RESEARCH NOTE

Computational identification of novel *PR-1*-type genes in *Oryza sativa*

QINGPO LIU and QINGZHONG XUE*

Department of Agronomy, College of Agriculture and Biotechnology, Zhejiang University, 310029 Hangzhou, China

Introduction

Plants are often threatened by pathogen attack or other external stresses during growth and development. To decrease loss in productivity, plants respond to biotic and abiotic stresses by triggering a series of biochemical reactions, termed hypersensitive reaction (HR), which is associated with synthesis of secondary metabolites, cell wall rigidification, lignin and suberin deposition, and induction of a variety of novel proteins. In the HR response, pathogenesis-related proteins (PRs) that are suggested to be effective in inhibiting growth, multiplication and/or spread of pathogens, and are responsible for the state of systemic acquired resistance (SAR; Van Loon 1997), are significantly induced and accumulated abundantly in host plants (Van Loon and Van Strien 1999).

PR proteins can be divided into 17 groups on the basis of their sequence characteristics, immunological relationships and biological activities (Liu *et al*. 2005). However, only 14 distinct families of PR proteins have been found in plants (Sarowar *et al*. 2005). Not all families are found in any one plant species (Van Loon and Van Strien 1999). Among the diverse PRs, PR-1 is the dominant group, whose function is not well known. The limited antifungal activity suggests function of PR-1 in plant defence (Kitajima and Sato 1999). However, the mode of action and its relationship to other types of proteins still remain unknown. The first PR-1 protein was discovered in 1970. Since then, a number of PR-1 proteins have been identified in *Arabidopsis* (Uknes *et al*. 1992), maize (Morris *et al*. 1998), wheat (Molina *et al*. 1999), tomato (Tornero *et al*. 1994), tobacco (Pfitzner and Goodman 1987), barley (Muradov *et al*. 1993), rice (Agrawal *et al*. 2000a,b; Kim *et al*. 2001) and pepper (Kim and Hwang 2000).

Rice is a vitally important crop for human consumption. The completion of the *Oryza sativa* genome (International Rice Genome Sequencing Project 2005) has made it possible to identify all the PR-1 family members in this plant species at the genome level. The present study is the first to provide the list of rice *PR-1* genes, which will be beneficial for studying *PR-1* gene expression patterns and further investigating the mechanism or mode of plant self-defence in rice.

Materials and methods

Collection of PR-1 genes

Several rice, *Arabidopsis*, maize and tobacco PR-1 amino acid sequences (accession numbers: U89895, AJ278436, M90508, U82200, Q00008, X06930) were obtained from the GenBank database. To obtain all of the PR-1s, the rice amino acid sequences collected in RGP (International Rice Genome Sequencing Project 2005), TIGR and NCBI were downloaded to construct a local rice protein database.

With the two rice PR-1 amino acid sequences as query, a PSI-BLAST was seeded to search the local protein database with e-value 10. In the second protocol, the multiple-sequence-aligned six cloned PR-1s were used to construct the HMM profile, and then used to search the in-house rice protein database, using software HMMER 2.3.2 (http://hmmer.wustl.edu/). In addition, the six amino acid sequences were also used to search, using the TBLASTN program, the complete rice genomic DNA database in RGP, and the whole-genome shotgun scaffold sequences of *indica* and *japonica* rice in NCBI. The program FgeneSH was used for gene prediction (http://www.softberry.com/). Redundant hits were removed by manual inspection.

Analysis of sequence characteristics

Signal peptide was predicted using the SignalP server (http://www.cbs.dtu.dk/services/SignalP/). The program InterProScan (http://www.ebi.ac.uk/InterProScan/) was used for

*For correspondence. E-mail: qzhxue@zju.edu.cn.

Keywords. rice; pathogenesis-related proteins; PR-1; phylogenetic analysis.

detecting conserved domains in rice PR-1 candidates. Prediction of protein secondary structure was performed by employing the PHD method (http://cubic.bioc.columbia.edu/predictprotein/).

Supports of EST and full-length cDNA sequences

A total of 406,790 rice EST sequences were downloaded from the GenBank database (release 102105). The coding region (CDS) of each *PR-1* candidate was used to perform BLASTN search against all of the ESTs. Matches with 95% identity and over 100 bases or more were considered as significant hits.

Furthermore, the cDNA sequences in NCBI and full-length cDNA sequence database in KOME (Kikuchi *et al*. 2003) were also searched, with the amino acid sequences of rice candidates. If the hits showed 100% identity, and over the entire polypeptide, it was considered as a positive, and to be an expressed gene in rice.

Multiple-sequence alignment and construction of phylogenetic tree

Multiple-sequence alignment of candidate PR-1 amino acid sequences was performed using the Clustal W program (http://www.molecularevolution.org/software/clustalw/). A phylogenetic tree was constructed by employing the neighbour-joining method wrapped in the PHYLIP software suite (http://evolution.genetics.washington.edu/phylip.html). The tree was viewed with MEGA3 (Kumar *et al*. 2004).

Results

Identification of rice PR-1-type genes

The PR-1-type proteins have specific structure characteristics (Van Loon and Van Strien 1999). First, the primary translation products of the *PR-1* genes contain a hydrophobic signal sequence. Second, the mature PR-1 proteins contain six conserved cysteine residues forming disulphide bridges. Third, the PR-1 proteins contain four α-helices and four β-strands. Accordingly, after carefully surveying the rice genome, 23 genes were defined as rice *PR-1*-type gene candidates, among which 21 were being reported for the first time (see figure 1, table 1). In addition, an InterProScan search was seeded, and conserved domains such as 'Allergen V5/Tpx-1 related', 'Ves allergen', and PR-1 family of extracellular SCP domains were detected. Notably, 'PATHOGENESIS-RELATED PROTEIN 1' domain was found to be in OsPR1-1, OsPR1-2, OsPR1-12 and OsPR1-17, which strongly suggested that these genes belong to the rice *PR-1* family.

Of the 23 computationally predicted OsPR1s, at least 11 had one significant EST hit. Moreover, seven OsPR1 genes were found to have their counterpart mRNA/cDNA sequences in the GenBank/KOME databases. Interestingly, although no significant EST hits were found in OsPR1-5 and OsPR1-23, there were two sequenced cDNAs in KOME that matched them quite well respectively (table 1). These results suggested that more than half of the predicted *PR-1* family members are expressed in the rice genome.

Chromosomal localization

The distribution of *PR-1* genes on rice chromosomes was extremely nonuniform. These genes were mainly located on chromosomes 1, 2, 4, 5, 6, 7 and 10 (figure 2). Compared with the only one *PR-1* lying on chromosomes 5, 6 and 10, about half of the *PR-1* genes were closely clustered on chromosome 7.

Phylogenetic analysis

Figure 3 shows a phylogenetic tree of PR-1 amino acid sequences. In this tree, PR-1 proteins in *Arabidopsis*, tobacco, pepper, maize, wheat and barley were selected as reference data. In addition, the *Schizophyllum commune* PR-1 protein (pSc7) was used as the outgroup sequence.

The phylogenetic tree shows that rice PR-1 proteins could be further divided into three distinct groups (figure 3). Group I was mainly composed of PR-1 proteins of dicot species with the exception of OsPR1-20 and OsPR1-21. The classification of second and third groups was based on the two types of PR-1 isoforms: group II is for acidic PR-1 whereas group III is for the basic form, suggestive of different evolutionary tracks of the two kinds of PR-1 proteins.

Notably, the last two groups were mostly monocot species, except for pSc7. This clearly different classification of dicot and monocot species suggested that the divergence of PR-1 proteins was subsequent to the divergence of dicotyledonous and monocotyledonous plants. However, the outgroup sequence pSc7 was well clustered with plant PR1s. In addition, rice OsPR1-20 and OsPR1-21 showed high homology to dicot PR1s. These results indicated that PR-1 proteins evolved from a common remote ancestor.

Discussion

Gene structure analysis showed that rice *PR-1*-type genes have no introns in their coding region (data not shown). Typically, a gene with no introns is considered a putative pseudogene (Lee *et al*. 1983). Thus, compared with the diverse gene structures of other plant families, little is known about the biological significance of only one exon in members of the rice *PR-1* family.

In the secondary structure of protein p14a, the four β-strands are arranged in antiparallel between helices I, III and IV and II (αI-βA-αII-βB-αIII-αIV-βC-βD; Fernández *et al*. 1997). However, the range order of helices and strands of rice PR-1 proteins (αI-αII-αIII-βA-αIV-βB-βC-βD) is apparently different from that of p14a. Nonetheless, study of wheat PR-1 gave the same conclusion as that in rice (Yu *et al*. 2001).

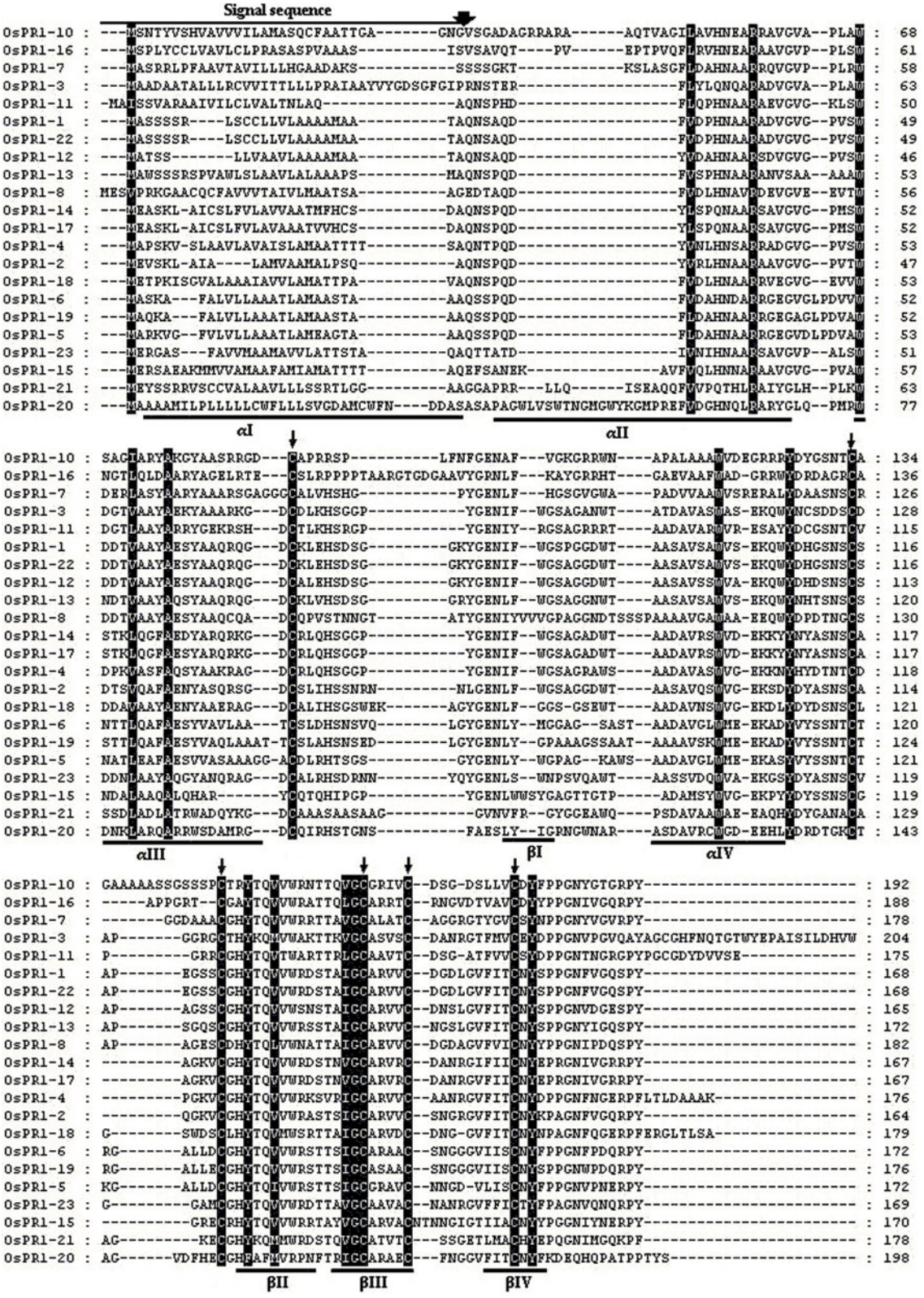

Figure 1. Amino acid swquence alignment of rice PR-1-type proteins. Alignment was performed using the Clustal W program. The positions of the cleavage site between the signal peptide and mature protein, and the six conserved cysteine residues, are indicated by thick and thin arrowhead respectively. The four α-helices and four β-strands are marked.

Table 1. List of *PR-1*-type genes in rice.

PR-1	Cloned PR-1	Protein ID	AA length	Chromosome	cDNA	EST number
OsPR1-1	OsPR1a	CAC03571	168	7	AJ278436	4
OsPR1-2	OsPR1b	AAB49685	164	1		5
OsPR1-3		BAD19616	204	2		0
OsPR1-4		AAP52566	176	10		9
OsPR1-5		BAC84827	172	7	AK062949	0
OsPR1-6		BAC56823	172	7		2
OsPR1-7		BAD19210	178	2		0
OsPR1-8		BAC84836	182	7		0
OsPR1-9		BAC22534	176	7		0
OsPR1-10		CAD40250	192	4		0
OsPR1-11		BAD34031	175	6		0
OsPR1-12		AAM45439	165	7	AF306651	7
OsPR1-13		BAC84837	172	7		1
OsPR1-14		BAB84473	167	1	AK121108	6
OsPR1-15		BAC84831	170	7		0
OsPR1-16		CAE02369	188	4		0
OsPR1-17		Predicted	167	1		6
OsPR1-18		BAC56830	179	7	AK060057	3
OsPR1-19		BAC84830	176	7		0
OsPR1-20		XP_476031	198	5	AK071326	1
OsPR1-21		BAD19213	178	2		0
OsPR1-22		BAC84842	168	7		4
OsPR1-23		BAC84818	169	7	AK060005	0

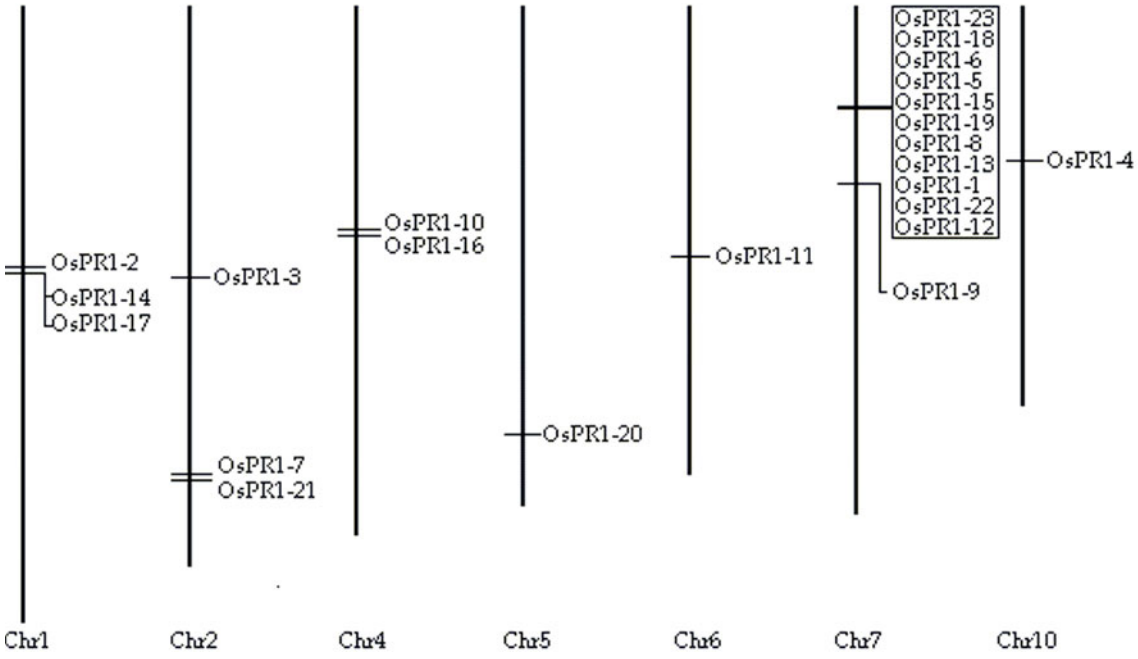

Figure 2. Chromosomal distribution of rice *PR-1*-type genes. The relative sizes of rice chromosomes were derived from RGP.

PR-1-type genes are commonly used as markers for SAR (Van Loon and Van Strien 1999). The computationally identified 23 rice *PR-1*-type genes would be potentially important markers in studying responses in rice against stress and disease. Importantly, the present study localized the 23 *PR-1*-type genes onto seven chromosomes, which would be significant in establishing *in vitro* studies. Notably, one possible explanation for rice *PR-1* genes preferentially clustering on chromosome 7 would be the occurrence of tandem gene duplication events.

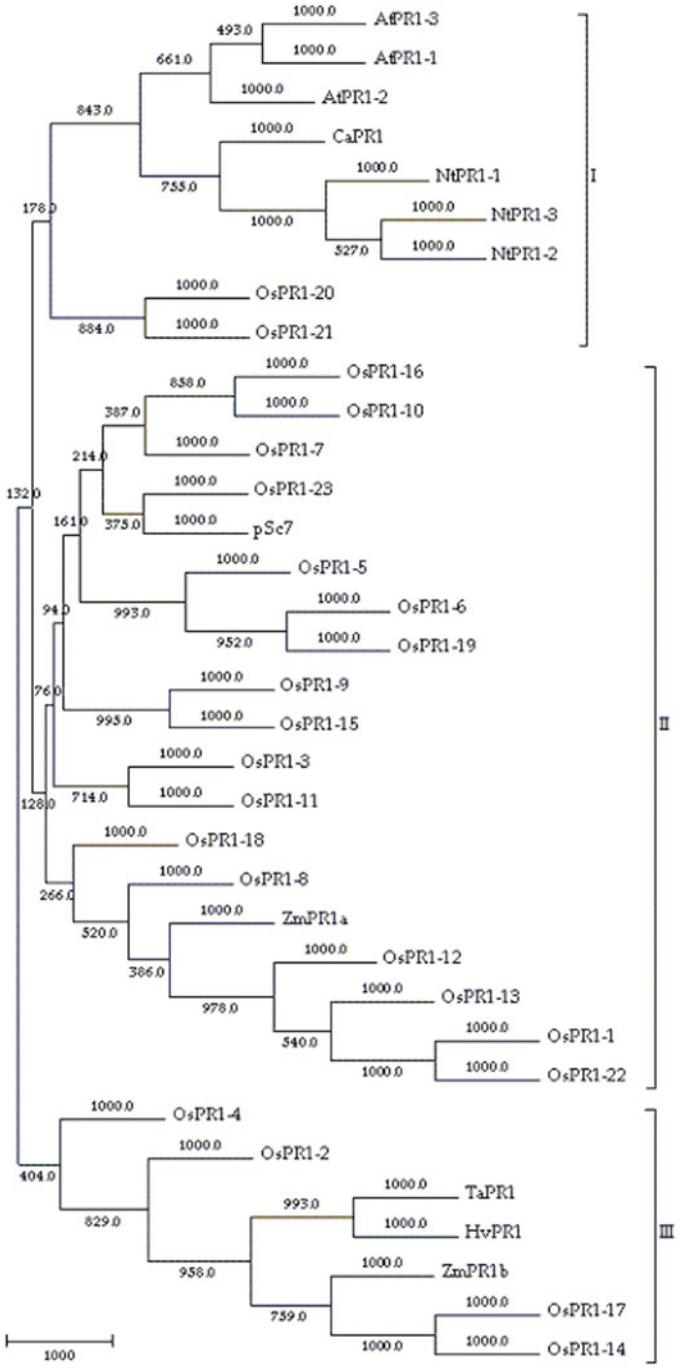

Figure 3. Phylogenetic tree of rice PR-1 amino acid sequences. Bootstrap values are shown at each node. AtPR1-1, AtPR1-2 and AtPR1-3, *Arabidopsis thaliana* PR-1 (accession number M90508, X96600); CaPR1, *Capsicum annuum* PR-1 (AF053343); NtPR1-1, NtPR1-2 and NtPR1-3, *Nicotiana tabacum* PR-1 (X06930, D90197 and X05454) ; pSc7, *Schizophyllum commune* fruiting body protein (M81722); ZmPR1a, *Zea mays* PR-1a (U82200); ZmPR1b, *Z. mays* PR-1b (Q00008); TaPR1, *Triticum aestivum* PR-1 (AF384143); HvPR1, *Hordeum vulgare* PR-1 (Z26320).

The PR-1 protein family belongs to a distinct and highly conserved group of proteins (Sarowar *et al*. 2005). Numerous studies have revealed that plant PR-1 proteins exhibit homologies and structural motifs in common with proteins from fungi, insects and vertebrates, demonstrating that *PR-1*-type genes have a common evolutionary origin (Van Loon and Van Strien 1999). The results of phylogenetic analysis strongly support this deduction.

After pathogen infection, PR-1 proteins are strongly induced (Manandhar *et al*. 1999) and seem to affect membranes (Liu *et al*. 2005), but their precise function still remains unclear. Our phylogenetic tree (figure 3) divided rice PR-1 proteins into three groups on the basis of sequence homology. Therefore, studies of homologous proteins identified in other organisms may provide information for clarifying the actual mechanism of the antifungal and unknown functions of rice PR-1. On the other hand, the expression pattern of known *PR-1* genes in rice would also greatly facilitate functional annotation of unknown homologues. Clearly, *in vivo* experiments and *in vitro* studies are necessary to establish the distinctive activities and biological roles of *PR-1* genes.

Acknowledgements

We are thankful for the valuable suggestions and comments of Dr Amitabh Joshi and the anonymous reviewer. This project was supported by China Postdoctoral Science Foundation (No. 20060390348).

References

Agrawal G. K., Jwa N. S. and Rakwal R. 2000a A novel rice (*Oryza sativa* L.) acidic PR1 gene highly responsive to cut, phytohormones, and protein phosphatase inhibitors. *Biochem. Biophys. Res. Commun.* **274**, 157–165.

Agrawal G. K., Rakwal R. and Jwa N. S. 2000b Rice (*Oryza sativa* L.) OsPR1b gene is phytohormonally regulated in close interaction with light signals. *Biochem. Biophys. Res. Commun.* **278**, 290–298.

Fernández C., Szyperski T., Bruyère T., Ramage P., Mosinger E. and Wutrich K. 1997 NMR solution structure of the pathogenesis-related protein p14a. *J. Mol. Biol.* **266**, 576–593.

International Rice Genome Sequencing Project 2005 The map-based sequence of the rice genome. *Nature* **436**, 793–800.

Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H. and Ooka H. 2003 Collection, mapping, and annotation of over 28000 cDNA clones from *japonica* rice. *Science* **301**, 376–379.

Kim S., Ahn H. P., Park C. H., Park S. G., Park S. Y., Jwa N. S. and Lee Y. H. 2001 Molecular characterization of the cDNA encoding an acidic isoform of PR-1 protein in rice. *Mol. Cells* **11**, 115–121.

Kim Y. J. and Hwang B. K. 2000 Pepper gene encoding a basic pathogenesis-related 1 protein is pathogen and ethylene inducible. *Physiol. Plant.* **108**, 51–60.

Kitajima S. and Sato F. 1999 Plant pathogenesis-related proteins: molecular mechanisms of gene expression and protein function. *J. Biochem.* **125**, 1–8.

Kumar S., Tamura K. and Nei M. 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**, 150–163.

Lee M. G. S., Lewis S. A., Wilde C. D. and Cowan N. J. 1983 Evolutionary history of a multigene family: an expressed human β-tubulin gene and three processed pseudogenes. *Cell* **33**, 477–487.

Liu Z., Du L. and Wan B. 2005 Pathogenesis-related proteins in higher plants. *Natur. Prod. Res. Dev.* **17**, 229–234.

Manandhar H. K., Mathur S. B., Smedegaard-Petersen V. and Thordal-Christensen H. 1999 Accumulation of transcripts for pathogenesis-related proteins and peroxidase in rice plants triggered by *Pyricularia oryzae, Bipolaris sorokiniana* and UV light. *Physiol. Mol. Plant Pathol.* **55**, 289–295.

Molina A., Gorlach J., Volrath S. and Ryals J. 1999 Wheat genes encoding two types of PR-1 proteins are pathogen inducible, but do not respond to activators of systemic acquired resistance. *Mol. Plant-Microbe Interact.* **12**, 53–58.

Morris S. W., Vernooij B., Titatarn S., Starrett M., Thomas S., Wiltse C. C., Frederiksen R. A., Bhandhufalck A., Hulbert S. and Uknes S. 1998 Induced resistance responses in maize. *Mol. Plant-Microbe Interact.* **11**, 643–658.

Muradov A., Petrasovits L., Davidson A. and Scott K. J. 1993 A cDNA clone for a pathogenesis-related protein 1 from barley. *Plant Mol. Biol.* **23**, 439–442.

Pfitzner U. M. and Goodman H. M. 1987 Isolation and characterization of cDNA clones encoding pathogenesis-related proteins from tobacco mosaic virus infected tobacco plants. *Nucl. Acids Res.* **15**, 4449–4465.

Sarowar S., Kim Y. J., Kim E. N., Kim K. D., Hwang B. K., Islam R. and Shin J. S. 2005 Overexpression of a pepper basic pathogenesis-related protein 1 gene in tobacco plants enhances resistance to heavy metal and pathogen stresses. *Plant Cell Rep.* **24**, 216–224.

Tornero P., Conejero V. and Vera P. 1994 A gene encoding a novel isoform of the PR-1 protein family from tomato is induced upon viroid infection. *Mol. Gen. Genet.* **243**, 47–53.

Uknes S., Mauch-Mani B., Moyer M., Williams S., Dincher S., Chandler D. *et al*. 1992 Acquired resistance in *Arabidopsis. Plant Cell* **4**, 645–656.

Van Loon L. C. 1997 Induced resistance in plants and the role of pathogenesis-related proteins. *Eur. J. Plant Pathol.* **103**, 753–765.

Van Loon L. C. and Van Strien E. A. 1999 The families of pathogenesis-related proteins, their activities, and comparative analysis of PR1 type proteins. *Physiol. Mol. Plant Pathol.* **55**, 85–97.

Yu L., Niu J. S., Ma Z. Q., Li X. Y., Chen P. D. and Liu D. J. 2001 Cloning and characterization of pathogenesis-related protein 1 from wheat - *H. villosa* translation lines. *Acta Genet. Sin.* **28**, 760–768.

Received 12 November 2005