Comparative analyses of pathogenesis-related protein-10 (PR10) in plants

Ibrahim Ilker Ozyigit¹, Recep Vatansever¹, Ertugrul Filiz²*

¹Marmara University, Faculty of Science and Arts, Department of Biology, 34722, Goztepe, Istanbul, Turkey, ²Duzce University, Department of Crop and Animal Production, Cilimli Vocational School, 81750, Cilimli, Duzce, Turkey

Received 3 March 2015; revised 1 September 2015; accepted 8 September 2015

In the present study, we have comparatively analyzed PR10 genes and proteins from 28 plant species in order to understand the relationship (conservation or divergence) between different PR10s in various plant species. In analyzed species, PR10 proteins were found to be small (157-166 aa long and 14.3-18.2 kDa weight) and acidic (4.69-6.17) in nature. Besides, PR10 sequences had highly conserved GxGxG motif (P-loop motif) structure at various positions. These positional variations in glycine (Gly) residues may become the result of substitution, deletions and insertions occurred during the course of PR10 evolution. In general, primary sequences of PR10s in various plant species may have a well conserved structure. Digital expression data of tomato and maize showed that expression of PR10 genes may significantly increase in plant parts (root, lateral root and root tips) where it is more open to the mechanical perturbation and pathogenic attack, supporting the involvement of PR10 in plant defense. In phylogenetic tree, a clear monocot/polycot and dicot separation were observed. This separation could have been arising from the well conserved structure of PR10 genes of monocots and polycots than dicots. All modeled species contained the same number of β-strands (seven) but α-helices varied between 2 and 4 depending on species. The results of this study will provide a theoretical reference regarding the primary, secondary and tertiary structure of PR10s in various plant species and will support the future studies that aiming to characterize the pathogenesis-related (PR) proteins.

**Key words:** In silico, pathogenesis-related protein 10, 3D structure, phylogenetic, polycot, divergence

**Introduction**

Pathogenic infection causes to induce various genes at the infection site and/or in other parts of the plant, which subsequently leads to developing a hypersensitive reaction (HR) and systemic acquired resistance (SAR)¹-². Pathogenesis-related (PR) genes are some of the first genes that overexpressed in different parts of the plants during an infection or under some stress-related conditions. Products of these genes comprise the very diverse group of proteins, including chitinases, glucanases, endoproteinases, peroxidases, proteinase inhibitors, osmotins, defensins, thionins and lipid transfer proteins (LTPs)³. The PR proteins are grouped into 17 families based on their primary structure, serological relationships and biological activities⁴-⁵. PR1, PR2, PR3 and PR4 have been characterized as β-1,3-glucanases, PR6 as proteinase inhibitors, PR9 as chitinases and PR11 as peroxidases, PR12 as defensins, PR13 as thionins, PR14 as lipid transfer proteins, and PR15 as oxalate oxidases, but precise biological functions of PR5, PR10, PR16 and PR17 remain unclear⁶-⁷. PR10 proteins, which was first identified as a major white birch pollen allergen (Bet_v_1)⁸, are small (15–19 kDa) acidic intracellular proteins and comprise the largest family of PRs with more than its 100 members⁹-¹⁰. They are induced by pathogenic attack¹¹-¹², extreme abiotic factors² or even by normal physiological changes in many plant species. Methyl jasmonate (MeJA), salicylic acid (SA), gibberellic acid (GA3), hydrogen peroxide (H₂O₂), sodium chloride (NaCl) and ethylene are reported to induce the transcription of PR10 genes in different plant species¹³-¹⁴. In addition, PR10 proteins have been demonstrated to exhibit ribonuclease activity and to have cytokinin-binding ability in various plant species¹⁵-¹⁶.

Although PR10 proteins are mainly known to be involved in plant defense under biotic/abiotic stress conditions, the overall biological function of many PR10 members still awaits to be elucidated. In the present study, we have comparatively analyzed the PR10 genes and proteins in 28 plant species in order to understand the evolutionary (conservation or divergence) relationship between these genes/proteins in various plants. The results of this study will provide a theoretical reference for the future studies.

*Author for correspondence
ertugrulfiliz@gmail.com
Materials and Methods

Sequence Retrieval, Domain and Interaction Partner Analysis of PR10s

The PR10 protein sequences were retrieved from the protein database of NCBI (http://www.ncbi.nlm.nih.gov/protein/). The query against the plant protein database resulted in 28 plant species, including 20 dicots, 6 monocots and 2 polycots after the partial, redundant and ambiguous sequences were removed (Table 1). Then, domain analysis was performed by Pfam server for retrieved sequences in order to confirm Bet_v_1 (PF00407) domain structure (http://pfam.sanger.ac.uk).17. The interaction partners of PR10 proteins were predicted using STRING 9.1 server (http://string-db.org).18.

Physicochemical and Conserved Motif Analysis of PR10s

Sequence length, molecular weight, and theoretical isoelectric point (pI) of proteins were analyzed by Expasy’s ProtParam server (http://web.expasy.org/protparam/). The conserved motif analysis was performed by MEME (Multiple Em for Motif Elicitation) tool20 (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) with following parameters: distribution of motifs, 0 or 1 per sequence; maximum number of motifs to find 5; minimum width of motif, 50.

Table 1 — Physicochemical properties and predicted sub-cellular localization of PR10s in 28 plant species

<table>
<thead>
<tr>
<th>Species name</th>
<th>Groups</th>
<th>Access. No. (NCBI)</th>
<th>Length (aa)</th>
<th>MW (kDa)</th>
<th>pI</th>
<th>Predicted subcellular localization</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Pisum sativum</em></td>
<td>Dicot</td>
<td>AAA9054</td>
<td>158</td>
<td>16.75</td>
<td>4.94</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Gossypium barbadense</em></td>
<td>Dicot</td>
<td>ACM17134</td>
<td>159</td>
<td>17.31</td>
<td>4.95</td>
<td>Cytoplasm/Chloroplast</td>
</tr>
<tr>
<td><em>Capsicum annumum</em></td>
<td>Dicot</td>
<td>ACB30364</td>
<td>166</td>
<td>18.25</td>
<td>5.95</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Arachis hypogaea</em></td>
<td>Dicot</td>
<td>AUA81922</td>
<td>150</td>
<td>16.22</td>
<td>5.34</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Nicotiana tabacum</em></td>
<td>Dicot</td>
<td>AYE11296</td>
<td>160</td>
<td>17.77</td>
<td>5.56</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Prunus domestica</em></td>
<td>Dicot</td>
<td>ABW9634</td>
<td>160</td>
<td>17.66</td>
<td>5.79</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Prunus persica</em></td>
<td>Dicot</td>
<td>ABW9628</td>
<td>160</td>
<td>17.65</td>
<td>5.79</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Capsicum baccatum</em></td>
<td>Dicot</td>
<td>ABC74797</td>
<td>159</td>
<td>17.24</td>
<td>5.21</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Lilium regale</em></td>
<td>Monocot</td>
<td>AHC68874</td>
<td>157</td>
<td>16.84</td>
<td>5.39</td>
<td>Chloroplast/Cytoplasm</td>
</tr>
<tr>
<td><em>Vicia faba</em></td>
<td>Dicot</td>
<td>AFD2983</td>
<td>133</td>
<td>14.34</td>
<td>5.53</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Crocus sativus</em></td>
<td>Monocot</td>
<td>ADL09408</td>
<td>162</td>
<td>17.40</td>
<td>6.08</td>
<td>Chloroplast/Chloroplast</td>
</tr>
<tr>
<td><em>Triticum aestivum</em></td>
<td>Monocot</td>
<td>ACC68733</td>
<td>160</td>
<td>17.06</td>
<td>5.19</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Vitis pseudoreticulata</em></td>
<td>Dicot</td>
<td>ABC86747</td>
<td>159</td>
<td>17.47</td>
<td>6.07</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Pinus pinaster</em></td>
<td>Polycot</td>
<td>ADJ53040</td>
<td>160</td>
<td>17.80</td>
<td>6.17</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Sorghum bicolor</em></td>
<td>Monocot</td>
<td>AAW83209</td>
<td>160</td>
<td>16.95</td>
<td>5.19</td>
<td>Chloroplast</td>
</tr>
<tr>
<td><em>Capsicum chinense</em></td>
<td>Dicot</td>
<td>CAI51309</td>
<td>159</td>
<td>17.31</td>
<td>4.81</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Lupinus albus</em></td>
<td>Dicot</td>
<td>CAO3926</td>
<td>158</td>
<td>17.92</td>
<td>4.83</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Tanacetum cinerariifolium</em></td>
<td>Dicot</td>
<td>AEL17175</td>
<td>157</td>
<td>17.18</td>
<td>5.79</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Pinus monticola</em></td>
<td>Polycot</td>
<td>AAL50007</td>
<td>161</td>
<td>17.80</td>
<td>5.34</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Pisum fulvum</em></td>
<td>Dicot</td>
<td>AAB07447</td>
<td>158</td>
<td>16.72</td>
<td>4.94</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Solanum lycopersicum</em></td>
<td>Dicot</td>
<td>AHC08074</td>
<td>160</td>
<td>17.37</td>
<td>5.44</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Ziziphus jujuba</em></td>
<td>Dicot</td>
<td>AGL07712</td>
<td>160</td>
<td>17.51</td>
<td>5.40</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Medicago truncatula</em></td>
<td>Dicot</td>
<td>XP_003594834</td>
<td>158</td>
<td>16.81</td>
<td>4.76</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Fragaria chiloensis</em></td>
<td>Dicot</td>
<td>ADN05762</td>
<td>157</td>
<td>17.27</td>
<td>5.52</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Glycine max</em></td>
<td>Dicot</td>
<td>XP_006582821</td>
<td>158</td>
<td>16.77</td>
<td>4.69</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Elaeis guineensis</em></td>
<td>Monocot</td>
<td>AEB96227</td>
<td>160</td>
<td>17.06</td>
<td>5.19</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Rheum australe</em></td>
<td>Dicot</td>
<td>ACH63224</td>
<td>160</td>
<td>17.81</td>
<td>5.87</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td><em>Oryza sativa</em></td>
<td>Monocot</td>
<td>BAD03969</td>
<td>160</td>
<td>16.90</td>
<td>4.88</td>
<td>Cytoplasm</td>
</tr>
</tbody>
</table>
Gene Structure and Expression Profile Analysis of PR10s

Exon/intron organization of PR10 genes were analyzed by GSDS 2.0 (Gene Structure Display Server) (http://gsds.cbi.pku.edu.cn/)22. Digital expression profile of PR10 genes were obtained from Genevestigator platform23 (https://www.genevestigator.com/gv/plant.jsp).

Phylogenetic Analysis of PR10s

PR10 protein sequences were aligned by ClustalW24 and phylogenetic tree was constructed by MEGA version 5.125 with following parameters; maximum likelihood (ML) method, Poisson correction, pairwise deletion and 1000 replicates bootstrap value.

3D Modeling of PR10s

Binding sites of PR10 proteins were predicted by 3DLigandSite server (http://www.sbg.bio.ic.ac.uk/3dligandsite/)26, and visualized by PyMOL27. Structural evaluation and stereochemical analyses were done with Rampage Ramachandran plot analysis (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php)28.

Results and Discussion

Sequence and Physico-chemical Analysis of PR10 Proteins

PR10 protein sequences were confirmed to contain Bet_v_1 (PF00407) domain structure, and found to have a close molecular weight (14.3-18.2 kDa but mainly 17-18 kDa) and amino acid length (157-166 aa), with a pI value of 4.69-6.17. The sub-cellular localization of PR10 proteins was predicted as cytoplasmic (Table 1). PR10s are reported to be small (17.1-18.4 kDa) acidic intra-cellular proteins10. This complies with the physico-chemical properties of PR10 proteins retrieved for this study.

The most conserved five motifs of PR10 proteins were detected. Motif 1 was observed in 24 of 28 species; motif 2 was observed in 25 of 28 species; motif 3 was observed in 26 of 28 species; motif 4 was observed in 27 of 28 species and motif 5 was present in all species. Only motif 1 was found to be related with the Bet_v_1 domain structure (Fig. 1 and Table 2). P-loop structure with an amino acid sequence GXGGXG in birch pollen Betv1 resembles...
OZYIGIT et al: COMPARATIVE ANALYSIS OF PATHOGENESIS RELATED PROTEIN -10 IN PLANTS

We found a similar motif with GNGGPG pattern in all species, except for Rheum australe. Overall, variations at different motif structures might be caused by deletion, insertion etc., causing the formation of new catalytic sites in PR10 family in terms of improving the survivability under the various biotic and abiotic stress conditions.

Putative retrotransposon protein (Ty1-copia subclass), endo-1,3; 1,4-β-D-glucanase precursor and ANTH/ENTH domain containing protein were predicted to be main interaction partners of rice PR10 protein (Fig. 2). Retrotransposons are mobile genetic elements affecting genome structure, function and evolution in eukaryotic organisms30. 1,3;1,4-β-glucanase type I (E.C. 3.2.1.73; glycosyl hydrolase family 17) evolved specifically in the grasses that might control growth responses in plants31. Many ANTH Sla2/HIP1 and ENTH Ent1-2/Epsin1, 2, 3 domains play role as lipid binding domains, functioning in phosphoinositide enrichment at the target membranes, membrane trafficking, and in metabolism and cytokinesis32. Interaction partners of rice PR10 proteins implicate that PR10 could be involved in a diverse metabolic pathways in plants.

Exon/intron Structure and Expression Profiles of PR10 Genes

Gene structure (exon/intron) analysis provides valuable information for understanding of genetic and evolutionary relationship between inter-and intra-species33,34. Exon-intron analysis was performed for six selected species, including G. max, M. truncatula, O. sativa, P. persica, S. lycopersicum and S. bicolor that have complete genome sequenced, thereby genomic DNA and CDS were available for exon-intron deduction. The analysis in six selected species showed that PR10 genes have similar exon/intron patterns (two exons) without any monocot/dicot separation (Fig. 3). In order to see how exons are related to each other, we obtained the nucleotide sequences of these potential two exons from the JBrowser of Phytozome and aligned by ClustalW (Fig. 4). Notably, some nucleotide residues in alignment were found to be highly conserved in both

Table 2 — The most conserved five motifs of PR10s in 28 plant species. Red residues in motif 4 indicate the phosphate-binding loop (P-loop; GXGGXG) (Saraste et al. 1990)

<table>
<thead>
<tr>
<th>Motif number</th>
<th>Width</th>
<th>Sequence</th>
<th>Pfam domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>FKYMHRIDFEEDENCVNYSLIEGDGLGD</td>
<td>Bet v I family</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>WEHEITSPVAPARLWKMMDWHLWPKLW</td>
<td>Not found</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>EEQHKQGKERADGLFLKAIEQYCLANPDYY</td>
<td>Not found</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>SIEIVGNGGPGTIRQMNFVE</td>
<td>Not found</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>EARPNNGCICKWTCHYHTKGD</td>
<td>Not found</td>
</tr>
</tbody>
</table>

Fig. 2 — Predicted interaction partners of rice PR10 protein

a motif found in many nucleotide-binding proteins29. We found a similar motif with GNGGPG pattern in all species, except for Rheum australe. Overall, variations at different motif structures might be caused by deletion, insertion etc., causing the formation of new catalytic sites in PR10 family in terms of improving the survivability under the various biotic and abiotic stress conditions.
exons, indicating that these nucleotide residues may encode the same functional residues of the proteins. Overall, this shows that PR10 genes might have been well preserved along the divergence of monocots and dicots during PR10 gene evolution in plants.

Digital gene expression profile of PR10 genes were analyzed for selected two species (tomato and maize) at developmental and anatomical part levels (Fig. 5). Tomato and maize represent for dicots and monocots, respectively. Tomato and maize genes were evaluated within seven developmental stages. Tomato gene (Solyc09g090980.2) was highly expressed in the early stage of flowering and fruit ripening (Fig. 5A). In maize, expression level of GRMZM2G075283 and GRMZM2G112524 genes relatively increased in seed germination period and GRMZM2G112524 was found to be highly expressed during seed development period (Fig. 5B). The reason of increased gene expression level in seed germination, flowering, fruit ripening and seed development could be related with increased plant hormones because plant PR10 proteins are known as more properly linked with general phytohormone-binding proteins (PhBP)35-37. At anatomical level, tomato gene (Solyc09g090980.2) was observed to mainly express in lateral root, root, exocarp (skin), pericarp and fruit, and the highest expression levels were found in lateral root and exocarp (skin) (Fig. 5C). In maize, GRMZM2G075283 gene was detected to be highly expressed in glume while GRMZM2G112524 gene was highly expressed in caryopsis (Fig. 5D). Also, GRMZM2G075283 and GRMZM2G112524 genes were observed to moderately be expressed in roots and lateral roots. This shows that PR10 gene
expression significantly increases in the parts (root, lateral root and root tips) of plant where it is more open to the mechanical perturbation and pathogenic attack.

**Phylogenetic Analysis of PR10s**

Phylogenetic tree was constructed for PR10 protein sequences of 28 species. First, constructed phylogenetic tree was divided into two main groups, namely as A and B. Then, group A and B were further subdivided into subgroups A1 and A2, and B1 and B2, respectively (Fig. 6). All monocots and polycots were observed to be clustered in group B (B1 and B2) with 67% bootstrap value while all dicots were in group A (A1 and A2) with 61% bootstrap value. In dicot group, *R. australe* (A2) was separated from other dicots and in monocot group, *O. sativa* (B1) was diverged from other monocots. A distinction between monocots/polycots (67% bootstrap value) and dicots (61% bootstrap value) could be explained with the well conserved structure of PR10 genes in monocots and polycots.

**3D Structure Prediction of PR10s**

*G. max*, *M. truncatula*, *O. sativa*, *P. persica*, *S. lycopersicum*, and *S. bicolor* plants selected for 3D modeling (Fig. 7). Selection based on the phylogenetic clustering of these species (Fig. 6). *O. sativa* and *S. bicolor* selected as representatives of monocots while *G. max*, *M. truncatula*, *P. persica* and *S. lycopersicum* selected as representatives of dicot species. In models, seven β-strands and three α-helices in *G. max*, *S. lycopersicum*, and *S. bicolor*; seven β-strands and four α-helices in *M. truncatula* and *O. sativa*; and seven β-strands and two α-helices in *P. persica* were predicted. Notably, all six models contained the same number of β-strands (seven). In 3D models, phosphate-binding loop (P-loop; GXGGXG), which is a strongly conserved region of nucleotide-binding proteins, were identified. P-loop is located at positions Gly46–48–49–51 in *G. max* and *M. truncatula*; Gly44–46–47–49 in *O. sativa*; Gly47–49–50–52 in *P. persica* and *S. lycopersicum* and Gly48–50–51–53 in *S. bicolor*. Positional variations in glycine (Gly) residues in P-loop motif suggest that mutations, deletions and insertions may cause the variations within PR10 genes during the course of evolution.

Protein-ligand binding sites have a vital importance in understanding the functional diversity among species. Thus, predicted binding sites were identified in PR10 proteins. These are; 23Phe-24Val-27Ile-39Leu-57Ile-59Phe-70His-82Tyr-84Tyr-101Ile-144Leu in *P. persica*; 23Ala-25Val-28Trp-56Arg-58Phe-60Phe-69Met-71Glu-141Val-144Ile in *S. bicolor*; 23Leu-24Val-27Phe-55Lys-57Met-59Phe-65Ile-68Leu-70His-84Tyr-91Val-101Tyr-141Val-144Leu in *S. lycopersicum*. 

---

**Fig. 5** — The digital expression profile analysis of tomato and maize PR10 genes by using gene investigator search engine. Heat map shows the expression profiles in different developmental stages (5A and 5B, respectively) of tomato and maize at different anatomical parts (5C and 5D, respectively).
Fig. 7 — The predicted 3D structure and putative active site of PR10 proteins generated by 3DLigandSite server. The green color indicates the predicted active site position of PR10s.

and 23Phe-24Ser-26Ala-54Met-66Phe-89Ala-92Lys-141Tyr-144Met in *O. sativa* (Fig. 8). Generally, in proteins, solvent-exposed residues show variational tendencies due to mutations\(^3\), and amino acid residues at such binding sites represent higher degree of variations\(^4\). Therefore, residues at predicted binding sites in PR10 proteins might be connected with deleterious or insertional mutations etc., leading to the formation of new catalytic sites with different functions in diverse metabolic pathways.

Model validation was done by Ramachandran plot analysis. It was found that 97.4%, 98.1%, 91.8%, 92.4%, 98.1% and 87.8% of residues were in favored region; 2.6%, 1.9%, 5.1%, 7.6%, 1.9% and 8.3% of residues were in allowed region and 0.0%, 0.0%, 3.2%, 0.0%, 0.0%, and 3.8% of residues were in outlier region in *G. max*, *M. truncatula*, *O. sativa*, *P. persica*, *S. lycopersicum*, and *S. bicolor*, respectively. This indicates that our 3D models were fairly in good quality.

Overall, *PR* genes are highly expressed in different anatomical parts of the plants under various biotic/abiotic stress conditions. Products of these genes comprise the very diverse group of proteins with various metabolic functions. However, we still have limited knowledge about many...
PR proteins of different plant species. Therefore, to have an insight about the structural and functional roles of PR proteins, we have comparatively analyzed the PR10 genes and proteins from 28 different plant species. Analyses revealed that PR10 proteins are structurally well conserved and functionally involve in plant defense. However, to better understand the physiological response of PR10 proteins in plant defense, further genome-wide analyses are required with various stress perturbations in various plant species. We believe that the results of this study will provide a theoretical reference regarding the primary, secondary and tertiary structure of PR10s in various plant species and will support the future studies that aims to characterize the pathogenesis-related (PR) proteins.

References