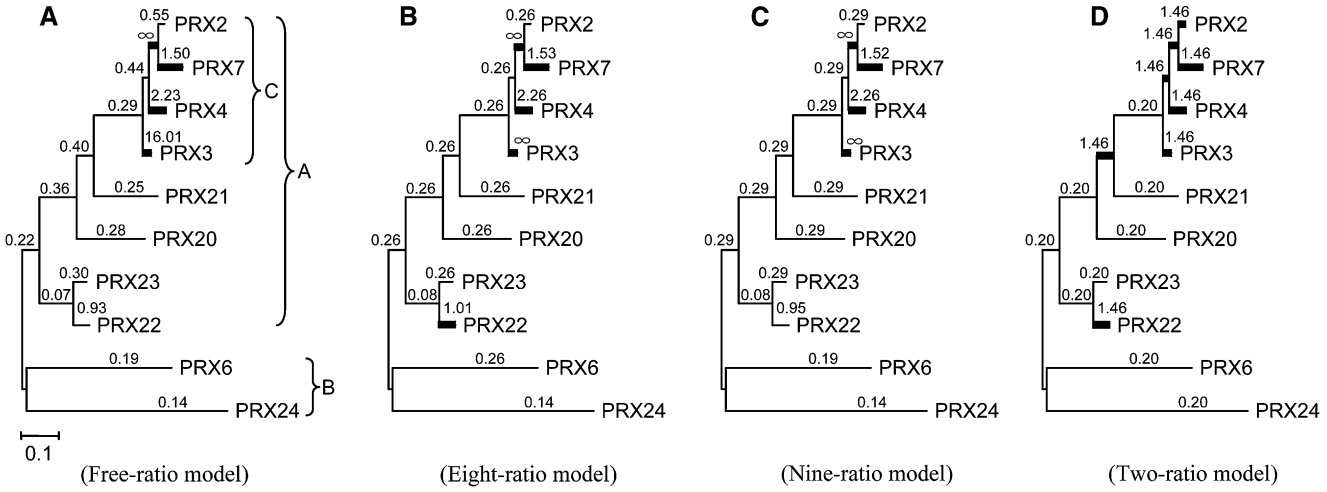


Figure 4. Molecular Evolution of PRX Genes in Clusters I and II as Revealed by Different Branch Models.

The phylogenetic relationship was reconstructed using the WAG model with 100 bootstrap replicates. The d_N/d_S (ω) values calculated using CODEML in the PAML package are shown in each branches. The bold lines illustrate branches with ω > 1. In some instances, zero synonymous substitutions lead to an apparent ω value of infinity.

the two PRX types were assigned a different ω ratio. The log-likelihood values under the one-ratio and two-ratio models were ln $L = -5617.378566$ and -5612.132522, respectively (Table 1). The likelihood ratio test showed that the two-ratio model rejected the null model (one-ratio model), indicating that the selective pressure differed significantly between the two PRX groups ($P < 0.001$). Under the two-ratio model, the ω values for eight vacuole *PRX*s and two cell wall *PRX*s were 0.31267 and 0.16888, respectively, indicating that the vacuole *PRX*s were under more relaxed selection constraints than the cell wall *PRX*s.

To test whether some branches in the phylogenetic tree of the 10 *PRX*s were under positive selection, we first applied a free-ratio model test using PAML (Yang, 2007). The free-ratio model allows an independent assignment of ω to each branch. This analysis showed three branches (*PRX3*, *4*, and *7*, clade C) had ω > 1 (bold branches in Figure 4A). We then applied a dynamic maximum likelihood–based procedure to globally search for optimal models to estimate substitution rates (Zhang et al., 2011). This procedure found three optimal branch-specific models for the 10 *PRX*s (Figures 4B to 4D; Supplemental Table 4). These models support the pattern revealed by the free-ratio model and consistently identified the same three branches in clade C as ω > 1.

To further assess the sites of selection in the sequences identified by the branch models (i.e., clade C in Figure 4), we performed site model tests using PAML (Table 1). This analysis identified the M8 model as the most probable, under which 13 amino acid sites were likely under positive selection (Table 1), including seven with posterior probabilities >0.95 (alignment positions 57, 92, 161, 167, 180, 200, and 247 in Figure 5). We also applied four codon substitution models (M0, M3, M3 + S1, and M3 + S2) in Fitmodel software (Guindon et al., 2004) to investigate the variation in selection among amino acid sites for the 10 *PRX*s in Figure 4. The likelihood ratio tests between nested models (M0 versus M3; M3 versus M3+S1; M3+S1 versus M3+S2) suggested that the M3 + S2 model was significantly better than the other models (Table 2). The switching rate between ω_2 (moderate purifying selection) and ω_3 (positive selection) ($R_{23} = 7.384$) was significantly higher than that between ω_1 (strong purifying selection) and ω_2 ($R_{12} = 0.545$) and that between ω_1 and ω_3 ($R_{13} = 0.002$) (Table 2). Under the M3 + S2 model, seven amino acid sites were inferred to be under positive selection ($\omega_3 = 4.149$) in the four PRXs in clade C (Figure 6). These sites corresponded to the seven positive selection sites (with posterior probabilities >0.95) identified by the site model in PAML (Table 1).

Structural Distribution of the Putative Positively Selected Sites

To understand the evolutionary changes of the putative positively selected sites, we reconstructed the most recent common ancestral protein (MRCAP) of the four PRXs (PRX2, 3, 4, and 7) using CODEML in the PAML package with an overall accuracy of 0.99171. Based on the x-ray structure of the *Armoracia rusticana* PRX protein (PDB accession number 7ATJ), we modeled the three-dimensional structures of the five PRX proteins (PRX2, 3, 4, 7, and MRCAP). The simulated structures of the five PRX proteins were superimposed to evaluate the goodness of fit of the overall topologies. The five PRX proteins shared the same conformation of structural elements for the α-helices and β-sheets (Supplemental Figure 3).

PRX proteins are heme-containing enzymes that can catalyze the oxidative coupling of phenolic compounds using H_2O_2 as the oxidizing agent (Henriksen et al., 1999). Based on sequence and structural comparisons, seven residues (corresponding to Arg-55, Ser-59, Arg-62, His-66, His-194, Arg-198, and Arg-200 of the MRCAP; red arrows in Figure 5) were predicted as active

Table 1. Results of the CODEML Analyses of the Selective Pattern for *Populus* PRXs

Model	Estimates of Parameters[a]	ln *L*	χ^2	P	Positively Selected Sites[b]
Branch model					
One ratio	ω = 0.26796 for all branches	−5617.378566			
Two ratios	ω_0 = 0.16888 for clade B ω_1 = 0.31267 for clade A	−5612.132522	10.492088	<<0.001	
Site model					
M0	ω = 1.25976	−1880.168952			
M3 (discrete)	p_0 = 0.69998, p_1 = 0.28721, p_2 = 0.01281 ω_0 = 0.00000, ω_1 = 4.52646, ω_2 = 20.98092	−1866.554246	27.229412	<<0.001	
M7 (β)	p = 0.48468, q = 0.00500	−1880.522937			
M8 (β and ω)	p_0 = 0.84471, p = 0.00743, q = 0.00993 (p1 = 0.15529) ω = 7.86361	−1866.698144	27.649586	<<0.001	57*, 92*, 140, 156, 161*, 167*, 180*, 200*, 213, 221, 236, 247*, 256

The asterisk indicates posterior probability >95% of having $\omega > 1$. Clade A and clade B refer to the two clades of PRXs shown in Figure 4.
[a]The proportion of sites (p_0, p_1, . . .) estimated to have ω_0, ω_1, . . .
[b]The numbering of residues identified by Bayes Empirical Bayes analysis corresponds to their alignment positions in Figure 5.

sites that could bind the substrate heme or ferulic acid through hydrogen bonds. Six residues (corresponding to Gly-93, Pro-163, Ala-164, Phe-166, Phe-176, and Thr-203 of the MRCAP; purple arrows in Figure 5) could interact with substrates within 4.0 Å through hydrophobic interactions. These six residues were also considered active sites. Among the seven putative positively selected sites, five (alignment positions 57, 92, 161, 167, and 180, corresponding to Gly-57, Ala-92, Ser-161, Leu-167, and Gly-180 of the MRCAP; Figures 5 and 7) were adjacent to the active sites, one (alignment positions 200, corresponding to Arg-200 of the MRCAP; Figures 5 and 7) was predicted to be an active site for binding heme, and another (alignment position 247, corresponding to Pro-247 of the MRCAP; Figures 5 and 7) was distant from active site regions.

Substrate Activity of the PRX Proteins in Clusters I and II

To examine the biochemical functions of the 10 PRX proteins in clusters I and II, we expressed and purified the corresponding proteins. Plant PRXs can catalyze the reduction of H_2O_2 by moving electrons to various donor molecules. In this study, six substrates (listed in Table 3) were selected to determine PRX protein activity.

All 10 purified PRX proteins showed enzymatic activities toward these six substrates with much higher enzymatic activities toward coniferyl alcohol than toward the other five substrates (paired-samples *t* test, $P < 0.003$), except for PRX21 toward sinapyl alcohol and PRX23 toward pyrogallol. We conducted a multiple response permutation procedure (MRPP) test to evaluate the significance of differences in the enzyme activity profiles between the cell wall and vacuole PRXs. MRPP is a nonparametric test that is flexible over unequal sample sizes and the violation of normality assumptions. The MRPP test showed no significant differences in the enzyme activity profiles between the two PRX types ($P > 0.55$). However, the Kendall's W nonparametric test for related samples indicated a significant difference in enzyme activity among the eight vacuole PRXs ($P = 0.00$). Thus, diversification in enzyme activity toward different substrates has apparently evolved among vacuole PRXs.

Kinetics of the PRX Proteins

The apparent kinetic constants of the 10 PRX proteins were determined using coniferyl alcohol, ferulic acid, and pyrogallol as substrates (Table 4). The 10 PRX proteins showed much higher affinities (higher $1/K_m$), turnover number (higher k_{cat}), and catalytic efficiency (higher k_{cat}/K_m) toward coniferyl alcohol and ferulic acid (paired-samples *t* test, $P < 0.01$) than toward pyrogallol. The MRPP test showed no significant difference in the enzyme kinetics between the cell wall and vacuole PRXs ($P > 0.30$). However, the apparent kinetic constants of the eight vacuole PRX proteins showed significant variation (Kendall's W test, $P = 0.00$) toward the substrates (Table 4). For example, the catalytic efficiency toward coniferyl alcohol varied from 600.83 to 64,632.39 $mM^{-1}\ s^{-1}$.

Functional Significance of Positive Selection

Branch model tests indicated that four vacuole PRXs (PRX2, 3, 4, and 7) in cluster I were under positive selection pressure (Figures 4 and 6). These four PRXs were derived after the recent whole-genome duplication in *Populus*. The enzymatic activities of these four vacuole PRXs varied 8-fold for sinapyl alcohol and 41-fold for ascorbic acid (Table 3). To understand the evolution of this enzymatic activity, we expressed and purified the MRCAP of the four vacuole PRXs. Wilcoxon nonparametric tests for two related samples indicated significant differences in enzyme activity profiles between the MRCAP and each of the four PRXs ($P < 0.046$). PRX2 and 3 showed much higher enzymatic activity toward all six substrates than the MRCAP (Table 3), except PRX2 toward ascorbic acid. The apparent kinetic constant values for coniferyl alcohol, ferulic acid, and pyrogallol were similar among MRCAP and PRX2 and 3 (Wilcoxon nonparametric tests,

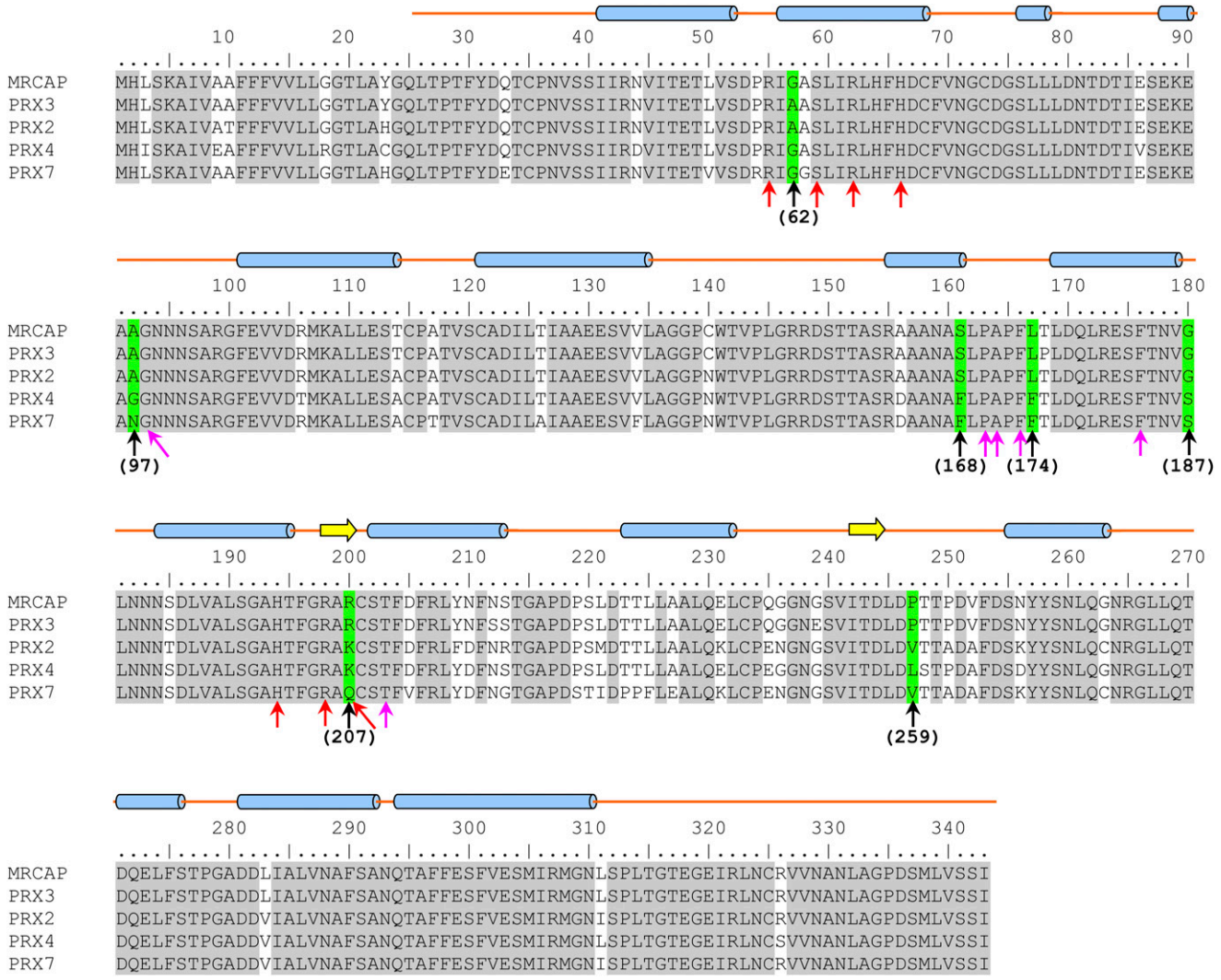


Figure 5. Alignment of PRX2, 3, 4, and 7 and Their MRCAP.

α-Helices and β-strands are represented as blue cylinders and yellow arrows, respectively. Conserved residues in these five PRXs are marked in gray. Positive selection sites (with posterior probabilities >0.95) calculated using the site models in the PAML package are marked in green. The predicted active sites for binding substrates are indicated with red and purple arrows. The numbers in parentheses correspond to positive selection sites (Figure 6) predicted by Fitmodel software.

P > 0.21). PRX4 and 7 showed much lower enzymatic activity toward all six substrates than MRCAP (Table 3). Significant differences in the apparent kinetic constants toward the three substrates listed in Table 4 were observed between the MRCAP and each of PRX4 and 7 (Wilcoxon nonparametric tests, P < 0.01). Compared with the MRCAP, except for affinity of PRX4 toward substrate coniferyl alcohol, PRX4 and 7 showed much lower affinities (lower $1/K_m$), turnover number (lower k_{cat}), and catalytic efficiency (lower k_{cat}/K_m) toward substrates coniferyl alcohol, ferulic acid, and pyrogallol (Table 4).

Table 2. Likelihood Ratio Test of the Models in the Fitmodel Program for *Populus* PRXs

	M0 (No Heterogeneity)	M3 (Heterogeneity across Sites)	M3+S1 (Shifting across Branches)	M3+S2 (Unequal Switching Rates)
ln *L*	−6136.029	−5950.128	−5875.428	−5863.143
$\omega_1\ \omega_2\ \omega_3$	0.340	0.002 0.314 1.417	0.000 0.305 1.713	0.001 0.184 4.149
$p_1\ p_2\ p_3$	1.000	0.302 0.474 0.224	0.496 0.295 0.208	0.416 0.471 0.113
$R_{12}\ R_{13}\ R_{23}$			1.605 1.605 1.605	0.545 0.002 7.384
χ^2 (models; P)		371.802 (M0 vs. M3; P < 0.001)	149.400 (M3 vs. M3+S1; P < 0.001)	24.570 (M3+S1 vs. M3+S2; P < 0.001)

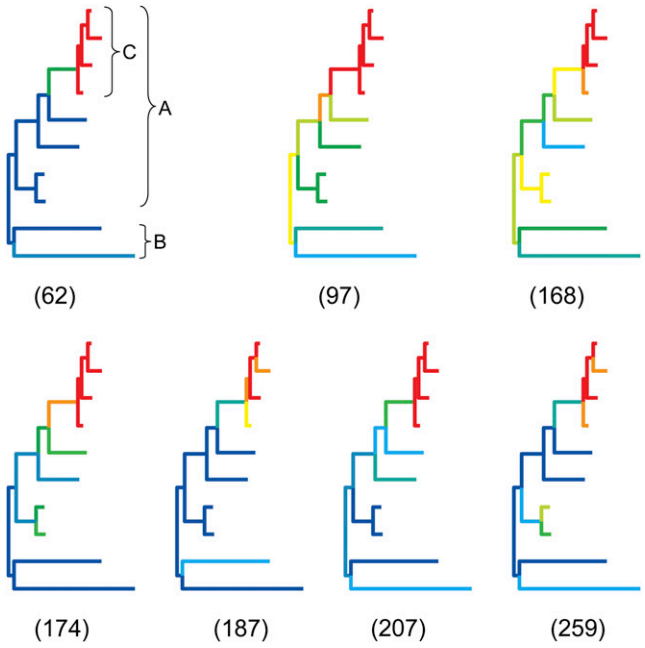


Figure 6. Site-Specific Selection Patterns on the PRX Genes in Clusters I and II.

The phylogenetic tree corresponds to the tree shown in Figure 4. Red branches with posterior probabilities >90% for selection class ω_3 were considered to be under positive selection. Blue branches with posterior probabilities lower than 20% for selection class ω_1 were considered to be under strong purifying selection. The branches with posterior probabilities from 21 to 89% for selection class ω_2 are shown with cool (light blue) to warm (orange) colors and were considered to be under moderate purifying selection. ω_1, ω_2, and ω_3 are presented in Table 2. The numbers in parentheses correspond to the amino acid sites in the alignment of *Populus* PRX genes in clusters I and II (Supplemental Figure 4).

Seven positive selection sites were detected in the four vacuole PRXs (PRX2, 3, 4, and 7) (Figures 5 and 6). To investigate the functional significance of these sites, we constructed 10 mutants using site-directed mutagenesis for the biochemical assays. For each of the positive selection sites, we mutated the corresponding sites in the MRCAP to the amino acid residues present in PRX2, 3, 4, and 7 (Table 3). The first positively selected site (alignment position 57 in Figure 5), corresponding to Gly-57 of MRCAP, was the Gly residue in PRX4 and 7, and the Ala residue in PRX2 and 3. When Gly-57 of MRCAP was replaced with Ala, the mutant G57A showed much higher enzymatic activity toward ascorbic acid, pyrogallol, guaiacol, ferulic acid, and sinapyl alcohol ($P < 0.05$, Mann-Whitney U test) and similar enzymatic activity toward coniferyl alcohol (Table 3). The fourth and fifth positively selected sites (alignment positions 167 and 180 in Figure 5) corresponded to Leu-167 and Gly-180 of MRCAP. PRX2 and 3 shared the same residues with MRCAP at these two sites, whereas these two sites were substituted with Phe and Ser in PRX4 and 7, respectively (Figure 5). When Leu-167 and Gly-180 of MRCAP were replaced with Phe and Ser, respectively, the L167F and G180S mutants showed much lower enzymatic activity toward all six substrates listed in Table 3 ($P < 0.05$, Mann-Whitney U test) (Figure 8). When the other four positively selected sites (alignment positions 92, 161, 200, and 247 in Figure 5) in the MRCAP were mutated to the corresponding amino acid residues present in PRX2, 3, 4, and 7, each of the MRCAP mutants (A92G, A92N, S161F, R200Q, R200K, P247V, and P247L) showed a diverse response to the tested substrates (Figure 8). For example, the P247V mutant showed decreased enzymatic activity toward ascorbic acid and coniferyl alcohol but increased enzymatic activity toward pyrogallol, guaiacol, ferulic acid, and sinapyl alcohol (Table 3).

DISCUSSION

Gene duplication is thought to play an important role in providing raw material for plant evolution (Ohno, 1970). However, in the event of a redundancy, duplicate genes may lose their functions and ultimately be removed by deletion over the course of evolution. In this study, we found 43 PRX fragments from the *Populus* genome, indicating that massive gene losses have occurred in this gene family. Nonetheless, 93 full-length PRXs have been retained in the *Populus* genome, bringing us to a central issue in the evolution of duplicate genes: determining which evolutionary mechanisms play a central role in the retention and functional diversification of duplicate genes. The evolutionary mechanisms of the retention of duplicate genes might vary among different gene families. Previous studies have shown that divergence in gene expression or protein function contributes to the retention of duplicate genes (Ganko et al., 2007; Lan et al., 2009; Yang et al., 2013). Protein subcellular relocalization has recently been considered another mechanism for gene retention because proteins could alter their functions when relocalized to a new subcellular structure (Byun-McKay and Geeta, 2007; Marques et al., 2008; Qian and Zhang, 2009). Plant PRXs are involved in a broad range of physiological processes throughout the plant life cycle (Passardi et al., 2005). These proteins are localized to the cell wall or vacuole. The cell wall PRXs are involved in cell wall metabolism, including lignin polymerization, whereas the vacuole PRXs are involved in defense responses to abiotic and biotic stresses and the oxidation of toxic reductants and other secondary metabolites (Ostergaard et al., 2000; Li et al., 2003; Gabaldón et al., 2005; Bindschedler et al., 2006; Costa et al., 2008). This divergence in subcellular localization might be an important first step in the evolution of new PRX functions following gene duplication. This study examined the functional evolution of the *Populus* PRX gene family and the possible drivers for it.

Divergence Aided by C-Terminal Signal Peptide

In the eukaryotic cell, most proteins are synthesized in the cytosol, and many need to be further sorted to the cell membrane or various membrane-bound cellular organelles. In the plant cell, when the final destination is the chloroplast, mitochondrion, or secretory pathway, sorting usually relies on the presence of an N- or C-terminal targeting sequence (signal peptide) (Emanuelsson et al., 2000). In most cases, the signal peptide sequence will be removed after the proteins are directed to their appropriate subcellular organelles. Thus, the signal peptide does not usually participate directly in protein function. The loss

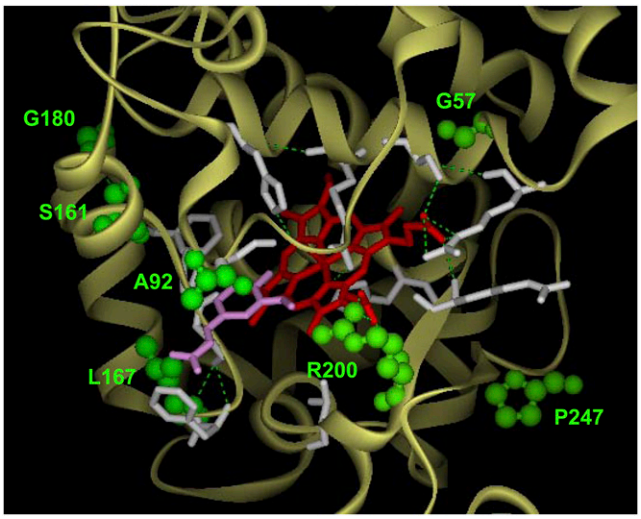


Figure 7. The Predicted Model of Interaction between MRCAP and Its Ligands.

Heme and ferulic acid are denoted by red and purple sticks, respectively. The predicted active sites for binding heme and ferulic acid are indicated with white sticks. The positively selected site residues are indicated by green balls and hydrogen bonds with green dashed lines.

or gain of a signal peptide can easily occur in duplicate genes. If an incomplete duplication event produced only a partial copy with a partial or lost signal peptide, the protein coded by this incomplete duplication gene may immediately alter its subcellular localization. Alternative translation initiation or point mutations occurring in the stop codon can also result in the addition of amino acids in the N- or C-terminal sequence of the proteins (Leissring et al., 2004). Because altering the subcellular target by changing signal peptide sequences is rapid, the resulting functional divergence among duplicate genes through subcellular relocalization should also be rapid. In plant PRXs, a highly hydrophobic C-terminal extension acts as a vacuolar sorting signal (Carter et al., 2004; Costa et al., 2008). In a horseradish PRX (HRP C1a), this vacuolar sorting signal consists of a 15–amino acid C-terminal peptide that includes a highly conserved hexapeptide that plays a key role in vacuolar sorting (Matsui et al., 2011). When the *Populus* PRXs in clusters I and II were aligned with HRP C1a and *Arabidopsis* vacuole PRXs, the vacuolar sorting signal peptide was not observed in the two cell wall PRXs (PRX6 and 24) (Figure 3B). The eight vacuole PRXs (PRX2, 3, 4, 7, 20, 21, 22, and 23) had the vacuolar sorting signal peptide as well as a highly conserved C-terminal hexapeptide (Figure 3B, black shading). Based on the putative duplication history of the PRXs in clusters I and II, the ancestor (Figure 2C, ancestral copy a) of clusters I and II was likely a cell wall type, and the ancestor (Figure 2C, ancestral copy d) of vacuole PRXs may have evolved by gaining the vacuolar sorting signal peptide.

Functional Divergence of Vacuole PRX Proteins

The *Populus* vacuole PRX proteins examined in this study showed different enzymatic activities and/or catalytic efficiency toward different oxidative substrates, indicating divergence in their biochemical properties. To understand the process and direction of the enzymatic divergence, we reconstructed the

Table 3. Specific Activities of the *Populus* PRXs and Mutants to Six Substrates (Mean ± SD Obtained from at Least Three Independent Determinations)

	Specific Activity (μmol/min per mg)					
	Ascorbic acid	Pyrogallol	Guaiacol	Ferulic Acid	Coniferyl Alcohol	Sinapyl Alcohol
PRX2	279.42 ± 8.93	2138.40 ± 44.55	821.06 ± 13.69	1993.73 ± 59.47	4192.93 ± 10.93	206.72 ± 10.63
PRX3	379.54 ± 15.58	1968.64 ± 32.13	561.22 ± 6.45	2037.17 ± 29.96	3369.13 ± 100.62	181.81 ± 1.70
PRX4	81.96 ± 4.91	671.70 ± 6.60	219.32 ± 6.52	865.33 ± 15.73	2394.40 ± 3.64	55.63 ± 0.72
PRX6	29.61 ± 0.76	520.02 ± 12.64	115.89 ± 2.73	704.60 ± 43.05	3379.71 ± 262.40	37.96 ± 2.61
PRX7	9.37 ± 0.45	58.59 ± 4.59	43.86 ± 1.11	218.30 ± 5.90	250.56 ± 1.82	26.57 ± 1.61
PRX20	8.83 ± 0.50	561.60 ± 10.80	598.50 ± 0.50	953.44 ± 28.32	4271.29 ± 36.44	31.93 ± 0.94
PRX21	120.98 ± 4.36	72.90 ± 8.10	171.08 ± 1.79	446.04 ± 2.36	507.49 ± 3.64	531.30 ± 15.46
PRX22	65.47 ± 1.92	3073.14 ± 53.46	3559.20 ± 55.20	1789.82 ± 22.66	5352.96 ± 104.96	579.60 ± 15.46
PRX23	61.15 ± 3.75	1248.75 ± 12.15	486.92 ± 2.92	364.62 ± 15.34	1120.67 ± 16.40	18.84 ± 1.93
PRX24	42.11 ± 1.37	171.12 ± 2.55	129.25 ± 5.90	189.19 ± 8.37	977.32 ± 31.03	41.26 ± 1.57
MRCAP	310.92 ± 5.06	1034.04 ± 13.79	428.40 ± 1.17	1768.84 ± 63.17	2899.64 ± 92.26	156.23 ± 2.31
MRCAP(G57A)	381.75 ± 10.47*	1276.02 ± 22.68*	493.75 ± 10.25*	2410.50 ± 34.46*	2988.63 ± 72.16	182.90 ± 5.80*
MRCAP(A92G)	151.54 ± 1.46*	1991.11 ± 45.46*	404.29 ± 2.96*	1541.22 ± 38.53*	3547.76 ± 66.94*	93.97 ± 2.30*
MRCAP(A92N)	263.84 ± 2.72*	1588.37 ± 16.97*	510.14 ± 1.14*	2184.16 ± 37.02*	2909.05 ± 27.33	137.01 ± 1.26*
MRCAP(S161F)	317.39 ± 10.59	781.71 ± 1.32*	277.81 ± 4.52*	1204.36 ± 14.46*	1700.51 ± 41.99*	175.33 ± 4.57*
MRCAP(L167F)	223.39 ± 7.07*	892.76 ± 10.91*	310.25 ± 8.76*	1293.56 ± 17.76*	1364.77 ± 11.37*	77.47 ± 2.87*
MRCAP(G180S)	221.64 ± 1.49*	541.69 ± 3.54*	197.03 ± 2.13*	1011.26 ± 4.72*	1327.63 ± 55.61*	129.93 ± 5.31*
MRCAP(R200Q)	80.89 ± 2.66*	1383.32 ± 19.53*	630.00 ± 1.91*	1408.47 ± 22.60*	3040.11 ± 82.87	322.69 ± 6.17*
MRCAP(R200K)	82.73 ± 2.40*	1529.65 ± 3.12*	558.92 ± 13.96*	2033.23 ± 36.31*	2880.51 ± 52.56	103.41 ± 2.17*
MRCAP(P247V)	262.34 ± 2.70*	2125.64 ± 8.59*	564.72 ± 13.81*	2370.73 ± 21.45*	2586.73 ± 59.64*	202.71 ± 16.83*
MRCAP(P247L)	246.54 ± 2.44*	1728.00 ± 37.38*	448.97 ± 0.90*	1791.18 ± 48.41	2410.94 ± 28.03*	133.75 ± 4.95*

Asterisk indicates significant difference ($P < 0.05$) in enzymatic activity between the MRCAP wild type and its mutants.

Table 4. Kinetic Constants of the *Populus* PRXs for the Substrates Coniferyl Alcohol, Ferulic Acid, and Pyrogallol

	Coniferyl Alcohol			Ferulic Acid			Pyrogallol		
	$1/K_m$ (mM^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})	$1/K_m$ (mM^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})	$1/K_m$ (mM^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1} s^{-1})
PRX2	3.70	11481.57	42481.81	3.57	14507.34	51791.20	0.36	776.96	279.71
PRX3	1.75	15897.90	27821.33	3.57	10595.86	37827.22	0.47	964.28	453.21
PRX4	2.94	5869.78	17257.15	2.33	6528.32	15210.99	0.15	296.97	44.55
PRX6	3.85	9225.69	35518.91	5.26	2606.66	13711.03	5.56	50.15	278.83
PRX7	0.95	632.45	600.83	3.13	494.36	1547.35	0.24	17.55	4.21
PRX20	1.89	12129.29	22924.36	2.13	8158.13	17376.82	0.07	1108.00	77.56
PRX21	5.00	878.46	4392.30	14.29	782.13	11176.64	0.05	92.84	4.64
PRX22	5.56	11624.53	64632.39	4.17	10498.26	43777.74	0.34	2813.50	956.59
PRX23	4.17	2713.45	11315.09	1.89	3039.35	5744.37	0.16	958.95	153.43
PRX24	3.23	4199.54	13564.51	7.69	963.65	7410.47	0.58	330.12	191.47
MRCAP	2.22	10298.32	22862.27	4.76	9321.98	44372.62	0.62	875.08	542.55

MRCAP of the four PRXs (PRX2, 3, 4, and 7) that were generated after the recent whole-genome duplication in *Populus* and examined the biochemical characteristics of the MRCAP. This paleo-molecular biochemistry approach has unraveled the evolutionary mysteries of some duplicate genes (Zhang and Rosenberg, 2002; Huang et al., 2012; Smith et al., 2013). We found that the enzymatic activities of the MRCAP toward six oxidative substrates were significantly lower than that of the PRX2 and 3 (except for PRX2 toward ascorbic acid) but higher than that of PRX4 and 7 (Table 3). The Escape from Adaptive Conflict (EAC) model might be the best explanation for this divergence pattern. Under the EAC model, after duplication each daughter gene is free to specialize for either the original or the novel function and to improve those functions (Des Marais and Rausher, 2008). If so, we might expect to observe that the different copies of PRXs perform different functions and that the ancestral protein MRCAP performed both of these functions, albeit with reduced efficiency. Increased enzymatic activities toward five oxidative substrates indicated that PRX2 and 3 improved the ancestral protein's response to these substrates. On the other hand, PRX4 and 7 showed reduced function for these substrates and might be specialized for other substrates not tested in this study. Thus, EAC seems a clear feature of vacuole PRX evolution.

The Gene Balance Hypothesis, also known as the gene dosage hypothesis, has been proposed as a general mechanism to explain long-term retention of duplicate genes (Birchler and Veitia, 2007, 2010). The Gene Balance Hypothesis posits that genes encoding components of multisubunit complexes are more likely to evolve in concert because the dosage change in the quantities of subunits affects the interaction and function of the whole complex (Birchler et al., 2011). Such genes, e.g., ribosomal proteins, protein kinases, and transcription factors that have extensive functional interactions with other biomolecules, are preferentially retained following whole-genome duplication (Birchler et al., 2001; Edger and Pires, 2009; Freeling, 2009). In contrast, genes encode more poorly connected products in a network or complex, such as PRX and those involved in stress responses in plants, would have an elevated probability of retention following tandem duplication (Hanada et al., 2008; Freeling, 2009). The *Populus* vacuole *PRX*s expanded by tandem duplication. The increase of PRX protein abundance dictated by increased gene copy number may be beneficial in defense responses especially in changing environments. Selection for increased gene dosage allows duplications to be selectively advantageous at birth, which increases the likelihood of subsequent functional divergence in duplicate genes (Conant and Wolfe, 2008). This dosage selection may also have played a role in the retention of the tandem duplicates for novel vacuole PRX genes.

An analysis of the biochemical functions of positive selection sites provided further evidence of adaptive changes in *Populus* vacuole PRXs. We used site-directed mutagenesis to investigate the functional profiles of seven positively selected sites in PRX2, 3, 4, and 7 relative to MRCAP. For example, one positive selection site (alignment position 57 in Figure 5) is Gly in PRX4, 7, and MRCAP, whereas it is Ala in PRX2 and 3. When this Gly site of MRCAP was replaced by Ala, the mutant protein G57A showed increased enzymatic activities toward five oxidative substrates (Table 3), indicating that the substitution of Ala for Gly at this amino acid site for PRX2 and 3 might confer an adaptive advantage. Conventional theory emphasizes changes in function produced by amino acid replacements at active sites (Chen et al., 1995). In this study, only one site (alignment position 200 in Figure 5) among the seven positive selection sites in PRX2, 3, 4, and 7 was an active site for binding heme, and five

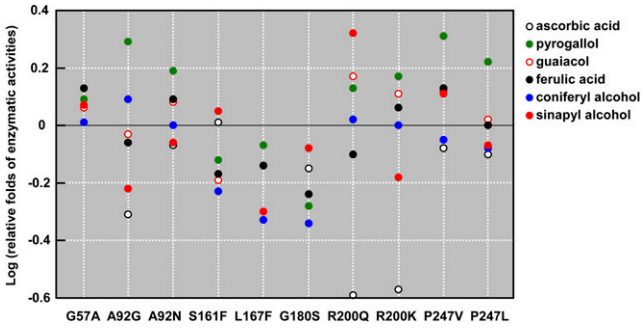


Figure 8. Changes in Enzymatic Activity Induced by Mutations at the Seven Selected Sites Relative to the MRCAP Protein.

The six substrates examined are each denoted by a different color. The enzymatic activity of the MRCAP toward each substrate is set as the baseline for comparison with the 10 mutants specified on the *x* axis.

sites (alignment positions 57, 92, 161, 167, and 180 in Figure 5) were adjacent to the active sites. Substitution of these residues resulted in the changes in the substrate activities. Our data showed that natural selection for amino acid replacements outside active sites could also produce significant functional changes in the enzyme family. Our previous study of pine glutathione *S*-transferase proteins also showed that positive selection has mainly acted on sites close to (but not directly at) the substrate binding site (Lan et al., 2013). Another case study found that positive selection for amino acid replacements outside the active site of the *Drosophila melanogaster* JGW protein produced a novel dehydrogenase with altered substrate specificity compared with the ancestral protein (Zhang et al., 2004). Mutations directly in the active sites might cause large changes that can disrupt catalysis or drastically alter the shape of the binding pocket, whereas mutations close to the active sites that would more subtly alter the substrate binding site might be more effective (Morley and Kazlauskas, 2005). Thus, positive selection acting on residues adjacent to, rather than directly at, a critical active site is likely a general mechanism for the functional diversification of enzyme families.

Protein loop regions are relatively flexible, compared with the rigid structural elements (α-helices and β-strands) that they connect. Flexible protein regions are not simply an outcome of looser packing or instability; rather, they are evolutionally selected (Sinha and Smith-Gill, 2002). Mutations in loop regions often change the catalytic characterization of the proteins but do not disrupt the native folded structure. Our previous study of pine glutathione *S*-transferase proteins showed that positive selection sites were mainly located in the loop regions of the protein structure (Lan et al., 2013). In this study, among the seven positive selection sites in PRX2, 3, 4, and 7, four were located in loop regions and two (alignment positions 161 and 200 in Figure 5) were located in terminals of the α-helix or β-strand. This indicates that the loop regions of the PRX protein might tolerate moderate structure modification. The adaptive functional diversification of enzyme families might be achieved through modifications of flexible regions more often than modifications of rigid structural elements, as they allow the acquisition of new functions without disrupting native folding structures.

Perennial plants are sessile organisms that cannot escape their environment and are constantly exposed to a multitude of environmental stresses. Thus, they have evolved various plastic mechanisms for responding to a wide range of potential threats. The expansion of a secondary metabolism with defense, communication, and protection roles is now considered particularly important (Ferreres et al., 2011). In this study, we selected four naturally occurring compounds (ascorbic acid, ferulic acid, coniferyl alcohol, and sinapyl alcohol) in the plant cell to study PRX activity. Ascorbic acid is a major metabolite in plants. It is an antioxidant and in association with other components of the antioxidant system protects plants against oxidative damage resulting from aerobic metabolism, photosynthesis, and a range of pollutants (Smirnoff, 1996). Ferulic acid can enhance the rigidity and strength of plant cell walls by cross-linking with pentosans, arabinoxylans, and hemicelluloses, thereby making the cell walls less susceptible to enzymatic hydrolysis when plants are attacked by pathogens (Henriksen et al., 1999). Coniferyl alcohol and sinapyl alcohol are monolignols that participate in the formation of the lignin polymers (Bonawitz and Chapple, 2010). These secondary compounds play important roles in plant defense responses to biotic and abiotic stresses. We found that most of the identified positively selected sites in PRX2, 3, 4, and 7 have significant effects on the enzymatic activities to these compounds. Hence, there is likely adaptive value in possessing PRXs with diverse activities to a wide range of secondary metabolites and high evolutionary flexibility, enabling perennial plants to respond to the diverse environmental challenges they are likely to encounter.

The analysis of complete genome sequences has revealed an unexpected prevalence of gene duplications, particularly in plants. The molecular mechanisms of retention and subsequent functional divergence of duplicate genes are still poorly understood. In this study, after conducting genome-wide gene family annotation and expression analyses, we focused on two tandem-arrayed PRX gene clusters in the *Populus* genome. Using molecular evolutionary analyses, subcellular localization experiments, reconstruction of the ancestral protein, site-directed mutagenesis, and enzyme functional assays of ancestral and current proteins, we elucidated the evolutionary process that has contributed to the divergence of the PRX gene family in *Populus* and possibly other plant species. Our findings suggest that protein subcellular relocalization and positive selection play key roles in the retention and subsequent functional divergence of duplicate genes.

METHODS

Identification of PRX Genes in the *Populus* Genome

To identify *Populus* PRX genes, the *Populus trichocarpa* genome database version 2.2 (http://www.phytozome.net/) was searched with 73 *Arabidopsis thaliana* PRX protein sequences (Supplemental Data Set 1) using the TBLASTN program with default algorithm parameters. This search found 93 full-length PRX candidates, of which 10 (*PRX5*, *8*, *18*, *32*, *38*, *47*, *60*, *66*, *69*, and *76*) had frame shifts disrupting the coding region or a stop codon occurring prematurely, resulting in a truncated protein. After removing these stop codons or revising the frame shifts by deleting one or two nucleotides, these 10 full-length sequences were included in the following analyses. These 93 PRX candidates were further analyzed by a search for conserved domains in the National Center for Biotechnology Information database to confirm the presence of peroxidase domains in their protein structures. All 93 *Populus* PRXs contained peroxidase domains, confirming that these genes belong to the peroxidase family. Also, we performed a joint phylogenetic analysis of all 73 *Arabidopsis* PRXs and 93 *Populus* PRXs (Supplemental Figure 5). The phylogenetic tree grouped PRXs from both species with no clear taxonomic division, further confirming that the 93 *Populus* PRXs belong to the PRX gene family. Then, the *Populus* PRX genes were amplified from mRNA from *P. trichocarpa*, cloned into the *pEASY*-T3 vector (TransGen), and sequenced in both directions to verify the gene sequences.

Phylogenetic and Molecular Evolution Analyses

The full-length PRX protein sequences were aligned using the MUSCLE version 3.8.31 (Edgar, 2004) and further adjusted manually using BioEdit version 7.0.5.3 (Hall, 1999) before all phylogenetic analysis. The optimal substitution model of amino acid substitution was selected using the program ModelGenerator version 0.84 (Keane et al., 2006). Phylogenetic

trees were constructed using the maximum likelihood method with PhyML software version 2.4.4 (Guindon and Gascuel, 2003). Two PRX proteins from *Ostreococcus lucimarinus* and *Chlamydomonas reinhardtii* were used as outgroups. One hundred bootstrap replicates were performed in each analysis to obtain the confidence support level.

To evaluate variation in selective pressures between the vacuole and cell wall *PRX*s in clusters I and II (Figure 4), maximum likelihood codon models in PAML version 4.4 (Yang, 2007) were used to estimate the ω ($=d_N/d_S$) ratio under two assumptions: a one-ratio model that assumes the same ω ratio for two types (clades) of PRXs and a two ratio model in which the two clades were assigned to different ω ratios. To test whether some branches in the phylogenetic tree of these PRXs were under positive selection, a free-ratio model in the PAML package was used. Furthermore, the maximum-likelihood procedure that globally searches for optimal substitution models for a given phylogeny (Zhang et al., 2011) was employed as an additional model search strategy. To detect variation in site-specific selection among branches, we performed the Fitmodel test (Guindon et al., 2004) on the phylogenetic tree of the vacuole and cell wall PRXs (Figure 6). We also performed the site model analyses in the PAML package to detect positive selection sites in *PRX2*, *3*, *4*, and *7*. Six site models in PAML were explored: the one-ratio model (M0), the discrete model (M3), the nearly neutral model (M1a), the positive-selection model (M2a), the beta model (M7), and the beta and ω model (M8). To verify which of the models best fit the data, likelihood ratio tests were performed by comparing twice the difference in log likelihood values between pairs of the models using a χ^2 distribution (Yang et al., 2000).

Expression of PRX Genes under Different Treatments

To investigate the expression patterns of the *Populus* PRX genes under normal and abiotic stress conditions, cuttings of *P. trichocarpa* were cultivated in potting soil at 25°C under a 14/10-h light/dark cycle in a growth chamber for 2 months before treatment. For H_2O_2, salicylic acid, and NaCl treatments, cuttings were then irrigated and sprayed using 0.5% H_2O_2 solution for 12 h, 3.5 mM salicylic acid for 24 h, and 150 mM NaCl solution once a day for 1 week. Drought stress was conducted by withholding water for 2 weeks. Each treatment consisted of three replicates. After treatment, total RNAs were isolated from root, shoot, leaf, bud, and phloem tissues of stem using an Aurum Total RNA kit (Bio-Rad Laboratories). Total RNAs were treated with RNase-free DNase I (Promega) and reverse transcribed into cDNA using an RNA PCR kit (AMV) version 3.0 (TaKaRa). Based on the multiple sequence alignment of the *Populus* PRX gene sequences, 101 specific primer pairs were designed for PCR analysis (Supplemental Table 2). PCR was performed in a volume of 25 μL containing 3 μL cDNA, 2.5 μL TaKaRa 10× PCR buffer, 0.125 μL TaKaRa ExTaq (5 units/μL), 1 μL deoxyribonucleotide triphosphate (2.5 mM each), and 10 pmol each primer. PCR conditions were optimized to consist of an initial denaturation of 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, 60°C annealing for 40 s, and 72°C for 1 min, with a final extension at 72°C for 5 min. In all PCR analyses, the *Populus Actin* gene (GenBank accession number XM_002316253.1) was used as an internal control. The PCR products from each sample were analyzed using 1% agarose gel electrophoresis and validated by DNA sequencing.

Subcellular Localization

To investigate the subcellular divergence, *Populus* PRXs were tested for subcellular localization using C-terminal GFP tagging. The full-length *Populus* PRX genes were amplified using the primers listed in Supplemental Table 5 and subcloned into a modified pBI121 vector (the original pBI121 vector GenBank accession AF485783; Supplemental Figure 6). Colonies containing the appropriate insert were identified by sequencing. The modified pBI121 vectors containing PRX sequences were transformed into *Agrobacterium tumefaciens* LBA4404. Cultures were infiltrated into epidermal cells of tobacco (*Nicotiana benthamiana*) leaves as described by Sparkes et al. (2006). Transformed tissues were harvested 3 to 6 d later and immediately observed with a confocal laser microscope (Olympus FV1000 MPE). GFP fusion fluorescence was excited with a 488-nm laser. Cell plasmolysis was induced by treatment with 0.3 g/mL sucrose.

Homology Modeling

To predict the tertiary structures of the *Populus* PRX proteins, the x-ray structure of an *Armoracia rusticana* PRX protein (PDB accession number 7ATJ) was used as template. Alignment of each target with the template sequence was achieved using the Align 2D program of the InsightII package. The structure of each target was automatically built using the Modeler Module. The resulting models were evaluated using the profile-3D program.

Expression and Purification of Recombinant *Populus* PRX Proteins

To investigate the enzymatic functions of the *Populus* PRXs, 10 members in clusters I and II and the MRCAP of PRX2, 3, 4, and 7 were selected for protein expression and purification (Table 3). Predictions of possible signal peptides and cleavage sites in the protein sequences were performed using the SignalP 3.1 server before gene cloning. Then, the coding sequences without the predicted N-terminal signal peptides were subcloned into the pET-30a expression vector (Novagen) using the primers listed in Supplemental Table 6. The MRCAP coding sequence was obtained using the *Populus PRX2* as a template with the primers listed in Supplemental Table 6. The methods of site-directed mutagenesis were based on our previous study (Zeng and Wang, 2005). All of the mutagenesis primers are listed in Supplemental Table 6. Colonies containing the appropriate insert were identified by sequencing. The resulting plasmids, pET-30a/PRXs, were transformed into *Escherichia coli* BL21 (DE3). The transformed *E. coli* cells were cultured at 37°C and grown until the optical density (A_{600}) reached 0.8. Then, a final concentration of 0.4 μM isopropyl-β-D-thiogalactopyranoside was added to induce synthesis of the recombinant PRX proteins. After 4 h induction, cells were harvested by centrifugation (8000*g*, 3 min, 4°C), resuspended in the buffer (20 mM Tris-HCl, pH 8.0, 2.0 M urea, 1% Triton X-100), and disrupted by cold sonication. Then, the homogenate was subjected to centrifugation (10,000*g*, 10 min, 4°C). The supernatant was discarded. The inclusion bodies were isolated, washed, and dissolved from the insoluble fractions as described by Teilum et al. (1999). Then, the dissolved PRX proteins were refolded into an active protein by diluting to a final concentration of 0.15 mg/mL in a refolding buffer containing 20 mM Tris-HCl (pH 8.5), 0.4 mM reduced glutathione, 0.6 mM oxidized glutathione, 1.5 μM hemin chloride, 6 mM $CaCl_2$, and 1.2 M urea, and incubated overnight at 4°C.

The refolded mixture was concentrated using the Amicon Ultra-15 centrifugal filter device (Millipore). The concentrated protein was desalted using a PD-10 desalting column (GE Healthcare UK) in 5 mM Tris-HCl buffer, pH 8.3. The desalted protein solution was filtered through a 0.22-μm syringe filter and subjected to a Macro-Prep High Q anion-exchange column (Bio-Rad). Then, the proteins were eluted with a linear gradient from 0 to 0.5 M NaCl in 5 mM Tris-HCl buffer, pH 8.3, over 20 column volumes. Peak fractions that contained active PRX were collected and concentrated using the Amicon Ultra-15 centrifugal filter device and desalted into 5 mM Tris-HCl, pH 8.3, via a PD-10 column.

Specific Activity and Kinetics of PRX Enzymes

PRX activity at 25°C with ascorbic acid, pyrogallol, and guaiacol were measured using the methods described by Kvaratskhelia et al. (1997), activity with ferulic acid was measured as described by Sanchez et al. (1996), activity with coniferyl alcohol was measured according to the method of Barceló and Aznar-Asensio (1999), and activity with sinapyl

alcohol was measured as described by Barceló et al. (2000). Protein concentrations were determined by measuring A_{280} (Layne, 1957).

The apparent K_m values for pyrogallol were determined using pyrogallol concentrations ranging from 2 to 20 mM and a fixed H_2O_2 concentration of 2 mM. The apparent K_m values for coniferyl alcohol were determined using coniferyl alcohol concentrations ranging from 0.02 to 0.20 mM and a fixed H_2O_2 concentration of 0.5 mM. The apparent K_m values for ferulic acid were determined using ferulic acid concentrations ranging from 0.016 to 0.160 mM at a fixed H_2O_2 concentration of 0.5 mM. The kinetic parameters were derived from nonlinear regression analysis using the Hyper32 program available at http://homepage.ntlworld.com/john.easterby/hyper32.html.

Accession Numbers

Sequence data from this study can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the accession numbers listed in Supplemental Data Set 1 for *P. trichocarpa* and *Arabidopsis*. Other sequences used in this study included the *Populus Actin* gene (GenBank accession number XM_002316253), the *A. rusticana* PRX gene (GenBank accession number AAA33377), the *O. lucimarinus* PRX gene (GenBank accession number XM_001419012), and the *C. reinhardtii* PRX gene (GenBank accession number XM_001695424).

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Summary of the *Populus* PRX Genes Expressed in Different Tissues under Normal Growth Conditions (NC) and Following H_2O_2 (HO), Salicylic Acid (SA), Salt (ST), and Drought (DR) Stress Treatments.

Supplemental Figure 2. Genomic Localization of the Full-Length *Populus* PRX Genes.

Supplemental Figure 3. Structural Comparison of Five PRX Proteins.

Supplemental Figure 4. Sequence Alignment of the *Populus* PRXs in Clusters I and II.

Supplemental Figure 5. Phylogenetic Relationships of the *Populus* and *Arabidopsis* PRXs.

Supplemental Figure 6. Modified pBI121 Vector.

Supplemental Table 1. PRX Fragments Identified from the *Populus trichocarpa* Genome.

Supplemental Table 2. PCR Primers Used to Detect the Expression of *Populus* PRX Genes.

Supplemental Table 3. K_S Values among Pairwise Comparisons of the PRX Genes in Clusters I and II.

Supplemental Table 4. Substitution Rates and ln *L* Values of the Three Optimal Branch-Specific Models Found by the Global Maximum-Likelihood Search Procedure of Zhang et al. (2011).

Supplemental Table 5. Primers Used to Construct the *Populus* PRX Subcellular Localization Vectors.

Supplemental Table 6. Primers Used to Construct the *Populus* PRX Protein Expression Vectors.

Supplemental Data Set 1. Full-length PRX Genes Identified from the *Populus trichocarpa* Genome and *Arabidopsis* PRX Protein Sequences

Supplemental Data Set 2. Text File of the Alignment Corresponding to the Phylogenetic Tree in Figure 1.

Supplemental Data Set 3. Text File of the Alignment Corresponding to the Phylogenetic Tree in Figure 2B.

Supplemental Data Set 4. Text File of the Alignment Corresponding to the Phylogenetic Tree in Figure 4.

Supplemental Data Set 5. Text File of the Alignment Corresponding to the Phylogenetic Tree in Supplemental Figure 5.

ACKNOWLEDGMENTS

We thank Chengjun Zhang of University of Chicago for assistance in branch model tests. This study was supported by grants from the Natural Science Foundation of China (NSFC 91231103 and 31330017).

AUTHOR CONTRIBUTIONS

Q.-Y.Z. conceived the project. L.-L.R., H.-J.L., T.-T.Q., and L.-W.Q. performed research. Y.-J.L. and Q.-Y.Z. analyzed data. X.-R.W. and Q.-Y.Z. wrote the article.

Received February 26, 2014; revised April 18, 2014; accepted May 24, 2014; published June 16, 2014.

REFERENCES

Barceló, A.R., and Aznar-Asensio, G.J. (1999). Coniferyl alcohol oxidase activity of a cell-wall-located class III peroxidase. Aust. J. Plant Physiol. **26:** 411–419.

Barceló, A.R., Pomar, F., and Pedreno, M.A. (2000). Competitive inhibitor-dissected histochemistry of the peroxidase responsible for syringyl lignin biosynthesis in *Zinnia elegans* xylem. Aust. J. Plant Physiol. **27:** 1101–1107.

Barkman, T.J., Martins, T.R., Sutton, E., and Stout, J.T. (2007). Positive selection for single amino acid change promotes substrate discrimination of a plant volatile-producing enzyme. Mol. Biol. Evol. **24:** 1320–1329.

Beisswanger, S., and Stephan, W. (2008). Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated *polyhomeotic* genes in *Drosophila.* Proc. Natl. Acad. Sci. USA **105:** 5447–5452.

Bindschedler, L.V., Dewdney, J., Blee, K.A., Stone, J.M., Asai, T., Plotnikov, J., Denoux, C., Hayes, T., Gerrish, C., Davies, D.R., Ausubel, F.M., and Bolwell, G.P. (2006). Peroxidase-dependent apoplastic oxidative burst in Arabidopsis required for pathogen resistance. Plant J. **47:** 851–863.

Birchler, J.A., and Veitia, R.A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. Plant Cell **19:** 395–402.

Birchler, J.A., and Veitia, R.A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol. **186:** 54–62.

Birchler, J.A., Bhadra, U., Bhadra, M.P., and Auger, D.L. (2001). Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. Dev. Biol. **234:** 275–288.

Birchler, J.A., Sun, L., Donohue, R., Sanyal, A., and Xie, W. (2011). Implications of the gene balance hypothesis for dosage compensation. Front. Biol. **6:** 118–124.

Bonawitz, N.D., and Chapple, C. (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. Annu. Rev. Genet. **44:** 337–363.

Byrne, K.P., and Wolfe, K.H. (2007). Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization

of yeast genes after whole-genome duplication. Genetics **175:** 1341–1350.

Byun-McKay, S.A., and Geeta, R. (2007). Protein subcellular relocalization: a new perspective on the origin of novel genes. Trends Ecol. Evol. (Amst.) **22:** 338–344.

Carter, C., Pan, S., Zouhar, J., Avila, E.L., Girke, T., and Raikhel, N.V. (2004). The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. Plant Cell **16:** 3285–3303.

Chen, R., Greer, A., and Dean, A.M. (1995). A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. Proc. Natl. Acad. Sci. USA **92:** 11666–11670.

Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. **9:** 938–950.

Cosio, C., and Dunand, C. (2009). Specific functions of individual class III peroxidase genes. J. Exp. Bot. **60:** 391–408.

Costa, M.M., Hilliou, F., Duarte, P., Pereira, L.G., Almeida, I., Leech, M., Memelink, J., Barceló, A.R., and Sottomayor, M. (2008). Molecular cloning and characterization of a vacuolar class III peroxidase involved in the metabolism of anticancer alkaloids in *Catharanthus roseus.* Plant Physiol. **146:** 403–417.

Daudi, A., Cheng, Z., O'Brien, J.A., Mammarella, N., Khan, S., Ausubel, F.M., and Bolwell, G.P. (2012). The apoplastic oxidative burst peroxidase in *Arabidopsis* is a major component of pattern-triggered immunity. Plant Cell **24:** 275–287.

Des Marais, D.L., and Rausher, M.D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature **454:** 762–765.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:** 1792–1797.

Edger, P.P., and Pires, J.C. (2009). Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. Chromosome Res. **17:** 699–717.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. **300:** 1005–1016.

Ferreres, F., Figueiredo, R., Bettencourt, S., Carqueijeiro, I., Oliveira, J., Gil-Izquierdo, A., Pereira, D.M., Valentão, P., Andrade, P.B., Duarte, P., Barceló, A.R., and Sottomayor, M. (2011). Identification of phenolic compounds in isolated vacuoles of the medicinal plant *Catharanthus roseus* and their interaction with vacuolar class III peroxidase: an H_2O_2 affair? J. Exp. Bot. **62:** 2841–2854.

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu. Rev. Plant Biol. **60:** 433–453.

Gabaldón, C., López-Serrano, M., Pedreño, M.A., and Barceló, A.R. (2005). Cloning and molecular characterization of the basic peroxidase isoenzyme from *Zinnia elegans*, an enzyme involved in lignin biosynthesis. Plant Physiol. **139:** 1138–1154.

Ganko, E.W., Meyers, B.C., and Vision, T.J. (2007). Divergence in expression between duplicated genes in Arabidopsis. Mol. Biol. Evol. **24:** 2298–2309.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:** 696–704.

Guindon, S., Rodrigo, A.G., Dyer, K.A., and Huelsenbeck, J.P. (2004). Modeling the site-specific variation of selection patterns along lineages. Proc. Natl. Acad. Sci. USA **101:** 12957–12962.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41:** 95–98.

Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., and Shiu, S.H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol. **148:** 993–1003.

Hatsugai, N., Iwasaki, S., Tamura, K., Kondo, M., Fuji, K., Ogasawara, K., Nishimura, M., and Hara-Nishimura, I. (2009). A novel membrane fusion-mediated plant immunity against bacterial pathogens. Genes Dev. **23:** 2496–2506.

Henriksen, A., Smith, A.T., and Gajhede, M. (1999). The structures of the horseradish peroxidase C-ferulic acid complex and the ternary complex with cyanide suggest how peroxidases oxidize small phenolic substrates. J. Biol. Chem. **274:** 35005–35011.

Hiraga, S., Yamamoto, K., Ito, H., Sasaki, K., Matsui, H., Honma, M., Nagamura, Y., Sasaki, T., and Ohashi, Y. (2000). Diverse expression profiles of 21 rice peroxidase genes. FEBS Lett. **471:** 245–250.

Huang, R., Hippauf, F., Rohrbeck, D., Haustein, M., Wenke, K., Feike, J., Sorrelle, N., Piechulla, B., and Barkman, T.J. (2012). Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. Proc. Natl. Acad. Sci. USA **109:** 2966–2971.

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and Mclnerney, J.O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6:** 29.

Kvaratskhelia, M., Winkel, C., and Thorneley, R.N. (1997). Purification and characterization of a novel class III peroxidase isoenzyme from tea leaves. Plant Physiol. **114:** 1237–1245.

Lan, T., Wang, X.R., and Zeng, Q.Y. (2013). Structural and functional evolution of positively selected sites in pine glutathione *S*-transferase enzyme family. J. Biol. Chem. **288:** 24441–24451.

Lan, T., Yang, Z.L., Yang, X., Liu, Y.J., Wang, X.R., and Zeng, Q.Y. (2009). Extensive functional diversification of the *Populus* glutathione *S*-transferase supergene family. Plant Cell **21:** 3749–3766.

Layne, E. (1957). Spectrophotometric and turbidimetric methods for measuring proteins. Methods Enzymol. **3:** 447–455.

Leissring, M.A., Farris, W., Wu, X., Christodoulou, D.C., Haigis, M.C., Guarente, L., and Selkoe, D.J. (2004). Alternative translation initiation generates a novel isoform of insulin-degrading enzyme targeted to mitochondria. Biochem. J. **383:** 439–446.

Li, Y., Kajita, S., Kawai, S., Katayama, Y., and Morohoshi, N. (2003). Down-regulation of an anionic peroxidase in transgenic aspen and its effect on lignin characteristics. J. Plant Res. **116:** 175–182.

Marques, A.C., Vinckenbosch, N., Brawand, D., and Kaessmann, H. (2008). Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol. **9:** R54.

Matsui, T., Tabayashi, A., Iwano, M., Shinmyo, A., Kato, K., and Nakayama, H. (2011). Activity of the C-terminal-dependent vacuolar sorting signal of horseradish peroxidase C1a is enhanced by its secondary structure. Plant Cell Physiol. **52:** 413–420.

Morley, K.L., and Kazlauskas, R.J. (2005). Improving enzyme properties: when are closer mutations better? Trends Biotechnol. **23:** 231–237.

Ohno, S. (1970). Evolution by Gene Duplication. (New York: Springer).

Ostergaard, L., Teilum, K., Mirza, O., Mattsson, O., Petersen, M., Welinder, K.G., Mundy, J., Gajhede, M., and Henriksen, A. (2000). *Arabidopsis* ATP A2 peroxidase. Expression and high-resolution structure of a plant peroxidase with implications for lignification. Plant Mol. Biol. **44:** 231–243.

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

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

Qian, W., and Zhang, J. (2009). Protein subcellular relocalization in the evolution of yeast singleton and duplicate genes. Genome Biol. Evol. **1:** 198–204.

Sakuma, S., Pourkheirandish, M., Hensel, G., Kumlehn, J., Stein, N., Tagiri, A., Yamaji, N., Ma, J.F., Sassa, H., Koba, T., and Komatsuda, T. (2013). Divergence of expression pattern contributed to neofunctionalization of duplicated HD-Zip I transcription factor in barley. New Phytol. **197:** 939–948.

Sanchez, M., Pena, M.J., Revilla, G., and Zarra, I. (1996). Changes in dehydrodiferulic acids and peroxidase activity against ferulic acid associated with cell walls during growth of *Pinus pinaster* hypocotyl. Plant Physiol. **111:** 941–946.

Shiu, S.H., Byrnes, J.K., Pan, R., Zhang, P., and Li, W.H. (2006). Role of positive selection in the retention of duplicate genes in mammalian genomes. Proc. Natl. Acad. Sci. USA **103:** 2232–2236.

Sinha, N., and Smith-Gill, S.J. (2002). Protein structure to function via dynamics. Protein Pept. Lett. **9:** 367–377.

Smirnoff, N. (1996). The function and metabolism of ascorbic acid in plants. Ann. Bot. (Lond.) **78:** 661–669.

Smith, S.D., Wang, S., and Rausher, M.D. (2013). Functional evolution of an anthocyanin pathway enzyme during a flower color transition. Mol. Biol. Evol. **30:** 602–612.

Sparkes, I.A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat. Protoc. **1:** 2019–2025.

Teilum, K., Ostergaard, L., and Welinder, K.G. (1999). Disulfide bond formation and folding of plant peroxidases expressed as inclusion body protein in *Escherichia coli* thioredoxin reductase negative strains. Protein Expr. Purif. **15:** 77–82.

Tognolli, M., Penel, C., Greppin, H., and Simon, P. (2002). Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana.* Gene **288:** 129–138.

Tuskan, G.A., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science **313:** 1596–1604.

Wang, X., Huang, Y., Lavrov, D.V., and Gu, X. (2009). Comparative study of human mitochondrial proteome reveals extensive protein subcellular relocalization after gene duplications. BMC Evol. Biol. **9:** 275.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24:** 1586–1591.

Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A.M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431–449.

Yang, Z.L., Liu, H.J., Wang, X.R., and Zeng, Q.Y. (2013). Molecular evolution and expression divergence of the *Populus* polygalacturonase supergene family shed light on the evolution of increasingly complex organs in plants. New Phytol. **197:** 1353–1365.

Zeng, Q.Y., and Wang, X.R. (2005). Catalytic properties of glutathione-binding residues in a tau class glutathione transferase (PtGSTU1) from *Pinus tabulaeformis.* FEBS Lett. **579:** 2657–2662.

Zhang, C., Wang, J., Xie, W., Zhou, G., Long, M., and Zhang, Q. (2011). Dynamic programming procedure for searching optimal models to estimate substitution rates based on the maximum-likelihood method. Proc. Natl. Acad. Sci. USA **108:** 7860–7865.

Zhang, J., and Rosenberg, H.F. (2002). Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. Proc. Natl. Acad. Sci. USA **99:** 5486–5491.

Zhang, J., Dean, A.M., Brunet, F., and Long, M. (2004). Evolving protein functional diversity in new genes of *Drosophila.* Proc. Natl. Acad. Sci. USA **101:** 16246–16250.

Subcellular Relocalization and Positive Selection Play Key Roles in the Retention of Duplicate Genes of *Populus* Class III Peroxidase Family

Lin-Ling Ren, Yan-Jing Liu, Hai-Jing Liu, Ting-Ting Qian, Li-Wang Qi, Xiao-Ru Wang and Qing-Yin Zeng

Plant Cell 2014;26;2404-2419; originally published online June 16, 2014;
DOI 10.1105/tpc.114.124750