

Phylogenetic Analyses Identify 10 Classes of the Protein Disulfide Isomerase Family in Plants, Including Single-Domain Protein Disulfide Isomerase-Related Proteins^{1[w]}

Norma L. Houston, Chuanzhu Fan², (Jenny) Qiu-Yun Xiang, Jan-Michael Schulze, Rudolf Jung, and Rebecca S. Boston*

Department of Botany, North Carolina State University, Raleigh, North Carolina 27695–7612 (N.L.H., C.F., Q.-Y.X., R.S.B.); and Pioneer Hi-Bred International, Inc., a DuPont Company, Johnston, Iowa 50131 (J.-M.S., R.J.)

Protein disulfide isomerases (PDIs) are molecular chaperones that contain thioredoxin (TRX) domains and aid in the formation of proper disulfide bonds during protein folding. To identify plant PDI-like (PDIL) proteins, a genome-wide search of *Arabidopsis* (*Arabidopsis thaliana*) was carried out to produce a comprehensive list of 104 genes encoding proteins with TRX domains. Phylogenetic analysis was conducted for these sequences using Bayesian and maximum-likelihood methods. The resulting phylogenetic tree showed that evolutionary relationships of TRX domains alone were correlated with conserved enzymatic activities. From this tree, we identified a set of 22 PDIL proteins that constitute a well-supported clade containing orthologs of known PDIs. Using the *Arabidopsis* PDIL sequences in iterative BLAST searches of public and proprietary sequence databases, we further identified orthologous sets of 19 PDIL sequences in rice (*Oryza sativa*) and 22 PDIL sequences in maize (*Zea mays*), and resolved the PDIL phylogeny into 10 groups. Five groups (I–V) had two TRX domains and showed structural similarities to the PDIL proteins in other higher eukaryotes. The remaining five groups had a single TRX domain. Two of these (quiescin-sulphydryl oxidase-like and adenosine 5'-phosphosulfate reductase-like) had putative nonisomerase enzymatic activities encoded by an additional domain. Two others (VI and VIII) resembled small single-domain PDIs from *Giardia lamblia*, a basal eukaryote, and from yeast. Mining of maize expressed sequence tag and RNA-profiling databases indicated that members of all of the single-domain PDIL groups were expressed throughout the plant. The group VI maize PDIL ZmPDIL5-1 accumulated during endoplasmic reticulum stress but was not found within the intracellular membrane fractions and may represent a new member of the molecular chaperone complement in the cell.

Proper folding of nascent polypeptides into functional proteins relies on a number of molecular chaperones and protein-folding catalysts that act to shield nonnative structures from aggregation until they fold into a native, stable state. One group of these folding catalysts, the protein disulfide isomerases (PDIs), interacts with nascent polypeptides to catalyze the formation, isomerization, and reduction/oxidation of disulfide bonds (for review, see Freedman et al., 1994; Wilkinson and Gilbert, 2004). Multiple PDI-related

genes have been identified in each eukaryotic genome surveyed by whole-genome sequencing. For most of the corresponding proteins, a demonstrated biochemical function is lacking; therefore, we refer to them as PDI-like (PDIL) proteins. PDIL proteins are members of a multigene family within the thioredoxin (TRX) superfamily, which includes, in addition, glutaredoxins, TRXs, ferredoxins, and peroxidoxins (Jacquot et al., 2002). All proteins in this ancient superfamily have at least one structural domain that functions through Cys residues in a CXXC tetrapeptide sequence (for review, see Ellgaard, 2004; Wilkinson and Gilbert, 2004). These domains, called TRX structural folds, have amino acids arranged in a conserved three-dimensional conformation (Kemmerling et al., 1997). The most prevalent and best-studied PDIL proteins are orthologs of an approximately 57-kD PDI that has been found in most eukaryotic organisms investigated, including plants. (In this report, we will refer to the approximately 57-kD protein as the major PDI whose counterparts in *Arabidopsis* [*Arabidopsis thaliana*], rice [*Oryza sativa*], and maize [*Zea mays*] have the GenBank accession nos. of NM_102024, AK068268, and L39014, respectively.) The disulfide isomerase function of PDI proteins is important in all cell types, especially those cells devoted to protein secretion and/or storage.

¹ This work was supported by the U.S. Department of Energy (grant no. DE-FG02-00ER150065 to R.S.B.), the National Science Foundation (grant no. DEB-0129069 to Q.-Y.X.), the North Carolina Agricultural Research Service (R.S.B. and Q.-Y.X.), and a fellowship (to N.L.H.) in the North Carolina State University Functional Genomics graduate program from National Science Foundation Integrative Graduate Education and Research Traineeship (grant no. 9987555).

² Present address: Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637.

* Corresponding author; e-mail boston@unity.ncsu.edu; fax 919-515-3436.

[w] The online version of this article contains Web-only data.

Article, publication date, and citation information can be found at www.plantphysiol.org/cgi/doi/10.1104/pp.104.056507.

Although there has been extensive biochemical characterization of a few PDIL proteins, most analysis has been based on similarities of amino acid sequences from organisms outside the plant kingdom. Adding further to the complexity of this gene family is the large size of the TRX superfamily. As part of an open-ended RNA-profiling study, we identified several sequences predicted to encode proteins that had functions related to the redox state of protein disulfide bonds and were up-regulated in endosperm of maize mutants associated with endoplasmic reticulum (ER) stress. The set of induced genes included the major PDI and also other sequences that contained TRX domains. We attempted to predict the function of these genes in silico with bioinformatics tools but found that misannotations in whole-genome databases and the absence of full-length sequences for PDIL proteins from other plants hindered a reliable analysis. Phylogenetic analyses available from previous studies of PDIL proteins included very few sequences from plants and considered only a narrow selection of likely PDIL proteins (Sahrawy et al., 1996; Kanai et al., 1998; McArthur et al., 2001). Furthermore, previous analyses that did include plant sequences focused primarily on subsets rather than on the global organization of the TRX superfamily, leaving an absence of reliable comparisons to define the PDIL gene space within it (Meyer et al., 1999; Balmer and Buchanan, 2002; Meiri et al., 2002; Brehelin et al., 2003; Takubo et al., 2003; Hanke et al., 2004; Rouhier et al., 2004).

Biochemical and cell fractionation studies have shown that PDI activity is generally associated with the ER, the entry point into the secretory pathway. Many PDIL proteins have NH₂-terminal sequences predicted to be signal peptides for ER targeting, and several have COOH-terminal KDEL motifs that serve as retrieval tags for ER resident proteins (Pelham, 1990). The ER is an important site of PDI function, in part due to the oxidizing environment that favors disulfide-bond formation. Mutations or pharmacological agents that perturb ER homeostasis lead to an ER stress response that includes induction of a number of PDIL genes. For example, in mammals the genes Erp57 and Erp72 and in yeast the PDIL genes EUG1, MPD1, MPD2, and EPS1 have been found to be induced (Wang and Chang, 1999; Travers et al., 2000; for review, see Wilkinson and Gilbert, 2004).

In maize, the major PDI accumulates to high levels in seeds producing mutant storage proteins that trigger induction of the ER stress response (Li and Larkins, 1996). This induction has been characterized in the maize mutants *floury-2* (*fl2*), *Mucronate* (*Mc*), and *Defective endosperm B30* (*De^{*}-B30*) that exhibit tissue-specific increases in molecular chaperones (including PDI) as a result of structural changes in storage proteins (Coleman et al., 1995; Li and Larkins, 1996; Wrobel, 1996; Kim et al., 2004). Because the ER stress response in these mutants is spatially and developmentally specific for endosperm during accumulation

of storage protein reserves, these mutants are excellent model systems for studying responses associated with ER stress. Even in the absence of ER stress, seeds provide a rich source of the major PDI, which is needed not only in response to ER stress but also for production of highly disulfide-bonded storage proteins (for review, see Shewry et al., 1995). The major PDI has been isolated and characterized from alfalfa (*Medicago sativa*) and seeds of castor (*Ricinus communis*), wheat (*Triticum aestivum*), and maize (for review, see Boston et al., 1996).

PDIL proteins from widely divergent organisms share similar functional building blocks and a common TRX domain organization (McArthur et al., 2001; Norgaard et al., 2001; Alanen et al., 2003; Wilkinson and Gilbert, 2004). Perhaps the best-studied PDIL multigene family is that of mammals with at least eight members that can be grouped into four classes based on position and number of TRX domains (Kanai et al., 1998). Class 1, the largest structural group, contains sequences with two active TRX domains, one near the NH₂ terminus and one in the COOH-terminal region of the protein. Class 3 proteins have these domains plus an additional TRX domain at the NH₂ terminus. Class 2 proteins (two domains) and class 4 proteins (three domains) have TRX domains that occur in tandem at the NH₂ or COOH terminus, respectively.

Single-celled organisms appear to have a less complex PDIL family than mammals. The PDIL family in *Saccharomyces cerevisiae* has five members, including a homolog of the major PDI that is necessary for viability (Norgaard et al., 2001). Three of the *S. cerevisiae* PDIL proteins have single TRX domains and thus fall into a fifth structural class (class 5), in which we will group all single-domain proteins (Norgaard et al., 2001). Like *S. cerevisiae*, *Giardia lamblia* contains five PDIL proteins, but all five have single TRX domains (McArthur et al., 2001).

In an effort to bring together a current data set of proteins encoding TRX domains from which we could identify PDIL members of the TRX superfamily in plants, we initiated a search to extract sequences encoding TRX domains from Arabidopsis genomic databases. From these data, we performed analyses to incorporate plant PDIL proteins into the existing PDI phylogeny and, after extensive sequencing of full-length PDIL cDNAs, compiled the comprehensive lists of PDIL gene sets presented here. Through this analysis, we introduced 49 additional sequences into the PDIL families from Arabidopsis, rice, and maize and identified five single TRX domain PDIL phylogenetic groups that arose prior to the split between monocots and eudicots, are evolutionarily distinct from each other, and are structurally distinct from the major PDI. The smallest member of the PDIL proteins (approximately 150 amino acids) had an NH₂-terminal signal sequence and showed a strong induction during ER stress but did not fractionate with organelles of the secretory pathway.

RESULTS

Phylogenetic Analysis of the Arabidopsis TRX Superfamily and Identification of the PDIL Clade

Altogether 117 TRX domains from 104 Arabidopsis-predicted amino acid sequences were compiled into a data matrix and aligned using ClustalX software (Thompson et al., 1997). The phylogenetic tree derived from analysis of the matrix using the Bayesian method is shown in Figure 1, and the matrix itself is available in Supplemental Table I. In general, subclades in the tree corresponded to members of putative functionally distinct subfamilies of the TRX superfamily (i.e. glutaredoxin, TRX, PDIL, ferredoxin, peroxidoxin) and provided the basis for selecting and further analyzing a likely complete list of Arabidopsis PDIL genes.

The phylogenetic analysis of the TRX domains identified a well-supported clade containing putative disulfide isomerases and oxidoreductases that act in the protein secretory pathway of plants (Fig. 1). This PDIL clade included two close homologs (At1g21750 and At1g77510; Fig. 1, arrowheads) of the functionally characterized castor PDI (accession no. AAB05641; Coughlan et al., 1996). Unexpectedly, this clade, henceforth designated PDIL, also included two protein groups for which enzymatic activities other than PDI had been shown. One of these, the adenosine 5'-phosphosulfate reductase-like (APRL) group, contained three sequences that had been shown to have reductase activity typical of TRXs as well as an adjacent domain responsible for adenylyl sulfate reductase (APR) activity (Gutierrez-Marcos et al., 1996; Setya et al., 1996; Wray et al., 1998; Prior et al., 1999). This APRL group formed a separate subclade (87% posterior probability; Fig. 1) and was strongly supported (96% posterior probability) to be a member of the PDIL clade. The other group consisted of four closely related sequences, two of which, At1g15020 and At2g01270, belong to the quiescin-sulfhydryl oxidase (QSOX) family. Members of this family, in addition to a TRX domain, possess an Erv1-like domain at the COOH terminus (Fig. 1). Interestingly, Erv1 domains have been independently implicated in cellular redox processes (Lange et al., 2001) and thus may function interdependently when fused with TRX domains. The two QSOX proteins are nested within the PDIL proteins and, together with the remaining set of 20 protein sequences in the Arabidopsis PDI-related clade (PDIL), form four well-supported groups on the tree (Fig. 1). Accession numbers of this nonredundant set of 22 Arabidopsis PDIL sequences are shown in Table I.

With the exception of two previously named groups (QSOX and APRL, see above), we have adopted a consolidating nomenclature for designating the individual plant PDIL proteins based on species and the five structural PDIL classes as defined by Kanai et al. (1998). All of the plant PDIL sequences had one or two active TRX domains and therefore fell into structural

classes 1, 2, or 5. The full nomenclature includes two lowercase letters for genus and species, a capital PDIL, and the structural class designation followed by an Arabic number initially based on prevalence of expression with subsequent numbers denoting precedence. (For example, the major PDI of Arabidopsis would be AtPDIL1-1.) For the proteins within this clade that have previously characterized nonisomerase activities, we adopted the published precedents of QSOX-like (QSOXL; Thorpe et al., 2002) and APRL (Gutierrez-Marcos et al., 1996; Setya et al., 1996; Wray et al., 1998).

Phylogenetic Analysis of Eukaryotic PDIL Domains

To investigate the evolutionary relationships among the TRX domains of PDIL proteins in plants and other eukaryotes, we produced a reliable dataset for resolving evolutionary relationships within the PDIL family by confirming (and correcting when necessary) each of the protein coding regions from the Arabidopsis genomic sequences with comparisons to cDNA sequences. We further added to the Arabidopsis amino acid data matrix TRX domains of PDIL sequences for the moss *Physcomitrella patens* and the green alga *Chlamydomonas reinhardtii*. These sequences were combined with a data matrix previously constructed by McArthur et al. (2001; Supplemental Table II) for TRX domains of PDIL proteins from both higher and lower eukaryotes. Because of the availability of the Arabidopsis complete genome sequence, the new matrix provided the highest confidence of having a complete plant complement of PDIL sequences. The combined data matrix included sequences from single-celled eukaryotes, plants and animals with all five PDIL structural classes represented, but did not include the Arabidopsis QSOX or APRL members, as no counterparts of these sequences from other organisms were present in the existing data matrix. The data matrix was analyzed using both maximum-likelihood (Felsenstein, 2002) and Bayesian inference methods (Huelsenbeck and Ronquist, 2001).

Results from the Bayesian analysis as shown in Figure 2 indicated that, in general, the introduction of the additional plant sequences to the data matrix of McArthur et al. (2001) caused few disruptions in the relationships between evolutionary lineages and structural arrangement of the TRX domains, as inferred previously (McArthur et al., 2001). Relationships among the PDIL proteins revealed by our analyses were also similar to those determined by Kanai et al. (1998). Our results show that the Arabidopsis PDIL TRX domains belong to 13 evolutionary lineages (Fig. 2; lineages are marked by I–VIII with N and C used as in Fig. 1 to mark the position of the TRX domain within the protein sequence or S to indicate single TRX domain PDIL proteins). These lineages are dispersed throughout the tree with most of the sequences of the two-domain groups (I–V) associated with TRX domains of PDIL proteins from outside of

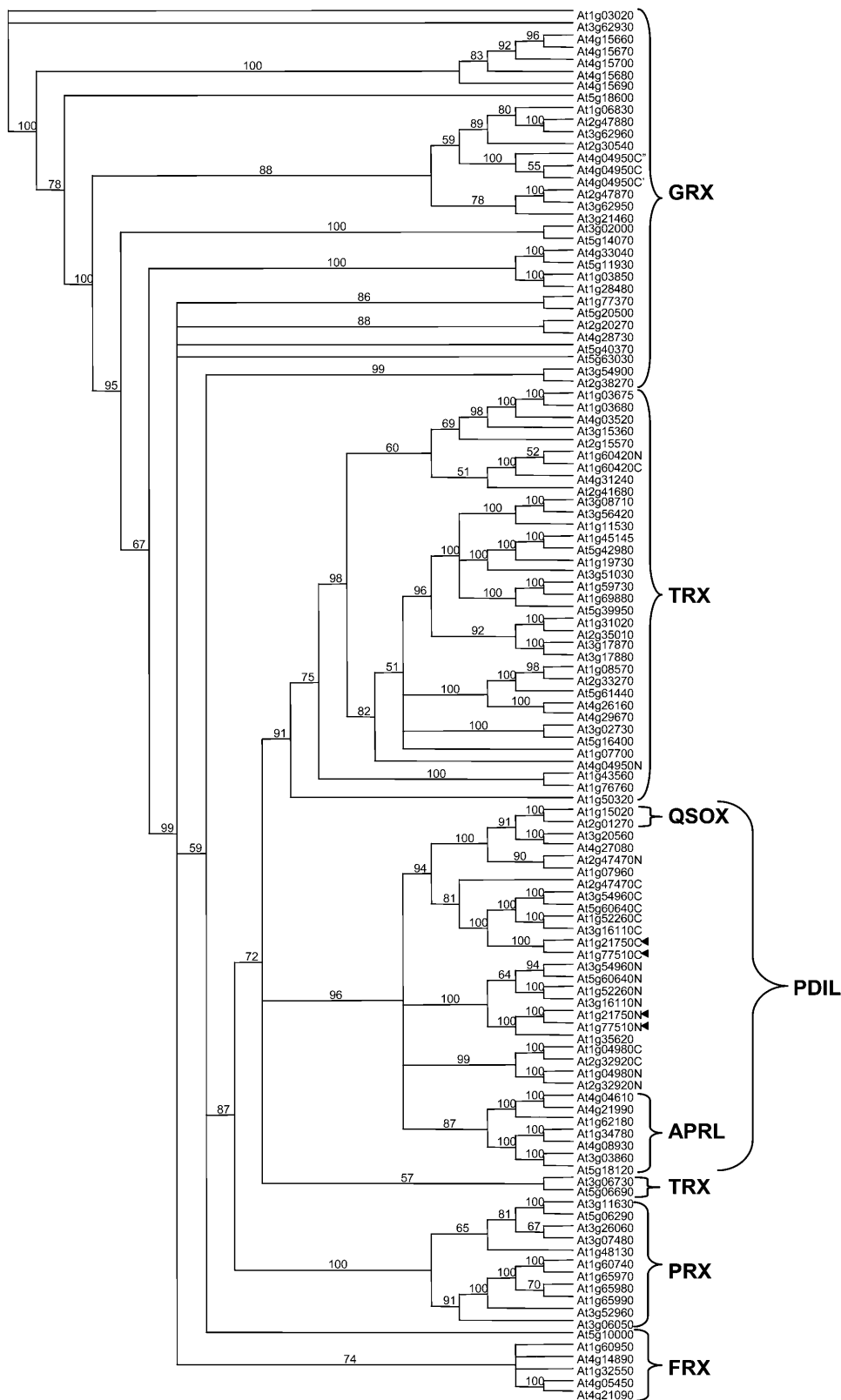


Figure 1. Phylogenetic tree of the Arabidopsis TRX domain superfamily resulting from Bayesian analysis of protein sequences using MrBayes 3.0 (Huelsenbeck and Ronquist, 2001). Posterior probability values (>50%) indicate nodal support and are shown above internodes. Brackets indicate members of the TRX superfamily based on annotation and functional data. GRX, Glutaredoxin; PRX, peroxidoxin; FRX, ferredoxin. Domain positions are represented as follows: C, COOH-terminal domain with C' and C'' denoting additional TRX domains; and N, NH₂-terminal domain. Arrowheads mark putative orthologs of castor PDI for which PDI enzymatic activity has been demonstrated.

Table 1. Properties of PDIL families from *Arabidopsis*, maize, and rice

Phylogenetic Group ^a	PDIL Designation ^b	No. of Exons ^c	Predicted Polypeptide Size ^d	Accession No. of Gene ^e	Map Location (Chromosome) ^f	Accession No. of cDNA ^{g,h}	Structural Class ⁱ	TRX Domains ^j	COOH-Terminal Tetrapeptide	Localization ^k
I	AtPDIL1-1	9	501	At1g21750	1	NM_102024	1	2	KDEL	S
I	AtPDIL1-2	10	508	At1g77510	1	NM_106400	1	2	KDEL	S
I	OsPDIL1-1	10	512		11	AK068268	1	2	KDEL	S
I	OsPDIL1-2	10	517		4	AY739308	1	2	KDEL	S
I	OsPDIL1-3	9	492		2	AK073161	1	2	KDEL	S
I	ZmPDIL1-1	ND ^l	514		ND	AY739284	1	2	KDEL	S
I	ZmPDIL1-2	ND	512		ND	AY739285	1	2	KDEL	S
II	AtPDIL1-3	12	579	At3g54960	3	NM_115353	1	2	KDEL	S
II	AtPDIL1-4	12	597	At5g60640	5	NM_180903	1	2	KDEL	S
II	OsPDIL1-4	13	563		2	AK071514	1	2	KDEL	S
II	ZmPDIL1-3	ND	568		ND	AY739286	1	2	KDEL	S
II	ZmPDIL1-4	ND	561		ND	AY739287	1	2	KDEL	S
III	AtPDIL1-5	12	537	At1g52260	1	NM_104105	1	2	KDEL	S
III	AtPDIL1-6	12	534	At3g16110	3	NM_112481	1	2	KDEL	S
III	OsPDIL1-5	12	533		6	AK073970	1	2	KDEL	S
III	ZmPDIL1-5	ND	529		ND	AY739295	1	2	KDEL	S
IV	AtPDIL2-1	10	361	At2g47470	2	NM_130315	2	2	VASS ^{m,n}	S
IV	OsPDIL2-1	11	366		5	NM_185280	2	2	TFSS ^{m,n}	S
IV	OsPDIL2-2	11	371		1	NM_183927	2	2	TFSS ^{m,n}	S
IV	ZmPDIL2-1	ND	367		ND	AY739288	2	2	TFSS ^{m,n}	S
IV	ZmPDIL2-2	ND	366		ND	AY739289	2	2	IFSS ^{m,n}	S
V	AtPDIL2-2	9	443	At1g04980	1	NM_100376	2	2	KDDL	S
V	AtPDIL2-3	9	440	At2g32920	2	NM_128852	2	2	KDEL	S
V	OsPDIL2-3	9	441		9	AK062254	2	2	NDEL	S
V	ZmPDIL2-3	ND	439		ND	AY739290	2	2	NDEL	S
VI	AtPDIL5-1	4	146	At1g07960	1	NM_202059	5	1	DKEL	S
VI	OsPDIL5-1	4	147		3	AK063663	5	1	LQDS ^m	S'
VI	ZmPDIL5-1	ND	150		ND	AY739291	5	1	LEAD ^m	S'
VII	AtPDIL5-2	6	440	At1g35620	1	NM_103262	5	1	KKED	S
VII	OsPDIL5-2	5	423		4	AK069367	5	1	AHEE ^m	S
VII	OsPDIL5-3	5	425		2	ND ^o	5	1	AHED ^m	S
VII	ZmPDIL5-2	ND	418		ND	AY739292	5	1	IHEE ^m	S
VII	ZmPDIL5-3	ND	420		ND	AY739293	5	1	IHEE ^m	S
VIII	AtPDIL5-3	15	483	At3g20560	3	NM_112948	5	1	GKNI ^m	O
VIII	AtPDIL5-4	15	480	At4g27080	4	NM_118842	5	1	GKNF ^m	O
VIII	OsPDIL5-4	15	485		7	AK099660	5	1	GKNI ^m	O
VIII	ZmPDIL5-4	ND	483		ND	AY739294	5	1	GKNI ^m	O
XI	AtQSOX1	12	528	At1g15020	1	AY062528	5	1	EKER ^m	S
XI	AtQSOX2	12	495	At2g01270	2	AY090364	5	1	PRRR ^m	S
XI	OsQSOXL1	ND	513		5	AK121660	5	1	KNWN ^m	S'
XI	ZmQSOXL1	ND	511		ND	AY739305	5	1	KNWN ^m	S'
X	AtAPR1	4	465	At4g04610	4	NM_116699	5	1	NLVR ^m	O
X	AtAPR2	4	454	At1g62180	1	NM_104899	5	1	NLLR ^m	O
X	AtAPR3	4	458	At4g21990	4	NM_118320	5	1	NLVR ^m	O
X	AtAPRL4	4	310	At1g34780	1	NM_103198	5	1	SSSQ ^m	S
X	AtAPRL5	4	300	At3g03860	3	NM_111257	5	1	SDQS ^m	S
X	AtAPRL6	4	295	At4g08930	4	NM_116962	5	1	SASQ ^m	S
X	AtAPRL7	4	289	At5g18120	5	NM_121817	5	1	SQSA ^m	S
X	OsAPRL1	ND	475		7	XM_478340	5	1	NSLR ^m	O
X	OsAPRL2	5	282		6	AY739306	5	1	PSTS ^m	O
X	OsAPRL3	4	311		2	XM_467860	5	1	NELR ^m	S
X	OsAPRL4	ND	264		8	AY739307	5	1	SNLS ^m	S
X	OsAPRL5	4	301		3	AK073308	5	1	SRQA ^m	S
X	OsAPRL6	4	300		12	ND ^o	5	1	AVLD ^m	S'
X	ZmAPRL1	ND	461		ND	AY739296	5	1	NSLR ^m	O
X	ZmAPRL2	ND	466		ND	AY739297	5	1	NSLR ^m	O
X	ZmAPRL3	ND	323		ND	AY739298	5	1	SELR ^m	S'
X	ZmAPRL4	ND	321		ND	AY739299	5	1	SELR ^m	S'
X	ZmAPRL5	ND	267		ND	AY739300	5	1	PSLS ^m	S'

(Table continues on following page.)

Table I. (Continued from previous page.)

Phylogenetic Group ^a	PDIL Designation ^b	No. of Exons ^c	Predicted Polypeptide Size ^d	Accession No. of Gene ^e	Map Location (Chromosome) ^f	Accession No. of cDNA ^{g,h}	Structural Class ⁱ	TRX Domains ^j	COOH-Terminal Tetrapeptide	Localization ^k
X	ZmAPRL6	ND	303		ND	AY739301	5	1	GSTI ^m	S'
X	ZmAPRL7	ND	297		ND	AY739302	5	1	LLLD ^m	O
X	ZmAPRL8	ND	299		ND	AY739303	5	1	SRQA ^m	S
X	ZmAPRL9	ND	300		ND	AY739304	5	1	SRQA ^m	S'

^aAssigned phylogenetic group based on nucleotide and protein phylogenetic analyses (Figs. 2 and 3). ^bPlant PDIL nomenclature as described in text. ^cNumber of predicted exons. ^dNumber of amino acid residues. ^eArabidopsis Genome Initiative accession number. ^fChromosomal location of genes obtained from MaTDB (<http://mips.gsf.de/proj/thal/db/index.html>; Schoof et al., 2002) and Gramene (<http://www.gramene.org/>; Ware et al., 2002). ^gGenBank cDNA accession numbers with submissions from this work shown in bold. ^hAK108878 was found in database searches annotated as *Oryza sativa*; however, it is a likely endophytic contamination (data not shown). ⁱAssigned structural class based on number and position of TRX domains. ^jNumber of TRX domains. ^kS, Secretory pathway assignment by TargetP (<http://www.cbs.dtu.dk/services/TargetP/>; Emanuelsson et al., 2000) reliability value ≥ 0.6 ; S', secretory pathway assignment reliability value < 0.6 ; O, other localization predicted. ^lND, Not determined. ^mNo evidence for tetrapeptide involvement in ER retention. ⁿPredicted polypeptide has a COOH-terminal 57-amino acid domain implicated in ER retention (Monnat et al., 2000). ^oNo cDNA sequence available; tBLASTn searches of Gramene identified predicted transcripts GRMT00000163510 (OsPDIL5-3) and GRMT00000182570 (OsAPRL6).

the plant kingdom. Of the three single-domain groups in the tree (VI, VII, and VIII), one (VI) contained sequences from *P. patens*, *C. reinhardtii*, *G. lamblia*, and human, in addition to AtPDIL5-1S, and formed a larger subclade with group VIII and the two TRX domains of a PDIL sequence from *Acanthamoeba castellanii* (Fig. 2). The remaining group (VII) consisted of only a single branch leading to AtPDIL5-2S. Results from a maximum-likelihood analysis (data not shown) were consistent with those from the Bayesian analysis shown in Figure 2.

The Arabidopsis proteins of structural class 1 (phylogenetic groups I–III) were grouped with structural class 1 proteins from other organisms (Fig. 2). Furthermore, the NH₂- and COOH-terminal domains of these groups were well separated on the tree. By contrast, the NH₂- and COOH-terminal domains of a given phylogenetic group of structural class 2 proteins (phylogenetic groups IV and V) grouped closely together (Fig. 2).

Bioinformatic and Sequence Analysis of the PDIL Gene Space in Maize and Rice

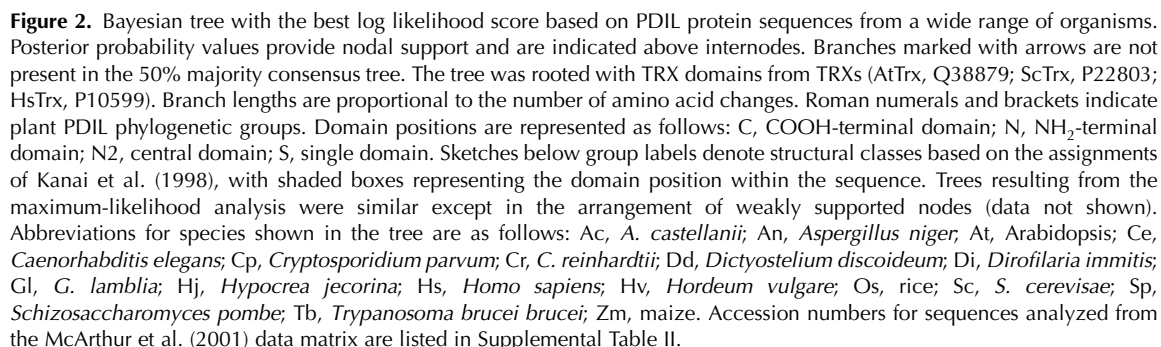
An unexpected discovery from the global phylogenetic tree was the presence of subgroups in which plant sequences were the only representatives of higher eukaryotes. This finding led us to investigate the sequential evolutionary events that led to the apparently greater diversity of PDIL proteins within the plant kingdom. We used the Arabidopsis sequences of the entire PDIL-related phylogenetic clade (QSOX, PDIL, and APRL) to identify counterparts in rice and maize for which we obtained full-length cDNA sequences. Accession numbers for the Arabidopsis, rice, and maize orthologs of this group are provided in Table I. From these, we constructed a data matrix (presented in Supplemental Table III) derived from nucleotide sequences of 88 TRX domains extracted from the 63 plant PDIL sequences of Arabidop-

sis (22), maize (22), and rice (19; Table I). A nucleotide sequence matrix was used because it included 3 times more characters in each sequence than the amino acid sequence matrix and thus would permit a more refined analysis.

Phylogenetic Analysis of Plant PDIL Nucleotide Sequences

The nucleotide-based gene phylogeny for the PDIL-related clade from Arabidopsis, rice, and maize (Fig. 3) generally agreed with the global amino acid phylogeny (Fig. 2) in topology and in having low support for basal nodes but much higher support for upper nodes. The tree topology showed clear paralogous and orthologous relationships and recognized the same phylogenetic groups. However, the nucleotide tree was better resolved and had greater support, especially for relationships among the phylogenetic groups. In the nucleotide phylogeny, the NH₂-terminal domains of phylogenetic groups I to III (structural class 1) and phylogenetic group IV (structural class 2) are grouped together in a clade separated from the COOH termini of these proteins. The single-domain groups, VI to VIII and QSOXL, are all nested within the NH₂-terminal clade containing groups I to IV, with VII being closely related to I_N, VI to IV_N, and VIII to QSOXL. By contrast, the two domains of group V (structural class 2) were resolved as sister groups (Fig. 3). The APRL proteins consist of a monophyletic group distinct from the other proteins (Fig. 3).

Our phylogenetic analyses suggested clearly that all 10 of the phylogenetic subgroups emerged before the divergence of monocots and eudicots, as a result of duplication of genes (e.g. groups I–III), duplication of domains (e.g. group V), and perhaps loss of domains (e.g. groups VI–VIII and QSOXL) that mostly may be traced back to an early eukaryotic evolution (Figs. 2 and 3). Recent duplications of different phylogenetic subgroups within a plant species also occurred in



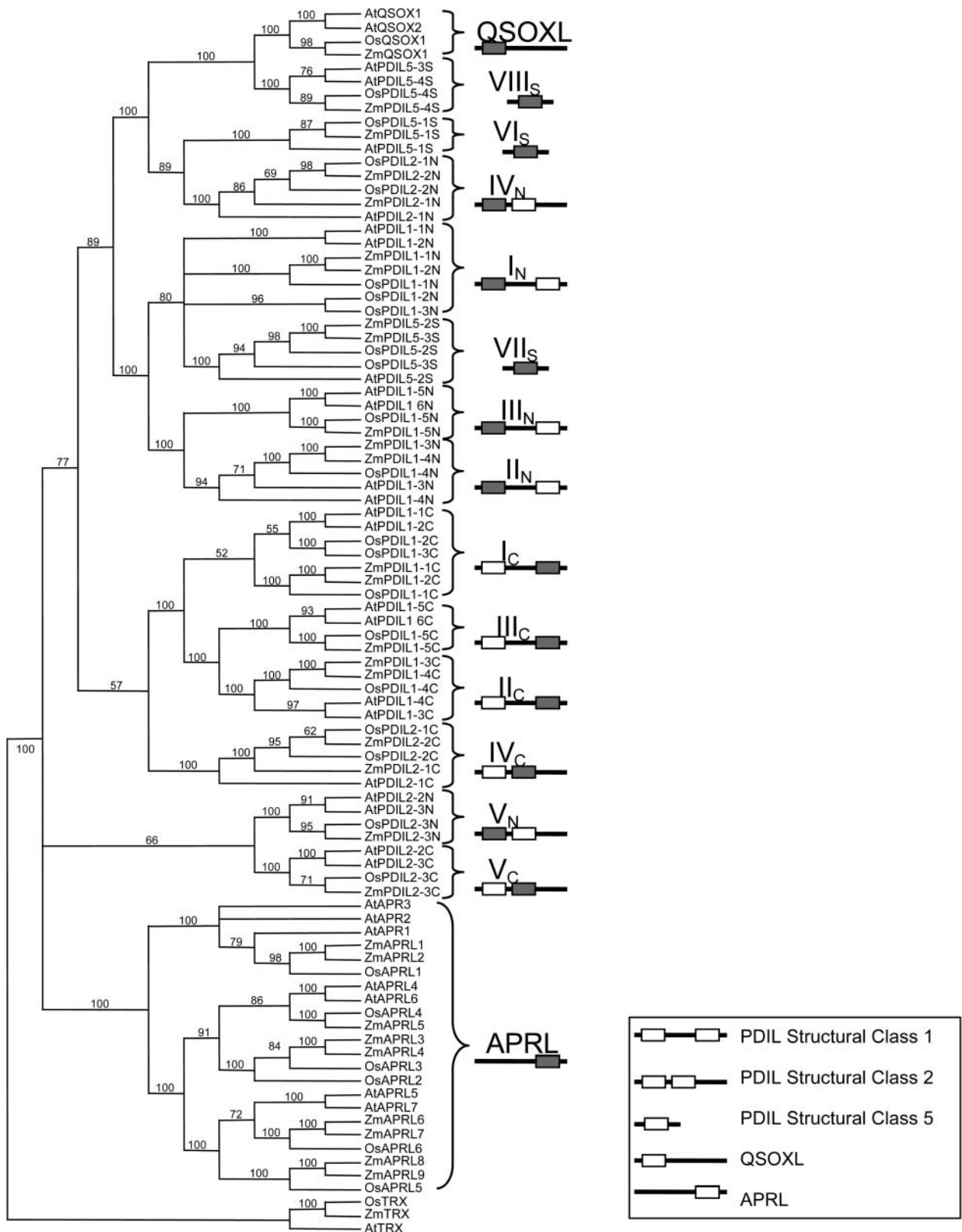


Figure 3. Bayesian consensus tree based on nucleotide sequences from the Arabidopsis plant PDIL clade. Posterior probabilities (>50%) are indicated above internodes. The tree was rooted with TRX domains from TRXs (AtTrx, At5g39950; OsTrx, NM_185032; ZmTrx, BM333382). Phylogenetic groups and domain organization are described in Figure 2. Trees resulting from the maximum parsimony analysis were similar except in the arrangement of nodes that are weakly supported (data not shown).

maize, rice, and Arabidopsis, as indicated by the presence of more than one sequence of the same species in a phylogenetic subgroup (Fig. 3; Table I).

Relationship between PDIL Phylogeny and Structural Features

The amino acid and nucleotide-based phylogenetic relationships were derived solely from the TRX domains. To determine if the domain relationships could be extended outside of these regions, we examined the physical characteristics predicted for the PDIL family members. Table I shows a comparison of the PDIL families from rice, maize, and Arabidopsis. Phylogenetic groups I, II, and III were similar in size, approximately 500, approximately 560, and approximately 530 amino acids, respectively, and made up structural class 1. The proteins in these phylogenetic groups were predicted by several analyses *in silico* to be secretory proteins with putative signal peptides and COOH-terminal KDEL-like ER retention sequences. Together with the data from amino acid and nucleotide phylogenetic analyses, such shared features in different PDIL groups offer further support for a similar evolutionary history. Proteins represented in subclade IV were approximately 360 amino acids in length but lacked a KDEL-like ER retention signal. Members of subclade V were longer (approximately 440 amino acids) and had KDEL-like ER retention signals at their COOH termini. The sequence and domain differences between subclades IV and V offer additional circumstantial support for the independent evolution suggested by the phylogenies.

Aside from having single TRX domains, class 5 PDIL proteins (groups VI, VII, VIII, APRL, and QSOXL) shared few structural features across groups. Such diversity is not surprising given the differing evolutionary origins suggested from the phylogenetic analysis. Group VI PDIL proteins were the smallest members of the plant PDIL family with only approximately 150 amino acids. Members of groups VII, VIII, and QSOXL were much larger (418–528 amino acids). PDIL proteins from groups VI and VII were predicted to be secretory proteins with signal peptides. None of the single TRX domain proteins had KDEL-like sequences, but one (AtPDIL5-2) of the five members of group VII had a COOH-terminal dilysine KKXX membrane-anchoring motif. All of the group VII proteins, as well as the QSOXL and group VIII proteins, were predicted to be membrane proteins by analysis with TMpred and TMHMM programs (Hofmann and Stoffel, 1993; Sonnhammer et al., 1998; Benghezal et al., 2000). The APRL clade contained two subclades. Members of one subclade (AtAPRL1, 2, and 3; ZmAPRL1 and 2; OsAPRL1) contained a 3'-phosphoadenylyl sulfate (PAPS) reductase domain in the full-length sequence, whereas members of the other APRL subclade lacked this domain. Proteins with PAPS reductase domains were predicted to be approximately 460 amino acids in length with NH₂-terminal

sequences predicted to target the proteins to the chloroplast; proteins in the other APRL subclade were smaller (approximately 300 amino acids), and most had NH₂-terminal ER signal sequences. A five-member terminal group from this putatively ER-targeted subclade (AtAPRL5 and 7; ZmAPRL6 and 7; OsAPRL6) was distinguished by the presence of a predicted membrane-spanning domain (data not shown). Overall, the phylogenetic groupings derived from the PDIL-active TRX domain sequences showed a positive correlation with both protein size and common targeting features within a group.

Distribution of PDIL Genes in the Arabidopsis Genome

Regional and genome duplication events are a major component of gene family development (Wendel, 2000). Members of a gene family can be distributed randomly throughout the genome; however, 17% of all duplicated genes in Arabidopsis have been reported as occurring in tandem (The Arabidopsis Genome Initiative, 2000). To determine the significance of duplication events in the plant PDIL gene family evolution, we investigated the chromosomal distribution of the plant PDIL sequences in Arabidopsis and rice. Of the 22 Arabidopsis sequences in the PDIL clade, 8 were found in duplicated regions of the genome (AtPDIL1-1, AtPDIL1-2, AtPDIL1-5, AtPDIL1-6, AtPDIL2-2, AtPDIL2-3, AtAPRL5, and AtAPRL7). One event was associated with regions duplicated on chromosome 1 (AtPDIL1-1 and AtPDIL1-2), two occurred in regions duplicated between chromosomes 1 and 2 (AtPDIL2-2 and AtPDIL2-3) and 1 and 3 (AtPDIL1-5 and AtPDIL1-6), and a third occurred between chromosomes 3 and 5 (AtAPRL5 and AtAPRL7). Five other PDIL sequences were found on nonduplicated regions of chromosome 1, and the remaining nine were found on other chromosomes. For rice, the 19 QSOXL-PDIL-APRL sequences mapped to 11 of the 12 chromosomes (Table I).

Maize PDI Gene Expression

Participation of the major PDIs in the essential cellular process of protein folding led us to assay for expression of PDIL mRNAs throughout the maize plant. We searched the massively parallel signature sequencing (MPSS) database (Brenner et al., 2000) from Pioneer Hi-Bred International for tissue-specific abundance of maize PDIL mRNAs in libraries from different tissue samples including a number of different stages of developing seeds. MPSS allows direct quantitative comparisons of mRNA abundance within and among samples based on generation of more than one million 17-nucleotide-long expressed sequence tags (ESTs) from each library. Table II shows a comparison of relative mRNA prevalence across representative tissues (B73 inbred) selected from a comprehensive gene expression data set of several hundred libraries. Eight of the maize PDIL genes were

Table II. Maize PDIL transcript levels detected by MPSS

Library Source ^a	ZmPDIL ^b												ZmAPRL ^b									ZmQSOXL ^b
	1-1	1-2	1-3	1-4 ^c	1-5	2-1	2-2	2-3	5-1	5-2	5-3	5-4	1	2 ^d	3	4	5	6	7	8	9	1
Immature ear	621	0	78	0	30	82	719	136	0	0	0	115	67	0	49	81	39	55	2	67	6	0
Embryo 21 DAP	1,648	0	86	0	0	28	245	46	0	0	0	80	0	0	177	0	6	0	0	0	0	0
Endosperm 21 DAP	6,747	0	117	0	0	70	91	246	17	0	17	145	19	0	36	14	24	0	0	18	30	0
Whole kernels 8 DAP	933	0	11	0	23	152	851	106	38	13	6	113	15	0	94	0	23	7	0	11	2	13
Mature leaf	1,019	0	26	0	0	200	81	217	20	5	0	37	101	0	46	0	32	10	6	28	10	13
Apical meristems before floral transition	734	0 ^e	43	0	47	61	337	71	9	3	0	77	23	0	19	15	38	41	4	16	21	0
Mature pollen	4	0	0	0	0	9	62	6	0	0	0	0	0	0	54	0	162	0	0	14	9	0
Primary root V2	501	0	54	0	12	149	296	106	9	5	0	65	176	0	90	36	31	0	0	34	12	0
Stalk	562	0	44	0	56	221	762	89	0	0	0	55	28	0	409	4	13	0	0	13	0	0
Tassel spikelets	2,639	0	297	0	18	129	441	186	31	4	0	106	23	0	147	21	16	5	0	0	35	0
quartet stage																						
Seedling	634	0	37	0	0	167	166	106	0	0	0	26	0	0	185	0	0	23	0	0	23	0

^aLibraries shown were from the common inbred line, B73. ^bmRNA abundance values are in part per million. ^cESTs found in cDNA libraries from callus. ^dESTs found in cDNA libraries from abiotic/biotic-stressed tissues. ^eFound at low level in specialized MPSS meristematic tissue libraries (data not shown).

expressed in at least 10 of the 11 libraries shown. The major PDI (ZmPDIL1-1) was highly expressed in all organs/tissues surveyed except for pollen, in which it was barely detectable. Other PDIL members were less abundant and showed marked differences in relative expression levels within and among groups. For paralogs in structural class 1 (phylogenetic groups I and II), usually one was expressed at a much higher level than others (Table II, compare ZmPDIL1-1 with 1-2 and 1-3 with 1-4). For structural class 2, by contrast, all three maize paralogs were expressed at similar levels across the range of organs/tissues surveyed (Table II, ZmPDIL2-1, 2-2, and 2-3). Even so, ZmPDIL2-1 and ZmPDIL2-2 had quantitatively different expression patterns, yet they were both in phylogenetic class V and shared sequence identity of 81% at the amino acid level. Expression of genes encoding single TRX domain proteins was quite diverse. Phylogenetic groups VI to VIII and QSOXL encoded rare mRNAs with the exception of ZmPDIL5-4 (a group VIII member). ZmPDIL5-4 expression resembled that of the major PDIL group, ZmPDIL1, which was dramatically underrepresented in pollen. The members of the APRL subclade also showed a variety of expression patterns, with some being prevalent across several organs/tissues (APRL1, 3, 5, 8, and 9) and others being quite rare (APRL2, 4, 6, and 7). Examination of PDIL MPSS data from Arabidopsis in the Pioneer Hi-Bred database and a public database of 17 available tissue libraries (<http://mpss.udel.edu/at/>; Meyers et al., 2004) reflected a similarly diverse expression pattern (N.L. Houston and R. Jung, unpublished data). Aside from MPSS experiments, ESTs of many of the maize PDILs were found at low levels in most tissue- or organ-specific cDNA libraries in the Pioneer Hi-Bred database but appeared to be prevalent in cDNA libraries derived from chemically perturbed Black Mexican Sweet maize cell suspension cultures (data not shown). Thus, at the level of mRNA

expression, a number of different PDIL genes follow a broad expression pattern expected for a generalized housekeeping function. Other PDIL mRNAs are rare, yet they may be highly expressed in specific tissues and/or in response to specific developmental or environmental cues.

Induction of a Single-Domain Group VI PDIL Protein during ER Stress

We were particularly interested in characterizing the proteins of the single-domain phylogenetic group VI, which had only one member from each plant. In addition, these proteins were the smallest members of the PDIL family, and the Arabidopsis representative grouped with the TRX domains from the primitive, nonvascular green plants *P. patens* and *C. reinhardtii*. Although members of this group had NH₂-terminal domains predicted to act as signal peptides, they lacked classical ER retention signals. To investigate whether or not accumulation of this PDIL group reflected the ER stress induction associated with ER molecular chaperones, we characterized accumulation of the maize member, ZmPDIL5-1, in tissues exhibiting an ER stress response. Figure 4 shows the results of replicate immunoblots probed for the small PDI, ZmPDIL5-1, and various marker proteins in the maize *fl2* mutant that exhibit ER stress during endosperm development (Boston et al., 1991). ZmPDIL5-1 cross-reacting material was increased between 8 and 15 d after pollination (DAP) and persisted through mid-maturation of the kernel (Fig. 4). The major PDI, ZmPDIL1, showed a similar accumulation pattern and induction in developing *fl2* endosperm (Fig. 4), as reported by Li and Larkins (1996). Likewise, the molecular chaperones BiP and calnexin were strongly induced between 8 and 15 DAP, whereas calreticulin produced a strong signal at 8 and 10 DAP, with a slight intensification at later developmental stages. Proteins

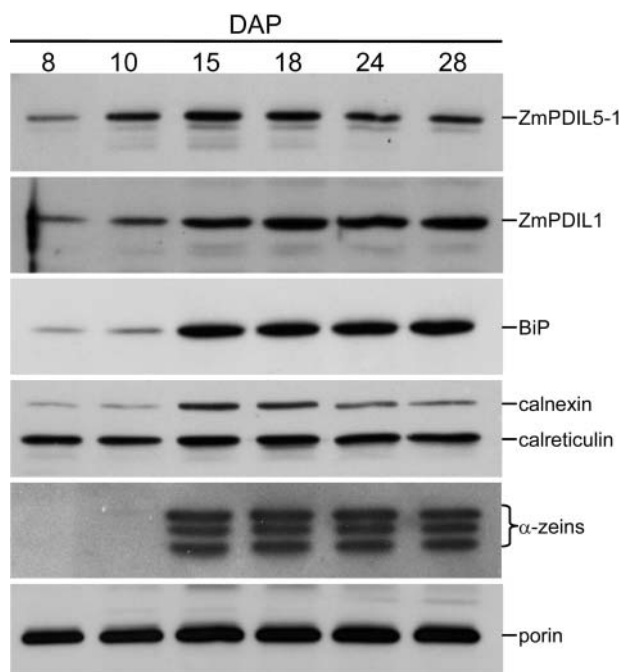


Figure 4. Accumulation of ZmPDIL5-1 protein during endosperm development. Crude protein extracts from equal fresh weights of *fl2* endosperm were fractionated by SDS-PAGE and immunoblotted to detect ZmPDIL5-1. The ZmPDIL5-1 blot was probed with antibody against an 18-amino acid COOH-terminal peptide. Duplicate blots were probed with antisera raised against the major wheat PDI (which cross-reacts with ZmPDIL1 proteins), BiP, calnexin/calreticulin, 22-kD α -zein, and porin. Developmental stages of endosperm samples are indicated at top in DAP.

cross-reacting with antibodies raised against α -zein were detected in samples harvested 15 DAP and later. The upper band of the α -zein triplet represents the mutant 24-kD zein, which is responsible for the ER stress response (Coleman et al., 1995, 1997). The mitochondrial outer membrane protein, porin, was used as a protein-loading control. Porin cross-reacting material was detected as a strong signal in each sample regardless of developmental stage.

To characterize the responsiveness of ZmPDIL5-1 expression to ER stress, we extended our investigation to include a normal maize inbred and endosperm mutants other than *fl2* that also exhibit an ER stress response (Coleman et al., 1995; Kim et al., 2004). Figure 5 shows an immunoblot analysis for endosperm samples. Blots were probed for PDIL5-1 and the molecular chaperone BiP for comparison prior to (10 DAP) and during (18 DAP) the ER stress response. Blots of 16-kD and 27-kD γ -zeins were used to track zein accumulation in all maize lines and ER stress in the mutant *Mc*, which has a mutation in the 16-kD γ -zein (J. Gillikin, R. Jung, and R.S. Boston, unpublished data). A blot probed with antisera against the mitochondrial α -ATPase subunit was used as a protein-loading control (Luethy et al., 1993). Increased signals of ZmPDIL5-1 and the molecular chaperone BiP were

detected after the onset of zein accumulation in the ER stress mutants *De*-B30* and *Mc* and the double mutant *Mc opaque-2* (*Mco2*), but not in the normal control.

ZmPDIL5-1 Is Not Associated with Endosperm Endomembrane Fractions

The observation that PDIL5-1 was induced in response to ER stress was confounded by the lack of an obvious ER retention signal and prompted us to investigate its subcellular location. Endosperm from normal and *fl2* maize lines was fractionated through linear Suc gradients and probed by immunoblot analysis for PDIL5-1 and various marker proteins (Fig. 6). PDIL5-1 was detected only in the upper fractions of the gradient. This localization was not a function of its participation in the ER stress response, as judged by detection of the protein in equivalent fractions regardless of whether the samples were extracted from normal or *fl2* endosperm. By contrast, all of the reference proteins showed an endomembrane association, as judged by immunological detection across the gradient.

The 22-kD α -zeins, found in the densest regions of the gradient, marked the protein body fractions. The porin marker was most abundant in fractions slightly less dense than the protein bodies, as expected for a mitochondrial protein. The molecular chaperones calnexin, calreticulin, and PDIL1 localized primarily in the denser portion of the gradient, as expected for ER and protein body proteins. PDIL1 and calreticulin were also detected in the upper portion of the gradient. In addition to cytosol, these fractions likely represent luminal con-

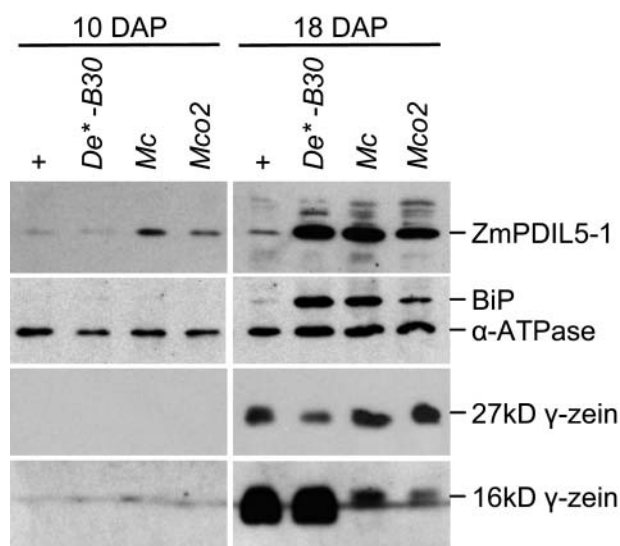


Figure 5. Immunoblot analysis of proteins from maize endosperm mutants *De*-B30*, *Mc*, and *Mco2*. Samples contained crude protein extracts from equal fresh weights of maize endosperm harvested 10 and 18 DAP. Samples from a normal inbred control are labeled as +. Replicate immunoblots were probed with antisera against ZmPDIL5-1, BiP and α -ATPase, 27-kD γ -zein, and 16-kD γ -zein.

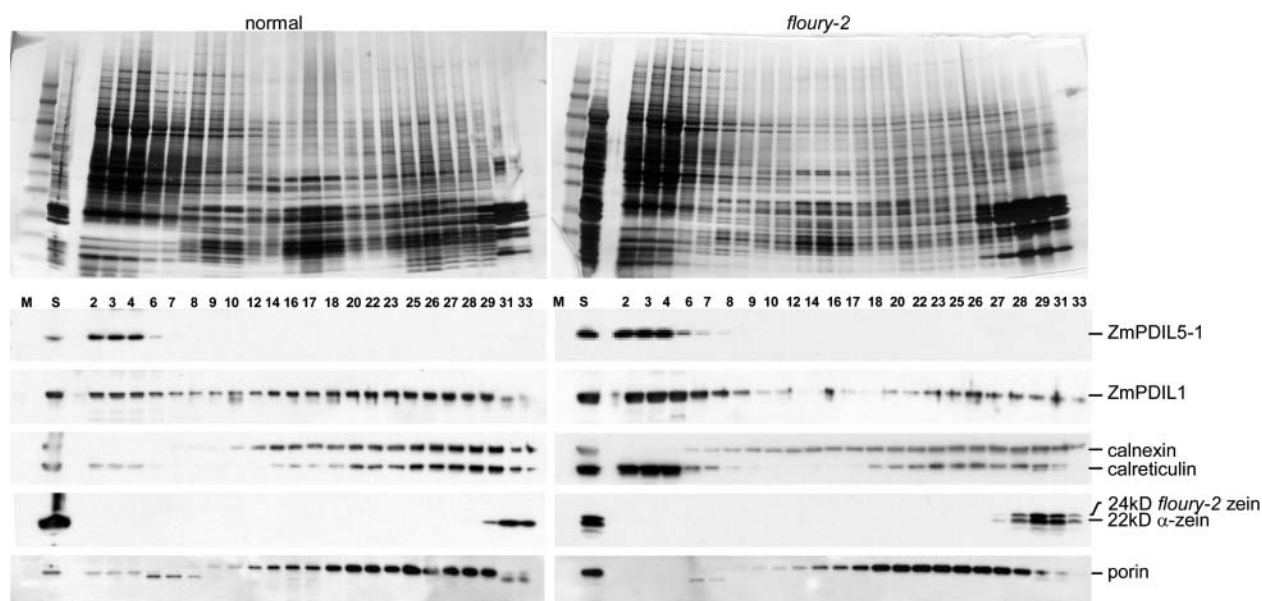


Figure 6. Subcellular localization of ZmPDIL5-1 in normal and *fl2* endosperm. Protein extracts from normal and *fl2* endosperm harvested 23 DAP were separated through 10% to 60% linear Suc gradients. Selected fractions from the gradients were analyzed by SDS-PAGE and immunoblotting. Proteins were visualized by silver staining. Replicate immunoblots were probed with antisera against ZmPDIL5-1, the major wheat PDI (which cross-reacts with ZmPDIL1), calnexin/calreticulin, 22-kD α -zein, and mitochondrial porin. Arabic numbers indicate selected gradient fractions. M, Molecular mass markers; S, protein extract loaded onto the gradient.

tents of lysed organelles, as judged by the absence of the membrane-associated calnexin from this region.

DISCUSSION

Classical distinctions within the TRX superfamily relied on size, number, and organization of TRX domains, subcellular location, and tetrapeptide-active site motifs along with enzymatic activities to separate proteins into functional groupings. As data from EST and genome sequencing projects became available, these groupings were confounded with more and more exceptions. We encountered this problem when BLAST searches for homologs of oxidoreductases uncovered in a large-scale RNA-profiling study returned large numbers of proteins with TRX domains. In an effort to develop a complete, phylogenetically supported data set, we mined Arabidopsis genomic and EST databases for sequences encoding TRX domains and organized the domains based on predicted phylogenetic relationships. Sequences within the phylogenetic tree grouped according to known or predicted enzymatic activities in both Bayesian (Fig. 1) and maximum-likelihood analyses (data not shown). Sequences known or predicted for different enzyme activities are placed in different clades. Such a phylogenetic pattern is suggestive that functional divergence after gene duplication may have played a critical role in the primary diversification of the gene family.

An unexpected relationship was the association of QSOXL and APRL proteins with the PDIL subclade.

This grouping, revealed in the Arabidopsis phylogeny and supported in higher resolution trees that included maize and rice (Fig. 3), demonstrates the utility of a combined bioinformatic/phylogenetic approach to characterize complex gene families. Furthermore, such an approach gave us a comprehensive data set that represents a near-complete, if not complete, catalog of PDI-related members from Arabidopsis, rice, and maize (Table I).

Our approach of verifying genomic data with cDNA sequences based on completely (and generally redundantly) sequenced cDNA clones for the PDIL group also led to correction of misannotations in previous curation efforts based on gene model/splicing predictions and in reports that relied on the resulting genomic predictions (Meiri et al., 2002). For maize, which has extensive sequence polymorphisms, cDNA sequences were compared to genome survey sequences (<http://www.maizegdb.org/>). This step allowed us to verify that closely related sequences originated from distinct genes and not from alleles that might have been isolated from different inbred lines.

The phylogenies revealed that major subclades corresponded to major orthologous groups containing sequences from both plants and animals. This early branching suggested that the groups were ancient within the evolution of PDIL proteins. Plant proteins containing two TRX domains formed five phylogenetic groups. The NH₂-terminal domains of structural class 1 (groups I–III) united as clades that were separated from the COOH-terminal clades from the same class 1 groups. The phylogenetic relationship among these

domains was preserved in both lower and higher eukaryotes, as expected for an emergence early in eukaryotic evolutionary history. In contrast with the domains from proteins in structural class 1, the NH₂- and COOH-terminal domains of group V (structural class 2) showed a close association. Such results are suggestive that members of group V emerged through independent domain duplication. The two domains of group IV (structural class 2) did not group together. Because subclades of the NH₂- and COOH-terminal domains of group IV lacked strong support, no inference could be made about their origins.

The distribution of the major orthologous groups across the tree is consistent with the hypothesis that duplications of the TRX domains occurred prior to the divergence of plants and made a major contribution to the evolution of this complex gene family from a common ancestral eukaryote (McArthur et al., 2001). Such preservation of major orthologs through time is suggestive that the different PDIL groups may have separate and important functions in the cell and that maintaining such functions has been a major evolutionary constraint. Among the recently duplicated paralogs, preservation of similar functions is expected to be more common than acquisition of a unique function. In such a case, functional redundancy might be avoided by temporal or spatial differences in gene expression or by maintenance of only a subset of the major function by each member. The eight AtPDIL proteins that lie in known duplicated regions of the genome would be good candidates for testing functional divergence or redundancy among paralogs.

It is intriguing that the TRX domains from QSOXL and APRL proteins are included in the PDIL clade. The QSOXL proteins are nested within the PDIL clade and are sisters with group VIII PDIL proteins. A phylogenetic relationship between QSOXL and PDIL was determined previously by Coppock et al. (1998); however, only one plant sequence was included in the analysis. Sister groups VIII and QSOXL share similar sequence characteristics outside of the TRX domain used in the phylogenetic analysis including a putative signal peptide, a membrane-spanning domain, and lack of a KDEL-like ER retention signal. These similarities in the primary sequence suggest that function may also have been evolutionarily conserved. Although QSOXL proteins contain only a single TRX domain, PDI activity may be possible if each PDI domain functions independently (Vuori et al., 1992). Although further study will be required to determine if QSOXL proteins have PDI activity, in RNase refolding studies *in vitro*, an avian QSOX protein and the major PDI appeared to act cooperatively to generate active RNase (Hoober et al., 1999).

APR proteins have been well studied in the sulfate assimilation pathway (Kopriva and Koprivova, 2004). The phylogenetic analysis presented here extends our knowledge of these proteins by showing that they have a strongly supported phylogenetic relationship with the PDIL family. Within the APRL group are

subgroups that differ in size, predicted subcellular localization, and domain composition (Fig. 3; Table I). Members of one subgroup have a PAPS reductase domain in addition to the TRX domain. Members of the other group have only a TRX domain. Because none of the other proteins in the PDIL clade have the PAPS reductase domain, it is likely that it was acquired by an ancestral TRX domain protein. Included in the APRL PAPS domain subgroup are three AtAPR proteins (AtAPR1, 2, and 3) that have been shown to have adenylyl sulfate activity (Gutierrez-Marcos et al., 1996; Setya et al., 1996; Wray et al., 1998). Whether this and/or the other APRL subgroup have retained isomerase activity will require further investigation.

Typically, PDIL proteins have multiple TRX domains capable of functioning independently (Vuori et al., 1992). Therefore, it was unexpected to find that 5 of the 10 phylogenetic groups in plants (VI, VII, VIII, QSOXL, and APRL) consisted of proteins with only 1 TRX domain. The close phylogenetic relationship between the nucleotide sequences encoding single-domain group VII proteins and the NH₂-terminal TRX domains of PDIL proteins of groups I_N, II_N, and III_N in structural class 1 is consistent with the group VII proteins having emerged by domain loss from a two-domain PDIL precursor. A similar domain loss from a different structural class, group IV_N, would explain the emergence of groups VI, VIII, and QSOXL in a subclade with the group IV_N members. After the domain loss, the QSOXL proteins would have evolved further by the addition of an Erv1-like domain. An alternative hypothesis for the evolution of groups VI and VIII, however, cannot be ruled out. This hypothesis, that proteins in groups VI and VIII retained an ancestral domain structure, is based on the relationship between groups VI and VIII and proteins from lower eukaryotes. Because at least one such single-domain protein from *Giardia* has been demonstrated to have isomerase activity, it seems feasible to speculate that members of groups VI and VIII might also be enzymatically active, either alone or as part of a redox chain requiring interacting partners (Knodler et al., 1999).

Examination of features other than the phylogenetic relationships revealed that the single-domain PDIL proteins had neither KDEL-like ER retention signals (groups I–III and V) nor the conserved COOH-terminal ER retention domains found on the multidomain PDIL proteins of group IV (Monnat et al., 2000). The significance of these tags is unclear, however, as other methods of ER retention have been identified. For example, retention by an additional domain allowed protein-protein interactions between a PDIL from human and other ER resident proteins (Russell et al., 2004). Eps1, a yeast PDIL protein, has a transmembrane domain and is associated with a function in the ER, having been shown to be involved in ER-associated protein degradation (Wang and Chang, 1999). Members of phylogenetic groups VII and VIII have COOH-terminal sequences predicted to serve as

transmembrane domains that could act as ER anchors (Wang and Chang, 1999; Matsuo et al., 2001).

Lack of a means for ER retention could also indicate alternative targeting. Despite the presence of canonical signal peptides and KDEL-like motifs, several PDIL proteins have previously been reported to be localized in the chloroplast and at the cell surface (Kim and Mayfield, 1997; Jiang et al., 1999; Zai et al., 1999; for review, see Turano et al., 2002). In normal maize, ZmPDIL5-1, a group VI protein, failed to colocalize with the molecular chaperones such as the major PDI in either protein bodies or ER (Fig. 6). Instead, it consistently stayed at the top of the gradient, regardless of whether the sample was derived from endosperm of stressed or unstressed seed. This spatial separation from the endomembrane system is particularly intriguing in light of experimental data that showed its strong induction during ER stress (Figs. 4 and 5). Whether this protein transiently resides in the lumen of the ER and/or exhibits an ER-related function remains to be determined.

The phylogenetic and comparative analyses presented here form the foundation for further investigation of the PDIL family, particularly with respect to protein localization and enzymatic activity. All members of the TRX superfamily contain a TRX domain that includes a CXXC motif. Traditionally, PDI proteins were characterized by a CGHC motif, which is found in all of the multidomain PDIL proteins as well as in those from group VII. This conservation is consistent with group VII having arisen by domain deletion from group I. Members of groups VI (CKHC) and VIII (CYWC/S) have motifs unlike any others in the plant PDIL family including orthologs in *P. patens* and *C. reinhardtii*. New insights into the single-domain PDIL proteins may well be found through determining if these changes in the tetrapeptide active site motifs preserved the ancestral isomerase function or caused functional divergence. Determining the enzymatic activities of these single-domain proteins and their capacity to function independently or act as links with other proteins in a redox chain will be important for understanding their biological function in production and/or isomerization of disulfide bridges, as well as their physiological roles.

MATERIALS AND METHODS

Database Searches and Sequence Analyses

An initial search for sequences encoding putative active TRX domains was carried out in Arabidopsis (*Arabidopsis thaliana*) by iterative BLAST and word searches of EST and nucleotide sequence databases (Altschul et al., 1997). After manual editing to correct putative coding sequence prediction errors and the removal of redundant sequences, a genomic locus-based set of 104 gene sequences was compiled encoding the likely near-complete set of Arabidopsis proteins containing TRX domains. The 117 predicted active-TRX domains were extracted from this protein set by homology to consensus sequences from the National Center for Biotechnology Information Conserved Domain Database (Marchler-Bauer et al., 2002), entered as separate sequences into a matrix (Supplemental Table I), and designated with an N or C after the gene name to denote the location of the domain (N, NH₂; C, COOH) within the

double or multiple TRX domain proteins. The final set of 22 AtPDIL sequences included the entire PDIL clade shown in Figure 1. These 22 AtPDIL sequences were used to identify orthologous sequence sets in rice (*Oryza sativa*) and maize (*Zea mays*) by iterative BLAST searching (Altschul et al., 1997) of the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>), Gramene (<http://www.gramene.org/>; Ware et al., 2002), and Pioneer Hi-Bred EST databases. Genomic and EST or cDNA sequences from Arabidopsis and rice were assembled, compared, and edited using Sequencher software (Genecodes, Ann Arbor, MI). Additionally, high-confidence reference sequences to confirm (and to correct when necessary) coding sequence predictions from Arabidopsis and rice were obtained from maize by systematically obtaining full-length cDNA sequence information from all identified maize PDIL genes. For further phylogenetic analyses, we constructed an amino acid data matrix that combined the 13 AtPDIL TRX domains with a data matrix constructed by McArthur et al. (2001) that is representative of TRX domains in eukaryotes (Supplemental Tables I and II). The combined data matrix did not include QSOX or APRL proteins. In addition, we included the TRX domains from the single-domain PDIL representative of *Physcomitrella patens* and *Chlamydomonas reinhardtii*. We also compiled a nucleotide data matrix of 88 plant PDIL TRX domains from 63 Arabidopsis, maize, and rice nucleotide sequences (Supplemental Table III).

Predicted plant PDIL amino acid sequences were further analyzed with bioinformatic tools to predict subcellular localization, transmembrane domains, and chromosomal location. Putative signal peptides were predicted with the neural network-based program TargetP (<http://www.cbs.dtu.dk/services/TargetP/>; Emanuelsson et al., 2000). Subcellular localization predictions were considered strong if the reliability value was greater than 0.6, and predictions below that threshold were confirmed with the Web-based program PSORT (<http://www.psort.org/>; Nakai and Kanehisa, 1991; data not shown). We used TMHMM (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>; Sonnhammer et al., 1998), a hidden Markov model-based program, and TMPred (ISREC Bioinformatics, Epalinges, Switzerland; http://www.ch.embnet.org/software/TMPRED_form.html; Hofmann and Stoffel, 1993) to predict transmembrane helices. Arabidopsis PDIL sequences were analyzed with the MATDB: Redundancy Viewer (http://mips.gsf.de/proj/thal/db/gv/rv/rv_frame.html) to identify sequences in duplicated chromosomal regions. Rice chromosome assignments were found by BLASTp searches in Gramene (<http://www.gramene.org/>; Altschul et al., 1997; Ware et al., 2002).

Phylogenetic Analysis

Nucleotide and protein data matrices were aligned using ClustalX multiple sequence alignment program (version 1.8; Supplemental Tables I–III; Thompson et al., 1997) with default gap penalties and further adjusted manually to minimize gap insertions. The Arabidopsis TRX and global PDIL protein sequences were analyzed with both maximum-likelihood and Bayesian methods using the PHYLIP program (Phylogeny Inference Package, version 3.6, Department of Genetics, University of Washington, Seattle; Felsenstein, 2002) and MrBayes 3.0 (Huelsenbeck and Ronquist, 2001). Both maximum-likelihood and Bayesian inference analyses of protein matrices were conducted under the Jones Taylor Thornton substitution model (Jones et al., 1992) with inclusion of observed amino acid frequencies, estimated proportion of invariant sites, and estimation of among-site rate variation for the remaining sites according to a gamma distribution. Analyses using a Bayesian inference procedure were carried out on MrBayes 3.0 (Huelsenbeck and Ronquist, 2001). Four Markov chains starting with a random tree were run simultaneously for 200,000 generations with sampling from the trees made every 100th generation. Altogether 700 trees (of 2,000) were discarded as the burn-in for the Arabidopsis TRX superfamily and 400 (of 2,000) trees were discarded for the global PDIL analyses, based on the stationary phase. The remaining trees were imported to PAUP (Phylogenetic Analysis Using Parsimony, version 4.0; Sinaur Associates, Sunderland, MA) to construct the 50% majority rule trees (Figs. 1 and 2; Swofford, 2002). The frequencies of clades on the 50% majority consensus tree provided the posterior probability support for the clades (e.g. Figs. 1 and 2).

The nucleotide sequences were analyzed using Bayesian methods implemented in MrBayes 3.0 (Huelsenbeck and Ronquist, 2001). The general time-reversible evolution model (Rodríguez et al., 1990) with a proportion of invariant characters and gamma distribution were implemented in the Bayesian analysis. This model was the best model chosen by an Akaike Information Criterion test conducted in modelTest (Posada and Crandall, 1998). Bayesian analysis was performed as described above for protein sequences, except that only 100,000 generations were run and 460 (of 1,000)

trees were discarded as the burn-in. Inference about relationships was based only on the remaining 540 trees. From these, the 50% majority consensus tree was constructed. The frequencies of clades on the 50% majority consensus tree provided posterior probability support for the clades (e.g. Fig. 3).

Plant Materials

Normal B37 and W64A inbreds, near isogenic mutant lines W64A *fl2*, B37 *Mc*, B37 *De*-B30*, and the double mutant B37 *Mco2* were grown during summer field seasons at the Central Crops Research Station (Clayton, NC). The near-isogenic B37 mutant lines were developed and provided by F. Salamini (Max-Planck-Institut für Züchtungsforschung, Cologne, Germany). The normal maize Pioneer R03 inbred was grown in the Johnston 2004 summer nursery. At the indicated days after pollination, ears were frozen in liquid nitrogen and kernels harvested from the ear for storage at -80°C until use.

Quantitative Expression Analysis of PDIL Genes by MPSS

The mRNA from a variety of maize tissue samples was previously isolated and MPSS performed by Lynx Therapeutics (Hayward, CA) as described (Brenner et al., 2000). The resulting MPSS ESTs, here defined as the first 17 base pairs, including and following downstream of the most 3' Sau3A site (GATC) of a gene transcript, are quantified and reported on a ppm basis (1–2 million sequencing reactions performed per sample) in a searchable database. The quantity of PDIL MPSS ESTs in each tissue sample were then obtained by queries of this database with the exact string of the conceptual MPSS ESTs identified for each of the PDIL gene transcripts (Meyers et al., 2004).

Protein Extraction and Immunoblot Analysis

Equal fresh weights of frozen endosperm from normal and mutant inbreds were ground with mortar and pestle. Buffer B (10 mM Tris-HCl, pH 8.5 at 25°C , 10 mM KCl, 5 mM MgCl_2 , and 7.2% [w/v] Suc) was added at a 1:2 (w/v) ratio (Shank et al., 2001). Cellular debris was removed by centrifugation at 325g for 5 min at 4°C .

Protein extracts were diluted 1:10 (v/v) in SDS sample buffer and boiled for 5 min before separation through 10% or 15% (w/v) SDS polyacrylamide gels (Laemmli, 1970) and then transferred to polyvinylidene difluoride membranes in a submerged blotting system (mini-trans blot; Bio-Rad Laboratories, Hercules, CA) with transfer buffer (48 mM Tris base, 39 mM Gly, pH 9.2) for 3 h at 50V. Membranes were blocked for 1 h with Tris-buffered saline containing 0.1% (v/v) Tween 20 and 5% (w/v) nonfat dry milk. Immunoblots of ZmPDIL5-1 were probed with a peptide antibody to the COOH terminus of the protein (NFVLNEAEKAGEAKLEAD) and anti-rabbit secondary antibody in Tris-buffered saline containing 0.1% (v/v) Tween 20. Other immunoblots were probed for the mitochondrial α -ATPase subunit (Luethy et al., 1993), porin, or others indicated below. Polyclonal antibodies raised against purified calreticulin (Pagny et al., 2000) also cross-reacted with maize calnexin. ZmPDIL1 proteins were detected with a polyclonal antibody raised against the major PDI from wheat (*Triticum aestivum*; Shimoni et al., 1995). This antibody detected the major ER stress-induced polypeptide of approximately 55 kD along with several minor polypeptides of similar size as judged by two-dimensional gel electrophoresis (N.L. Houston and R.S. Boston, unpublished data). We cannot determine whether these additional polypeptides are isoforms encoded by the major ZmPDIL1-1 gene or products of other members of the ZmPDIL1 gene family. Anti-BiP antibody (ID9) was obtained from StressGen Biotechnologies (Victoria, British Columbia, Canada). Antibody raised against a 16-kD γ -zein peptide also weakly cross-reacted with maize 27-kD γ -zein proteins (Woo et al., 2001). The α -zeins were detected with a polyclonal antibody (R166) against the 22-kD class of α -zeins (Fig. 4; Esen, 1988) or an antibody made against a bacterially expressed 22-kD α -zein peptide (Fig. 6; Woo et al., 2001). Proteins that cross-reacted with antibodies were detected with chemiluminescent substrates (Pierce, Rockford, IL) and visualized on x-ray film.

Linear Suc Gradients

Maize ears were harvested 23 DAP. All extractions and centrifugations were conducted at 4°C in buffer X (50 mM Tris-HCl, pH 8.0; 100 mM KCl; 30 mM

MgCl_2 ; 1 mM EGTA-NaOH; 1 mM EDTA). Mortar and pestle were used to carefully homogenize 6 g of endosperm tissue in 21 mL of buffer X containing 0.2 M Suc. The homogenate was filtered through two layers of cheesecloth (Veratec, Walpole, MA) and one layer of Miracloth (Calbiochem, La Jolla, CA). Equal amounts of the filtrate were layered over two 3-mL pads of buffer X containing 2 M Suc. After centrifugation at 160g for 10 min, 1 mL of the 0.2 M Suc supernatant was aspirated and immediately loaded onto Suc density gradients.

Linear Suc density gradients (10%–60%) in buffer X were prepared in SW40 ultracentrifuge tubes (Beckman Coulter Instruments, Fullerton, CA) using the BIOCAMP Gradient Maker 107ip (BioComp Instruments, New Brunswick, Canada) per the manufacturer's instructions. One milliliter of tissue homogenates was applied to the top of the prepared and chilled gradients. Organelles were then fractionated by centrifugation at 34,000 rpm (SW40 rotor) for 3 h. Following centrifugation, gradients were fractionated using a BIOCAMP Piston Gradient Fractionator-151 (BioComp Instruments) at 0.2 mm s^{-1} and collected using a Frac-200 fraction collector (Pharmacia LKB, Uppsala) set to collect 12 drops (approximately 300 μL) per fraction. The Suc density in every other gradient fraction was determined using 20 μL on a Milton Roy Abbe-3 L refractometer (Milton Roy, Rochester, NY) to ensure gradient quality.

For SDS-PAGE analysis and immunoblotting, 8 μL of sample was incubated at 100°C for 5 min with 2 μL of SDS-PAGE loading buffer (250 mM Tris, pH 6.8; 500 mM dithiothreitol; 2% [w/v] SDS; 0.5% [w/v] bromophenol blue; 50% [v/v] glycerol). Samples were then subjected to electrophoresis through 26-well 4% to 20% gradient Tris-HCl mini-gels (Bio-Rad) at 150V following a technical step to suppress a band artifact as described in Yokato et al. (2000). Protein was visualized by silver staining (Bollag et al., 1991). Immunoblotting and antigen detection were carried out as described by Gruis et al. (2002) and as described above. Prestained molecular mass protein standards (Page Ruler; Fermentas, Hanover, MD) were used to determine the apparent molecular masses of proteins.

Sequence data from this article have been deposited with the EMBL/GenBank data libraries and are shown in bold in Table I. Accession numbers for these sequences are AY739306, AY739307, AY739296, AY739297, AY739298, AY739299, AY739300, AY739301, AY739302, AY739303, AY739304, AY739305, AY739308, AY739284, AY739285, AY739286, AY739287, AY739295, AY739288, AY739289, AY739290, AY739291, AY739292, AY739293, and AY739294.

ACKNOWLEDGMENTS

We thank T. Elthon, A. Esen, and G. Galili for providing antisera against the mitochondrial protein subunits, α -zein, and the wheat PDI, respectively, and F. Salamini for providing the B37 mutant maize lines. We further thank colleagues in the DuPont-Pioneer Bioinformatics and Analytical and Genomics Technologies departments for creating a comprehensive and searchable gene database of maize and for performing the sequence analysis of full-length cDNA clones of maize and rice PDILs. Special thanks are extended to J. Gillikin, D. Thomas, and S. Yans for their excellent technical support through the course of the work. We would also like to thank B. Wiegmann and members of the Xiang and Boston laboratories for helpful discussions.

Received November 12, 2004; returned for revision December 2, 2004; accepted December 2, 2004.

LITERATURE CITED

- Alanen HI, Williamson RA, Howard MJ, Lappi AK, Jantti HP, Rautio SM, Kellokumpu S, Ruddock LW (2003) Functional characterization of ERp18, a new endoplasmic reticulum-located thioredoxin superfamily member. *J Biol Chem* **278**: 28912–28920
- Altschul SE, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402
- Balmer Y, Buchanan BB (2002) Yet another plant thioredoxin. *Trends Plant Sci* **7**: 191–193
- Benghezal M, Wasteneys GO, Jones DA (2000) The C-terminal dilysine motif confers endoplasmic reticulum localization to type I membrane proteins in plants. *Plant Cell* **12**: 1179–1201

- Bollag DM, Rozycki MD, Edelstein SJ** (1991) Protein Methods. Wiley-Liss, New York
- Boston RS, Fontes EB, Shank BB, Wrobel RL** (1991) Increased expression of the maize immunoglobulin binding protein homolog b-70 in three zein regulatory mutants. *Plant Cell* **3**: 497–505
- Boston RS, Viitanen PV, Vierling E** (1996) Molecular chaperones and protein folding in plants. *Plant Mol Biol* **32**: 191–222
- Brehelin C, Meyer EH, de Douris J-P, Bonnard G, Meyer Y** (2003) Resemblance and dissemblance of Arabidopsis type II peroxiredoxins: similar sequences for divergent gene expression, protein localization, and activity. *Plant Physiol* **132**: 2045–2057
- Brenner S, Johnson M, Bridgman J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M, et al** (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**: 630–634
- Coleman CE, Clore AM, Ranch JP, Higgins R, Lopes MA, Larkins BA** (1997) Expression of a mutant alpha-zein creates the floury2 phenotype in transgenic maize. *Proc Natl Acad Sci USA* **94**: 7094–7097
- Coleman CE, Lopes MA, Gillikin JW, Boston RS, Larkins BA** (1995) A defective signal peptide in the maize high-lysine mutant *floury-2*. *Proc Natl Acad Sci USA* **92**: 6828–6831
- Coppock DL, Cina-Poppe D, Gilleran S** (1998) The quiescin Q6 gene (QSCN6) is a fusion of two ancient gene families: thioredoxin and ERV1. *Genomics* **54**: 460–468
- Coughlan SJ, Hastings C, Winfrey RJ** (1996) Molecular characterization of plant endoplasmic reticulum: identification of protein disulfide-isomerase as the major reticuloplasmin. *Eur J Biochem* **235**: 215–224
- Ellgaard L** (2004) Catalysis of disulphide bond formation in the endoplasmic reticulum. *Biochem Soc Trans* **32**: 663–667
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G** (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**: 1005–1016
- Esen A** (1988) Immunological cross-reactivity among alpha-zeins of maize (*Zea mays* L.). *J Cereal Sci* **8**: 93–109
- Felsenstein J** (2002) PHYLIP (Phylogeny Inference Package). Department of Genome Sciences, University of Washington, Seattle
- Freedman RB, Hirst TR, Tuite MF** (1994) Protein disulphide isomerase: building bridges in protein folding. *Trends Biochem Sci* **19**: 331–336
- Gruis DE, Selinger DA, Curran JM, Jung R** (2002) Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. *Plant Cell* **14**: 2863–2882
- Gutierrez-Marcos JF, Roberts MA, Campbell EI, Wray JL** (1996) Three members of a novel small gene-family from *Arabidopsis thaliana* able to complement functionally an *Escherichia coli* mutant defective in PAPS reductase activity encode proteins with a thioredoxin-like domain and “APS reductase” activity. *Proc Natl Acad Sci USA* **93**: 13377–13382
- Hanke GT, Kimata-Arigo Y, Taniguchi I, Hase T** (2004) A post genomic characterization of Arabidopsis ferredoxins. *Plant Physiol* **134**: 255–264
- Hofmann K, Stoffel W** (1993) TMbase: a database of membrane spanning proteins segments. *Biol Chem Hoppe Seyler* **374**: 166
- Hoover KL, Sheasley SL, Gilbert HF, Thorpe C** (1999) Sulfhydryl oxidase from egg white: a facile catalyst for disulfide bond formation in proteins and peptides. *J Biol Chem* **274**: 22147–22150
- Huelsenbeck JP, Ronquist F** (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755
- Jacquot JP, Gelhaye E, Rouhier N, Corbier C, Didierjean C, Aubry A** (2002) Thioredoxins and related proteins in photosynthetic organisms: molecular basis for thiol dependent regulation. *Biochem Pharmacol* **64**: 1065–1069
- Jiang XM, Fitzgerald M, Grant CM, Hogg PJ** (1999) Redox control of exofacial protein thiols/disulfides by protein disulfide isomerase. *J Biol Chem* **274**: 2416–2423
- Jones DT, Taylor WR, Thornton JM** (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**: 275–282
- Kanai S, Toh H, Hayano T, Kikuchi M** (1998) Molecular evolution of the domain structures of protein disulfide isomerases. *J Mol Evol* **47**: 200–210
- Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE** (1997) The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Curr Biol* **7**: 239–245
- Kim CS, Hunter BG, Kraft J, Boston RS, Yans S, Jung R, Larkins BA** (2004) A defective signal peptide in a 19-kD alpha-zein protein causes the unfolded protein response and an opaque endosperm phenotype in the maize *De⁻-B30* mutant. *Plant Physiol* **134**: 380–387
- Kim JM, Mayfield SP** (1997) Protein disulfide isomerase as a regulator of chloroplast translational activation. *Science* **278**: 1954–1957
- Knodler LA, Noiva R, Mehta K, McCaffery JM, Aley SB, Svard SG, Nystul TG, Reiner DS, Silberman JD, Gillin FD** (1999) Novel protein-disulfide isomerases from the early-diverging protist *Giardia lamblia*. *J Biol Chem* **274**: 29805–29811
- Kopriva S, Koprivova A** (2004) Plant adenosine 5'-phosphosulphate reductase: the past, the present, and the future. *J Exp Bot* **55**: 1775–1783
- Laemmli UK** (1970) Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* **227**: 680–685
- Lange H, Lisowsky T, Gerber J, Muhlenhoff U, Kispal G, Lill R** (2001) An essential function of the mitochondrial sulfhydryl oxidase Erv1p/ALR in the maturation of cytosolic Fe/S proteins. *EMBO J* **2**: 715–720
- Li CP, Larkins BA** (1996) Expression of protein disulfide isomerase is elevated in the endosperm of the maize *floury-2* mutant. *Plant Mol Biol* **30**: 873–882
- Luethy MH, Horak A, Elthon TE** (1993) Monoclonal antibodies to the [alpha]- and [beta]-subunits of the plant mitochondrial F1-ATPase. *Plant Physiol* **101**: 931–937
- Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH** (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* **30**: 281–283
- Matsuo Y, Akiyama N, Nakamura H, Yodoi J, Noda M, Kizaka-Kondoh S** (2001) Identification of a novel thioredoxin-related transmembrane protein. *J Biol Chem* **276**: 10032–10038
- McArthur AG, Knodler LA, Silberman JD, Davids BJ, Gillin FD, Sogin ML** (2001) The evolutionary origins of eukaryotic protein disulfide isomerase domains: new evidence from the amitochondriate protist *Giardia lamblia*. *Mol Biol Evol* **18**: 1455–1463
- Meiri E, Levitan A, Guo F, Christopher DA, Schaefer D, Zryd JP, Danon A** (2002) Characterization of three PDI-like genes in *Physcomitrella patens* and construction of knock-out mutants. *Mol Genet Genomics* **267**: 231–240
- Meyer Y, Verdoucq L, Vignols F** (1999) Plant thioredoxins and glutaredoxins: identity and putative roles. *Trends Plant Sci* **4**: 388–394
- Meyers BC, Lee DK, Vu TH, Tej SS, Edberg SB, Matvienko M, Tindell LD** (2004) Arabidopsis MPSS: an online resource for quantitative expression analysis. *Plant Physiol* **135**: 801–813
- Monnat J, Neuhaus E, Pop MS, Ferrari DM, Kramer B, Soldati T** (2000) Identification of a novel saturable endoplasmic reticulum localization mechanism mediated by the C-terminus of a Dictyostelium protein disulfide isomerase. *Mol Biol Cell* **11**: 3469–3484
- Nakai K, Kanehisa M** (1991) Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* **11**: 95–110
- Norgaard P, Westphal V, Tachibana C, Alsos L, Holst B, Winther JR** (2001) Functional differences in yeast protein disulfide isomerases. *J Cell Biol* **152**: 553–562
- Pagny S, Cabanes-Macheteau M, Gillikin JW, Leborgne-Castel N, Lerouge P, Boston RS, Faye L, Gomord V** (2000) Protein recycling from the Golgi apparatus to the endoplasmic reticulum in plants and its minor contribution to calreticulin retention. *Plant Cell* **12**: 739–756
- Pelham HR** (1990) The retention signal for soluble proteins of the endoplasmic reticulum. *Trends Biochem Sci* **15**: 483–486
- Posada D, Crandall KA** (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818
- Prior A, Uhrig JE, Heins L, Wiesmann A, Lillig CH, Stoltze C, Soll J, Schwenn JD** (1999) Structural and kinetic properties of adenyllyl sulfate reductase from *Catharanthus roseus* cell cultures. *Biochim Biophys Acta* **1430**: 25–38
- Rodríguez E, Oliver JL, Marin A, Medina JR** (1990) The general stochastic model of nucleotide substitution. *J Theor Biol* **142**: 485–501
- Rouhier N, Gelhaye E, Jacquot J-P** (2004) Plant glutaredoxins: still mysterious reducing system. *Cell Mol Life Sci* **61**: 1266–1277
- Russell SJ, Ruddock LW, Salo KEH, Oliver JD, Roebuck QP, Llewellyn DH, Roderick HL, Koivunen P, Myllyharju J, High S** (2004) The primary substrate binding site in the b' domain of ERp57 is adapted for endoplasmic reticulum lectin association. *J Biol Chem* **279**: 18861–18869
- Sahrawy M, Hecht V, Lopez-Jaramillo J, Chueca A, Chartier Y, Meyer Y**

- (1996) Intron position as an evolutionary marker of thioredoxins and thioredoxin domains. *J Mol Evol* **42**: 422–431
- Schoof H, Zaccaria P, Gundlach H, Lemcke K, Rudd S, Kolesov G, Arnold R, Mewes HW, Mayer KF** (2002) MIPS *Arabidopsis thaliana* Database (MATDB): an integrated biological knowledge resource based on the first complete plant genome. *Nucleic Acids Res* **30**: 91–93
- Setya A, Murillo M, Leustek T** (1996) Sulfate reduction in higher plants: molecular evidence for a novel 5'-adenylylsulfate reductase. *Proc Natl Acad Sci USA* **93**: 13383–13388
- Shank KJ, Su P, Brglez I, Boss WF, Dewey RE, Boston RS** (2001) Induction of lipid metabolic enzymes during the endoplasmic reticulum stress response in plants. *Plant Physiol* **126**: 267–277
- Shewry PR, Napier JA, Tatham AS** (1995) Seed storage proteins: structures and biosynthesis. *Plant Cell* **7**: 945–956
- Shimoni Y, Zhu XZ, Levanony H, Segal G, Galili G** (1995) Purification, characterization, and intracellular-localization of glycosylated protein disulfide-isomerase from wheat grains. *Plant Physiol* **108**: 327–335
- Sonnhammer EL, von Heijne G, Krogh A** (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175–182
- Swofford DL** (2002) PAUP: Phylogenetic Analysis Using Parsimony, Version 4.0b10. Sinaur Associates, Sunderland, MA
- Takubo K, Morikawa T, Nonaka Y, Mizutani M, Takenaka S, Takabe K, Takahashi M, Ohta D** (2003) Identification and molecular characterization of mitochondrial ferredoxins and ferredoxin reductase from *Arabidopsis*. *Plant Mol Biol* **52**: 817–830
- The Arabidopsis Genome Initiative** (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG** (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882
- Thorpe C, Hooper KL, Raje S, Glynn NM, Burnside J, Turi GK, Coppock DL** (2002) Sulfhydryl oxidases: emerging catalysts of protein disulfide bond formation in eukaryotes. *Arch Biochem Biophys* **405**: 1–12
- Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P** (2000) Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* **101**: 249–258
- Turano C, Coppari S, Altieri F, Ferraro A** (2002) Proteins of the PDI family: unpredicted non-ER locations and functions. *J Cell Physiol* **193**: 154–163
- Vuori K, Myllylä R, Pihlajaniemi T, Kivirikko KI** (1992) Expression and site-directed mutagenesis of human protein disulfide isomerase in *Escherichia coli*: this multifunctional polypeptide has 2 independently acting catalytic sites for the isomerase activity. *J Biol Chem* **267**: 7211–7214
- Wang Q, Chang A** (1999) Eps1, a novel PDI-related protein involved in ER quality control in yeast. *EMBO J* **18**: 5972–5982
- Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, et al** (2002) Gramene: a resource for comparative grass genomics. *Nucleic Acids Res* **30**: 103–105
- Wendel J** (2000) Genome evolution in polyploids. *Plant Mol Biol* **42**: 225–249
- Wilkinson B, Gilbert HF** (2004) Protein disulfide isomerase. *Biochim Biophys Acta* **1699**: 35–44
- Woo YM, Hu DW, Larkins BA, Jung R** (2001) Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *Plant Cell* **13**: 2297–2317
- Wray JL, Campbell EI, Roberts MA, Gutierrez-Marcos JF** (1998) Redefining reductive sulfate assimilation in higher plants: a role for APS reductase, a new member of the thioredoxin superfamily? *Chem Biol Interact* **109**: 153–167
- Wrobel R** (1996) Expression of molecular chaperones in endoplasmic reticulum of maize endosperm. PhD thesis. North Carolina State University, Raleigh, NC
- Yokato H, Mori K, Kaniwa H, Shibamura T** (2000) Elimination of artifactual bands from polyacrylamide gels. *Anal Biochem* **208**: 188–189
- Zai A, Rudd MA, Scribner AW, Loscalzo J** (1999) Cell-surface protein disulfide isomerase catalyzes transnitrosation and regulates intracellular transfer of nitric oxide. *J Clin Invest* **103**: 393–399