ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2011) **9**(2); 342–346
doi:10.1017/S1479262111000232

Protein disulphide isomerase family in bread wheat (*Triticum aestivum* L.): genomic structure, synteny conservation and phylogenetic analysis

E. d'Aloisio[1], A. R. Paolacci[2], A. P. Dhanapal[1], O. A. Tanzarella[2], E. Porceddu[2] and M. Ciaffi[2]*

[1]*Scuola Superiore Sant'Anna, Pisa, Italy and* [2]*Dipartimento di Agrobiologia e Agrochimica, Università della Tuscia, Viterbo, Italy*

Abstract

Eight genes encoding protein disulphide isomerase (PDI)-like proteins in bread wheat were cloned and characterized and their genomic structure was compared with that of homoeologous genes isolated from other plant species. Fourteen wheat cDNA sequences of PDI-like genes were amplified and cloned; eight of them were relative to distinct PDI-like genes, whereas six corresponded to homoeologous sequences. Also, the genomic sequences of the eight non-homoeologous genes were amplified and cloned. Phylogenetic analysis, which included eight genes encoding PDI-like proteins and the gene encoding the typical PDI, assigned at least one of them to each of the eight major clades identified in the phylogenetic tree of the PDI gene family of plants. The close chromosome synteny between wheat and rice was confirmed by the location of the homoeologous genes of the PDI family in syntenic regions of the two species. Within the same phylogenetic group, a high level of conservation, in terms of sequence homology, genomic structure and domain organization, was detected between wheat and the other plant species. The high level of conservation of sequence and genomic organization within the PDI gene family, even between distant plant species, might be ascribed to the key metabolic roles of their protein products.

Keywords: chromosome synteny; genomic structure; phylogenetic analysis; protein disulphide isomerase; *Triticum aestivum*

Introduction

The protein disulphide isomerase (PDI) gene family includes the 'typical PDI', which catalyzes formation, reduction and isomerization of disulphide bonds in secretory proteins within the lumen of the endoplasmic reticulum (ER). The proteins encoded by the genes of the plant PDI gene family cluster into eight phylogenetic classes (Houston *et al*., 2005) and differ for number and position of the active thioredoxin-like site (CGHC: Cysteine, Glycine, Histidine, Cysteine) and for presence/absence of specific domains and of the Lysine, Aspartic acid, Glutamic acid, Leucine (KDEL) signal of retention in the ER. They are involved in folding and deposition of seed storage proteins in several species (Li and Larkins, 1996; Takemoto *et al*., 2002). Since wheat flour quality is strongly affected by composition and structure of seed storage proteins, the potential involvement of proteins of the wheat PDI family in their folding and in the formation of intra- and inter-molecular disulphide bonds makes their study particularly interesting. Genomic, cDNA and promoter sequences of the three homoeologous gene encoding the 'typical' PDI have been cloned and characterized (Ciaffi *et al*., 2006). Here, we report the isolation and

* Corresponding author. E-mail: ciaffi@unitus.it

characterization of eight new non-homoeologous genes coding for PDI-like proteins, their assignment to the eight phylogenetic groups of the plant PDI family, their chromosome location and the organization of their genomic sequences.

Material and methods

Single plants of bread wheat cv. Chinese Spring (CS) and its nulli-tetrasomic (NT) lines were used for DNA and RNA extractions. The Dana-Faber Cancer Institute (DFCI) wheat gene index database (TaGI, version 11) was BLAST-searched with the available sequences of PDI-like genes of rice (Houston *et al.*, 2005) and the non-redundant wheat PDI-like tentative consensus (TC) sequences identified were used as template for 5′ and 3′ Rapid Amplification of cDNA Ends (RACE) extension. Gene-specific primer pairs, designed on the 5′ and 3′ Un Translated Region (UTR) sequences, were used for cDNA and genomic amplifications, followed by cloning and sequencing of the resulting amplicons. The evolutionary relationships between PDI and PDI-like genes of wheat and other plants were studied by phylogeny reconstruction based on the alignment of 92 deduced amino-acid sequences of the following genes: 9 of wheat, 13 of *Arabidopsis*, 12 of poplar, 10 of grapevine, 5 of soybean, 12 each of maize and rice, 14 of *Physcomitrella patens* and 5 of *Chlamydomonas reinhardtii*. The chromosome location of the newly identified PDI-like gene sequences was determined through Southern analyses of CS and its NT lines using digoxigenin-labelled probes.

Results and discussion

The BLAST search using 12 PDI-like gene sequences of rice in the DFCI wheat gene index database fetched nine TC sequences, one of them encoding the typical PDI, whose three homoeologous genes had previously been cloned and characterized (Ciaffi *et al.*, 2006). The eight TC sequences encoding PDI-like genes (Table 1) were used as RACE template to isolate their 5′ and 3′ extensions, subsequently validated by sequence analysis and the corresponding full-length cDNAs were cloned by RT-PCR of RNA from various wheat tissues using specific primer pairs designed in the 5′ and 3′ UTRs. Fourteen wheat cDNA sequences encoding PDI-like proteins were identified by sequence analysis; eight of them derived from distinct genes, whereas six corresponded to homoeologous sequences (Table 1).

Phylogenetic analysis included the nine non-homoeologous sequences of the wheat PDI family into the eight phylogenetic groups identified in plants, then at least one wheat gene had been cloned for each group (Fig. 1). On the basis of the modular structure of the proteins, six of the eight subfamilies were clustered into two major clades, whereas the proteins of the subfamilies VI and VIII being highly diversified were considered as outgroups (Fig. 1). Since genes from *P. patens*, monocots and dicots formed three distinct sub-clusters within each of the eight PDI phylogenetic groups (Fig. 1), the eight subfamilies would have emerged before the divergence of bryophytes and angiosperms. Only three of the five PDI-like encoding genes from *C. reinhardtii* were included in plant phylogenetic groups (CrPDI-4 in group V, CrPDI-5 in group VIII and CrPDI-3 in group VI), indicating that only three PDI-like genes would be common to both chlorophytes and streptophytes, which diverged over 1 billion years ago. The presence of multiple genes of the same species within single phylogenetic groups can be explained by duplication events occurred either after the separation of the angiosperms from the briophytes or later, after the diversification of monocots and dicots. It is noteworthy that only group VII hosted two paralogous genes of wheat not related to the allopolyploid origin of its genome.

The eight non-homoeologous genomic sequences encoding the novel PDI-like proteins were amplified by the same primer combinations previously used to clone the cDNA sequences; their lengths were: (1) *TaPDIL2-1*: 5213 bp; (2) *TaPDIL3-1*: 4294 bp; (3) *TaPDIL4-1*: 3820 bp, (4) *TaPDIL5-1*: 5326 bp; (5) *TaPDIL6-1*: 2162 bp; (6) *TaPDIL7-1:* 2887 bp; (7) *TaPDIL7-2*: 2475 bp; (8) *TaPDIL8-1*: 7034 bp. The exon–intron structure was determined by their alignment with the corresponding cDNA sequences, which showed an almost perfect nucleotide match between cDNAs and exonic regions of the genomic sequences. The eight genes showed a complex genomic organization with following exon numbers: 12 in *TaPDIL2-1* and *TaPDIL3-1*; 11 in *TaPDIL4-1*; 9 in *TaPDIL5-1*; 4 in *TaPDIL6-1*; 5 in *TaPDIL7-1* and *TaPDIL7-2*; 15 in *TaPDIL8-1* (Supplementary Fig. S1, available online only at http://journals.cambridge.org). Consistent with previous studies (Kersanach *et al.*, 1994; Petersen *et al.*, 2006), the genes of wheat, *Arabidopsis*, rice and *P. patens* clustering into the same phylogenetic group revealed a high level of conservation of their structural features (exon–intron pattern and number, size and position of the protein active sites; Supplementary Fig. S1, available online only at http://journals.cambridge.org), whereas the intron–exon structure of the genes of the alga *C. reinhardtii* was very different (data not shown).

The genes encoding the typical PDI had been located in chromosome arms 4AL, 4BS and 4DS of bread wheat (Ciaffi *et al.*, 2006). The chromosome locations of the eight wheat genes encoding PDI-like proteins, which were determined by Southern analysis

https://doi.org/10.1017/S1479262111000232 Published online by Cambridge University Press

Table 1. Characteristics of the full-length cDNA sequences coding for wheat PDI-like proteins cloned in this study and their chromosome location

	T. aestivum cv. CS					*O. sativa*				
	Full-length cDNA					Orthologous rice gene				
Clone[a]	UTR5′ (nt)	UTR3′ (nt)	Open Reading Frame (ORF) (nt)	TC sequence (DFCI wheat gene index)	Chromosome location[b]	Previous name[c]	This study	Acc. number of cDNA	Protein identity	Chromosome location[b]
TaPDIL2-1a	63	109	1767	TC301880	W6	OsPDIL1-4	OsPDIL2-1	AK071514	408/561 (72.73%)	R2
TaPDIL3-1a	141	110	1626	TC353685	W7	OsPDIL1-5	OsPDIL3-1	AK073970	437/529 (82.61%)	R6
TaPDIL4-1a	74	139	1104	TC300461	W1	OsPDIL2-1	OsPDIL4-1	AK103944	316/366 (86.34%)	R5
TaPDIL4-1b	74	138	1104	TC300461	W1	OsPDIL2-1	OsPDIL4-1	AK103944	317/366 (86.61%)	R5
TaPDIL5-1a	55	150	1323	TC317379	W5	OsPDIL2-3	OsPDIL5-1	AK062254	391/439 (89.07%)	R9
TaPDIL5-1b	55	151	1323	TC317379	W5	OsPDIL2-3	OsPDIL5-1	AK062254	397/439 (90.43%)	R9
TaPDIL6-1a	13	240	456	TC294820	W4	OsPDIL5-1	OsPDIL6-1	AK063663	121/146 (82.88%)	R3
TaPDIL6-1b	13	255	450	TC294820	W4	OsPDIL5-1	OsPDIL6-1	AK063663	119/146 (81.51%)	R3
TaPDIL7-1a	37	289	1242	TC287269	W2	OsPDIL5-2	OsPDIL7-1	AK069367	355/411 (86.37%)	R4
TaPDIL7-1b	37	289	1254	TC287269	W2	OsPDIL5-2	OsPDIL7-1	AK069367	359/416 (86.30%)	R4
TaPDIL7-1c	37	289	1242	TC287269	W2	OsPDIL5-2	OsPDIL7-1	AK069367	356/411 (86.62%)	R4
TaPDIL7-2a	53	186	1257	TC287749	W6	OsPDIL5-3	OsPDIL7-2	ND	311/416 (74.76%)	R2
TaPDIL7-2b	53	186	1257	TC287749	W6	OsPDIL5-3	OsPDIL7-2	ND	311/416 (74.76%)	R2
TaPDIL8-1a	122	226	1458	TC301351	W2	OsPDIL5-4	OsPDIL8-1	AK099660	450/485 (92.78%)	R7

ND, no cDNA sequence available, tBLASTn searches of Gramene identified predicted transcript GRMT00000163510 (OsPDIL5-3/OsPDIL7-2) and the amino acid sequence was predicted from the available genomic sequence Os02g34530 (Houston *et al.*, 2005).

[a] A code of two letters (Ta = *T. aestivum*) followed by the suffix PDIL and by an Arabic number indicating the corresponding phylogenetic group was assigned to each sequence. Multiple sequences clustering into the same subfamily were designed by an additional number (1 and 2) and putative homoeologous gene sequences were distinguished with an additional letter (a–c). Corresponding TC sequences, identified in DFCI wheat gene index database, orthologous rice genes and their chromosomal location have also been reported. [b] The seven wheat homoeologous groups (w1–w7) are syntenic to the 12 rice chromosomes (r1–r12), in particular w1 = r5 + r10, w2 = r7 + r4, w3 = r1, w4 = r3 + r11, w5 = r12 + r9 + r3, w6 = r2 and w7 = r6 + r8 (La Rota and Sorrells, 2004). [c] Nomenclature used by Houston *et al.* (2005).

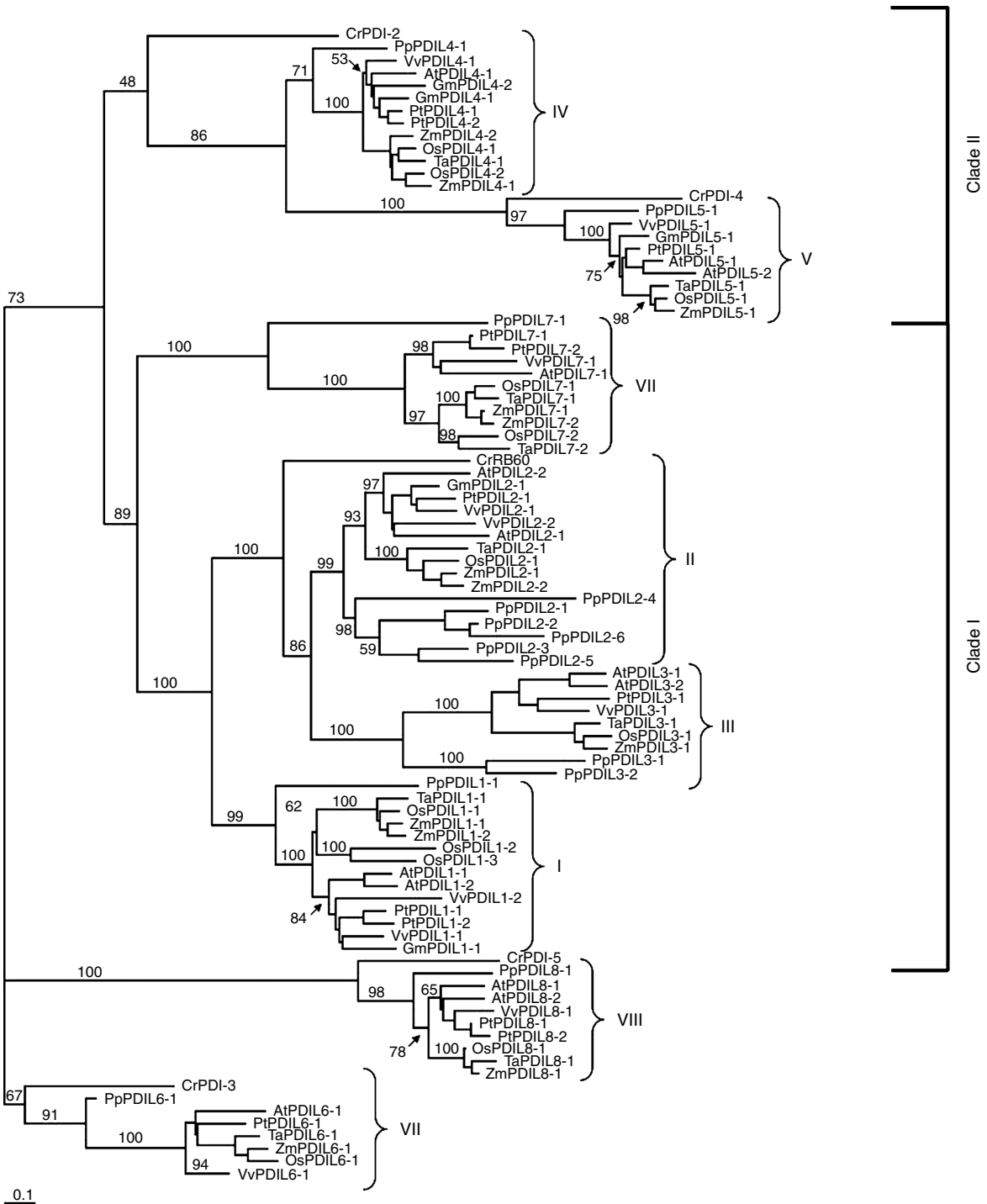

Fig. 1. Phylogenetic tree based on deduced amino-acid sequences of 92 PDI and PDI-like genes: nine of wheat (Ta), 13 of *Arabidopsis* (At), 12 of poplar (Pt), 10 of grapevine (Vv), 5 of soybean (Gm), 12 each of maize (Zm) and rice (Os), 14 of *P. patens* (Pp) and 5 of *C. reinhardtii* (Cr). Multiple alignment was performed by ClustalX 1.83 software using the Gonnet series as protein weight matrix and parameters set to three gap open penalty, 1.6 gap extension penalty, negative matrix on and divergent sequence delay set at 36% followed by manual adjustment. The phylogenetic tree was constructed by NEIGHBOR (PHYLIP 3.6). Distance matrices were estimated by PRODIST using the Percent Accepted Mutation (PAM) model of amino acid transition. To evaluate statistical significance, 1000 bootstrap replicates were generated by SEQBOOT. Numbers on main branches indicate bootstrap percentages for 1000 replicates. The most divergent PDI-like sequences of groups VI and VIII were set as outgroups. The major clades (I and II) and the eight phylogenetic groups (I–VIII) are enclosed by curly and square brackets, respectively.

https://doi.org/10.1017/S1479262111000232 Published online by Cambridge University Press

of CS and its aneuploid NT lines, were compatible with those of their rice orthologs, as expected on the basis of the main chromosome synteny between rice and wheat (Table 1). In fact, the seven wheat homoeologous groups (w1–w7) are syntenic to the 12 rice chromosomes (r1–r12): w1 = r5 + r10, w2 = r7 + r4, w3 = r1, w4 = r3 + r11, w5 = r12 + r9 + r3, w6 = r2 and w7 = r6 + r8 (La Rota and Sorrells, 2004), even though their collinearity has often been disrupted by extensive duplications and rearrangements of chromosome segments. The outputs of the sequence-based macro-collinearity between the Michigan State University (MSU) rice pseudomolecules (Release 6.1) and the wheat Expressed Aequence Tags (ESTs), whose locations on the wheat Bin map is reported in GrainGenes, were exploited to assess the syntenic relationships between the regions of the wheat and rice chromosomes flanking the PDI-like encoding genes (Supplementary Table S1, available online only at http://journals.cambridge.org). Besides confirming the chromosome locations, the analysis based on Wheat Bin Mapped Markers assigned the PDI-like genes to specific wheat chromosome arms. The close syntenic relationships between wheat and rice confirmed by this study might be helpful for high-density mapping of specific wheat chromosome regions, which would be useful for chromosome walking and gene positional cloning.

References

Ciaffi M, Paolacci AR, D'Aloisio E, Tanzarella OA and Porceddu E (2006) Cloning and characterization of wheat PDI (protein disulfide isomerase) homoeologous genes and promoter sequences. *Gene* 366: 209–218.

Houston NL, Fan C, Xiang QY, Schulze JM, Jung R and Boston RS (2005) Phylogenetic analyses identify 10 classes of the protein disulfide isomerase family in plants, including single-domain protein disulfide isomerase-related proteins. *Plant Physiology* 137: 762–778.

Kersanach R, Brinkmann H, Liaud MF, Zhang DX, Martin W and Cerff R (1994) Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature* 367: 387–389.

La Rota M and Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. *Functional and Integrative Genomics* 4: 34–46.

Li CP and Larkins BA (1996) Expression of protein disulfide isomerase is elevated in the endosperm of the maize floury-2 mutant. *Plant and Molecular Biology* 30: 873–882.

Petersen J, Teich R, Brinkmann H and Cerff R (2006) A "green" phosphoribulokinase in complex algae with red plastids: evidence for a single secondary endosymbiosis leading to haptophytes, cryptophytes, heterokonts and dinoflagellates. *Journal of Molecular Evolution* 62: 143–157.

Takemoto Y, Coughlan SJ, Okita TW, Satoh H, Ogawa M and Kumamaru T (2002) The rice mutant esp2 greatly accumulates the glutenin precursor and deletes the protein disulfide isomerase. *Plant Physiology* 128: 1212–1222.