Leucine-rich repeat (LRR) proteins feature tandem leucine-rich motifs that form a protein–protein interaction domain. Plants contain diverse classes of LRR proteins, many of which take part in signal transduction. We have identified a novel family of nine Arabidopsis LRR proteins that, based on predicted intracellular location and LRR motif consensus sequence, are related to Ras-binding LRR proteins found in signaling complexes in animals and yeast. This new class has been named plant intracellular Ras group-related LRR proteins (PIRLs). We have characterized PIRL cDNAs, rigorously defined gene and protein annotations, investigated gene family evolution and surveyed mRNA expression. While LRR regions suggested a relationship to Ras group LRR proteins, outside of their LRR domains PIRLs differed from Ras group proteins, exhibiting N- and C-terminal regions containing low complexity stretches and clusters of charged amino acids. PIRL genes grouped into three subfamilies based on sequence relationships and gene structures. Related gene pairs and dispersed chromosomal locations suggested family expansion by ancestral genomic or segmental duplications. Expression surveys revealed that all PIRL mRNAs are actively transcribed, with three expressed differentially in leaves, roots or flowers. These results define PIRLs as a distinct, plant-specific class of intracellular LRR proteins that probably mediate protein interactions, possibly in the context of signal transduction. T-DNA knock-out mutants have been isolated as a starting point for systematic functional analysis of this intriguing family.

Keywords: Arabidopsis thaliana — G-proteins — Gene knockouts — Leucine-rich repeats — Linker protein — Signal transduction.

Introduction

Leucine-rich repeat (LRR) proteins constitute a large and widespread protein superfamily found in animals, yeast, protists and bacteria as well as plants (Kobe and Deisenhofer 1994, Kaijawa 1998). Their defining feature is an LRR domain composed of tandemly repeated leucine-rich motifs, each 18–29 amino acids in length. While found on diverse proteins, LRR domains serve a common purpose: they form platforms that mediate specific protein–protein interactions (Kobe and Deisenhofer 1995, Kobe and Kaijawa 2001). Specific recognition and binding activities have been established experimentally for LRR domains in numerous proteins (Kobe and Deisenhofer 1995, Haberland and Gerke 1999, Leckie et al. 1999, Ellis et al. 2000, Howitt et al. 2004). Binding specificity is determined in part by the sequence characteristics of the leucine-rich unit motifs that make up the LRR domain; therefore, the leucine-rich motif sequence is a key determinant of LRR protein function.

In plants, LRR proteins carry out diverse functions in signal transduction (Vernon and Forsthoefel 2002, and references therein). For example, LRR-receptor-like kinases (RLKs), and related receptor-like proteins (RLPs), constitute a large family of plasma membrane proteins with extracellular N-terminal LRR domains. They take part in numerous developmental, environmental and defense-related pathways (Lease et al. 1998, Jeong et al. 1999, Shiu and Bleecker 2001, Morris and Walker 2003, Dievart and Clark 2004). Two distinct classes of LRR proteins mediate plant pathogen resistance: intracellular nucleotide binding site (NBS)-LRR proteins and Cf-type receptors with extracellular LRR domains (Jones et al. 1994, Mindrinos et al. 1994, Dangle 1995, Ellis et al. 2000, McDowell and Dangle 2000, Flurh 2001). Another example of plant LRR proteins involved in signaling are F-box/LRR proteins, which mediate targeted protein degradation in response to developmental and hormonal cues (Ruegger et al. 1998, Hellmann and Estelle 2002, Kuroda et al. 2002). In all of these major plant families, LRR domains have been shown to be critical for protein function, either for ligand recognition or for docking with other pathway components. Other plant LRR proteins involved in such diverse processes as pollen tube growth, root development, Rsp GTPase activation, transcription regulation and meristem cell organization have also been identified (Rose and Meier 2001, Stratford et al. 2001, Baumberger et al. 2001, Acevedo et al. 2004, Guyon et al. 2004, Suzuki et al. 2004).
Thus, plant LRR proteins are important in a variety of developmental and physiological contexts.

LRR proteins function in cell signaling in animals, yeast and protists, as well as plants. For example, other eukaryotes contain F-box/LRR proteins. In addition, receptors with extracellular LRR domains, such as TOLL-like and SLIT family receptors, are found in animals ranging from Drosophila to humans (Schuster and Nelson 2000, Aruga et al. 2003, Howitt et al. 2004). Another class of LRR proteins found in animals and lower eukaryotes are the Ras group LRR proteins (Claudianos and Campbell 1995). Ras-group LRR proteins are intracellular (or have intracellular LRR domains) and take part in signal transduction; several are known to interact directly with Ras GTPases. One, Caenorhabditis elegans SUR8, was identified in developmental Ras mutant suppressor screens, and directly interacts with Ras and Raf in a signaling complex (Sieburth et al. 1998, Li et al. 2000). Another, mouse/human RSP1/RSU1, is a tumor suppressor that can counter the effects of oncogenic Ras mutations in mammalian cell cultures (Cutler et al. 1992, Tsuda and Cutler 1993). Yeast adenylate cyclase is regulated by Ras through a cytoplasmic Ras-binding C-terminal LRR domain (Suzuki et al. 1990). Other examples of Ras group LRR proteins include Drosophila FLI, which binds Ras and regulates cytoskeletal organization in response to developmental signals, and Dictyostelium PATS1, which is involved in signaling during cytokinesis (Claudianos and Campbell 1995, Liu and Yin 1998, Abyssal et al. 2003).

Here, we describe a novel class of plant LRR proteins more closely related to Ras group LRR proteins than to any previously identified plant proteins. We have characterized sequence features, isolated and characterized cDNAs, rigorously defined gene structures, surveyed mRNA expression and determined evolutionary relationships for this new family. Based on their structural characteristics, we propose that these proteins interact with other cellular components via their LRR domains, potentially functioning in intracellular signal transduction. This hypothesis provides a testable framework for detailed elucidation of plant intracellular Ras-group-related LRR protein (PIRL) function. We also report identification of T-DNA knock-out mutants that will provide a starting point for defining the organismal contexts of PIRL activity.

## Results

### PIRLS: a novel class of intracellular plant LRRs

The completed Arabidopsis genome provides a valuable resource for identification of novel genes with prospective roles in signal transduction. During analysis of a region of Arabidopsis chromosome 5 (Tax and Vernon 2001), we identified a gene encoding a novel LRR protein.

Subsequent searches of the Arabidopsis genome identified eight more genes encoding related LRR proteins in this model plant system (Table 1). Their products were clearly related by a conserved central LRR domain (Suzuki et al. 1990). Other examples of Ras group LRR proteins include Drosophila FLI, which binds Ras and regulates cytoskeletal organization in response to developmental signals, and Dictyostelium PATS1, which is involved in signaling during cytokinesis (Claudianos and Campbell 1995, Liu and Yin 1998, Abyssal et al. 2003).

Here, we describe a novel class of plant LRR proteins more closely related to Ras group LRR proteins than to any previously identified plant proteins. We have characterized sequence features, isolated and characterized cDNAs, rigorously defined gene structures, surveyed mRNA expression and determined evolutionary relationships for this new family. Based on their structural characteristics, we propose that these proteins interact with other cellular components via their LRR domains, potentially functioning in intracellular signal transduction. This hypothesis provides a testable framework for detailed elucidation of plant intracellular Ras-group-related LRR protein (PIRL) function. We also report identification of T-DNA knock-out mutants that will provide a starting point for defining the organismal contexts of PIRL activity.

#### Table 1  PIRL locus and protein information

<table>
<thead>
<tr>
<th>PIRL AGI locus</th>
<th>cDNA accession no.</th>
<th>Protein Length a</th>
<th>LRR b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 At5g05850</td>
<td>AY849571</td>
<td>506</td>
<td>9</td>
</tr>
<tr>
<td>2 At3g26500</td>
<td>AY849572</td>
<td>470</td>
<td>9</td>
</tr>
<tr>
<td>3 At1g12970</td>
<td>AY849573</td>
<td>463</td>
<td>9</td>
</tr>
<tr>
<td>4 At4g35470</td>
<td>AY849574</td>
<td>549</td>
<td>11</td>
</tr>
<tr>
<td>5 At2g17440</td>
<td>AY849575</td>
<td>526</td>
<td>11</td>
</tr>
<tr>
<td>6 At2g19330</td>
<td>AY849576</td>
<td>380</td>
<td>9</td>
</tr>
<tr>
<td>7 At4g39880</td>
<td>AY849577</td>
<td>374</td>
<td>9</td>
</tr>
<tr>
<td>8 At4g26050</td>
<td>AY849578</td>
<td>382</td>
<td>9</td>
</tr>
<tr>
<td>9 At3g11330</td>
<td>AY849579</td>
<td>537</td>
<td>9</td>
</tr>
</tbody>
</table>

a Protein length in number of amino acids.

b Number of full-length leucine-rich motif units in the LRR domain.

AGI, Arabidopsis Genome Initiative.

#### Table 2  BLASTP alignments between PIRLs and mammalian Ras group LRR SUR8 (SOC2)

<table>
<thead>
<tr>
<th>Amino acid range of alignment a</th>
<th>e-value (byte score)</th>
<th>% Identity (% similarity) b</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIRL1</td>
<td>198–434</td>
<td>2e –26 (122)</td>
</tr>
<tr>
<td>PIRL2</td>
<td>146–380</td>
<td>6e –29 (131)</td>
</tr>
<tr>
<td>PIRL3</td>
<td>150–390</td>
<td>1e –30 (130)</td>
</tr>
<tr>
<td>PIRL4</td>
<td>197–498</td>
<td>2e –37 (159)</td>
</tr>
<tr>
<td>PIRL5</td>
<td>188–498</td>
<td>3e –26 (122)</td>
</tr>
<tr>
<td>PIRL6</td>
<td>83–304</td>
<td>3e –20 (101)</td>
</tr>
<tr>
<td>PIRL7</td>
<td>80–304</td>
<td>1e –16 (88)</td>
</tr>
<tr>
<td>PIRL8</td>
<td>85–283</td>
<td>8e –18 (94)</td>
</tr>
<tr>
<td>PIRL9</td>
<td>202–418</td>
<td>1e –29 (133)</td>
</tr>
</tbody>
</table>

a Amino acid region in the PIRL query sequence that exhibited BLASTP alignment, corresponding to the positions of the LRR domains, shown in Fig. 1.

b Percentage amino acid identity within LRR regions shown in column 2. Similarity with conservative amino acid changes is indicated in parentheses.
leucine-rich unit motifs, typically 23 amino acids in length. N- and C-terminal regions varied in length and sequence but shared some general features. They were hydrophilic and relatively rich in charged amino acids. Low complexity stretches, such as clusters of serines and/or serine and proline, were present in N-terminal regions, and short E/D-rich or K-rich
charged clusters were present in both N- and C-terminal domains (Fig. 1). All lacked prospective hydrophobic N-terminal signal sequences and candidate transmembrane or anchor domains. This combination of features defined this family as a distinct class of plant LRR proteins. Based on the LRR sequence comparisons described below, we named these proteins PIRLs. Fig. 1 shows a multiple alignment of the PIRLs, illustrating their domain organization, LRR domain conservation and other notable sequence features.

We carried out BLASTP database searches to investigate the relationship between PIRLs and other LRR proteins. Preliminary searches with the first identified protein, PIRL1, suggested a relationship to LRR proteins from animals and fungi involved in signal transduction, including mammalian SUR8 (SOC2) and RSP1, and the cytoplasmic C-terminal domain of yeast adenylate cyclase (Suzuki et al. 1990, Cutler et al. 1992, Sieburth et al. 1998). These are intracellular LRR proteins that directly interact with the small GTPase Ras, and which have been categorized as members of a distinct class of ‘Ras group’ LRR proteins (Claudianos and Campbell 1995). BLASTP analysis with each of the PIRLs confirmed this: the most closely related proteins with known functions were animal LRR proteins involved in signal transduction, most consistently SUR8 (SOC2) and RSP1. Information on BLASTP alignments between PIRLs and a representative Ras group LRR protein, SUR8, are summarized in Table 2. In all cases, PIRL alignments with animal LRR proteins were limited strictly to the LRR domain, and they extended the length of the PIRL LRR region (~225–300 amino acids; see Fig. 1 for location of LRR domains within PIRL sequences). PIRL LRR domains were shorter than those of the detected animal LRR proteins, and thus limited the range of alignment. Amino acid identities within LRR domains ranged from 26 to 37%; similarities were higher, allowing for conservative amino acid changes (Table 2), reflecting significant similarity across kingdoms for these proteins. PIRLs also aligned to a similar degree with receptor-associated animal LRR proteins in the Scribble family (Scribble and Erbin) and/or with mammalian Densin, which interacts with calcium-dependent protein kinase II (Borg et al. 2000, Strack et al. 2000, Jaulin-Bastard et al. 2002), both of which also take part in signal transduction.

PIRL sequence features confirm a relationship to Ras group LRRs

The LRR domain was the dominant structural feature of the PIRLs and appeared to distinguish PIRLs from previously defined classes of plant LRR proteins and drive alignments to animal LRR proteins. BLASTP searches detected similarity between PIRLs and animal LRR proteins, but were not sufficient to demonstrate a close relationship. LRR domain lengths vary widely between proteins in this superfamily, and relationships between LRR proteins are better reflected by unit motif sequences, rather than large-scale domain alignments. To investigate this further, we determined the PIRL LRR consensus motif. All of the individual leucine-rich motifs from the nine PIRLs were identified, compiled and compared. The 23 amino acid PIRL consensus motif is shown in Figure 2. We then compared this with the consensus motifs from other classes of LRR proteins. Several features of the PIRL consensus resembled features characteristic of the Ras group, including LP and IG amino acid combinations, and most prominently a distinct L\textsubscript{2}V\textsubscript{2}S sequence preceding the highly conserved N residue found at position 19 (Fig. 2).

Another intriguing resemblance to RAS group LRR proteins was the presence of a GxxxVxxYxxxxW (‘GVYW’) motif immediately following the LRR domain. Caenorhabditis elegans Ras-binding LRR SUR8, mouse RSP1 and human RSU1 all contain this sequence or derivations of it in precisely the same position, with the initial G located exactly 13 amino acids beyond the highly conserved N at position 19 of their final LRR motif. Other classes of LRR proteins do not contain this motif. This motif, or degenerate versions of it (GVFW or GGY), was present at precisely this location in PIRLs 1, 2, 4, 5 and 9 (Fig. 1). More degenerate versions (VY or GLY) were present in PIRLs 6, 7 and 8. While the functional significance of the GVYW motif and its degenerate forms is not known, it at the very least may reflect an evolutionary relationship between PIRLs and the Ras group LRR proteins.

While resembling LRR proteins from diverse animal and fungal sources, PIRLs differed from most plant LRR classes with respect to the Ras consensus motif, the cellular location of the LRR domain, or both (Fig. 2). Some plant motifs shared individual characteristics of the PIRLs, such as the serine at position 20 in the polygalacturonase inhibitor protein (PGIP) consensus (Fig. 2). However, none contained the combination of consensus features and predicted intracellular location shared by the PIRLs and Ras group LRR proteins.

Importantly, there were structural differences that distinguished the PIRLs from Ras group LRR proteins and set them apart as a novel, plant-specific family. PIRLs had considerably shorter LRR domains than some Ras group proteins, in which LRR domain length varies. For example, SUR8 (with 18) and adenylate cyclase (with 21) have considerably more LRR units.
PIRLs, a novel class of intracellular LRR proteins

PIRLs are a novel class of intracellular LRR proteins, feature more extensive N-terminal regions than PIRLs (Sieburth et al. 1998, Suzuki et al. 1990). PIRLs also featured more extensive N-terminal regions than some Ras group LRR proteins. SUR8 and RSP1, for example, lack extensive N-terminal domains, and consist almost entirely of leucine-rich repeats. Thus, PIRLs overall do not appear to be highly related to Ras group proteins in all respects.

Arabidopsis cDNA sequences define PIRL gene boundaries and exon/intron structure

Arabidopsis genome annotations and corresponding protein sequence predictions originally suggested the presence of short, unconserved ‘non-leucine-rich islands’ within the LRR domains of the nine PIRL proteins and compared with published consensus motifs for representatives of the indicated LRR classes (Cutler et al. 1992, Jones et al. 1994, Mindrinos et al. 1994, Li and Chory 1997). Consensus sequences are shown aligned, with amino acid position numbers indicated at the top and bottom. Animal and yeast Ras group LRR motifs are highlighted with a gray background. Upper case letters, amino acids conserved in that position in >50% of motifs; staggered upper case letters, >50% of motifs contain one of the indicated amino acids; lower case ‘a’, a hydrophobic amino acid is present in >50% of motifs; other lower case letters, amino acids present in that position at a frequency of 25–49%. Dots mark positions of variable amino acids. Intracellular vs. extracellular designation refers to the location of the LRR domain in each class of protein.

cDNA sequences define PIRL gene boundaries and exon/intron structure

Arabidopsis cDNA sequences were then compared with previously predicted Arabidopsis gene structures to refine models of PIRL transcription units. Fig. 3 illustrates the resulting empirically determined PIRL intron/exon structures.

For most PIRL genes, annotation predictions were correct. However, in PIRL3 and PIRL7, we identified small, in-frame introns that had been overlooked by bioinformatic annotation. Their position corresponded to the location of the putative non-LRR islands that had been predicted in those proteins, thus indicating that LRR domains were in fact uninterrupted in these proteins. Consistent with this, introns were also present at corresponding positions in the most closely related PIRL genes (Fig. 3). In PIRL9, the cDNA sequence revealed the presence of a single large intron spanning a region that had been annotated as an exon in original genome annotation. GenBank accession numbers for cDNAs are provided in Table 1. The arrows mark sites of T-DNA inserts in knock-out alleles.

Arabidopsis cDNA resources. cDNAs were then compared to previously predicted Arabidopsis gene structures to refine models of PIRL transcription units. Fig. 3 illustrates the resulting empirically determined PIRL intron/exon structures.

For most PIRL genes, annotation predictions were correct. However, in PIRL3 and PIRL7, we identified small, in-frame introns that had been overlooked by bioinformatic annotation. Their position corresponded to the location of the putative non-LRR islands that had been predicted in those proteins, thus indicating that LRR domains were in fact uninterrupted in these proteins. Consistent with this, introns were also present at corresponding positions in the most closely related PIRL genes (Fig. 3). In PIRL9, the cDNA sequence revealed the presence of a single large intron spanning a region that had been annotated as an exon and two smaller introns. This established that PIRL9 intron/exon organization paralleled that of closely related PIRL1, and ruled out the presence of a non-LRR island in the PIRL9 gene product (Fig. 3). Thus, uninterrupted LRR domain structures are conserved throughout the PIRLs, suggesting conserved functions for this domain in the protein family.
PIRLs, a novel class of intracellular LRR proteins

PIRLs are divided into three subfamilies, with PIRLs 1, 2, 3 and 9 comprising one group, PIRLs 6, 7 and 8 another, and PIRLs 4 and 5 the third. These relationships were also reflected in gene exon/intron organization (Fig. 3), suggesting ancient divergence between the three subfamilies. Within these subfamilies, more closely related gene pairs, such as PIRL2 and PIRL3, with products sharing >60% amino acid identity, could be distinguished (Fig. 4).

Gene distribution can provide information on family evolution. Members of some large plant LRR families, for example, are found in large tandem gene clusters, indicating that they expanded in part via local gene duplication events (Ellis et al. 2000, Fluhr 2001, Shiu and Bleecker 2001, Shiu and Bleecker 2003). This was not the case for PIRL genes: PIRL loci were distributed throughout the genome, with none found in tandem clusters (Fig. 4B). PIRL5 and PIRL6 on chromosome 2 were the closest genes, and these were separated by 189 other loci and are not in the same subfamily. Thus, tandem duplications apparently did not contribute to PIRL gene expansion.

**PIRL transcripts are widely expressed in adult plants**

To confirm that PIRL loci are actively expressed, we qualitatively surveyed PIRL transcript expression in leaves, roots, flowers and developing seeds by RT–PCR, in conjunction with sequencing to confirm RT–PCR specificity and product identity. In general, PIRL RNAs were widely expressed, most being detected in all of the organs surveyed (Fig. 5). However, some were differentially expressed: PIRL7 transcripts were detected only in flowers, while RNA from PIRL8 was not detected in flowers, but was present in roots and leaves. Northern blots with equal amounts of RNA from leaves, roots and flowers revealed differential expression of PIRL2 and PIRL5, showing substantially higher expression of these transcripts in leaves and roots than in flowers. For other genes, RNA blots mirrored the results obtained with RT–PCR (data not shown). Thus, PIRL genes as a group were widely expressed during development, but PIRLs 2, 5, 7 and 8 exhibited organ-specific differences in RNA expression.

**Identification of T-DNA knock-out alleles for eight PIRL genes**

To lay the groundwork for a functional analysis of the PIRL genes, we have taken a reverse-genetic approach and identified pirl knock-out alleles. Using standardized PCR screening procedures (Krysan et al. 1999, Sussman et al. 2000), we screened T-DNA mutagenized genomic DNA pools from the Arabidopsis knock-out facility (University of Wisconsin). Prospective mutant lines containing insertion alleles of PIRL2, 3, 4, 5, 7, 8 and 9 were identified. T-DNA insert positions are located within PIRL transcription units and have been confirmed by PCR. T-DNA insert positions are shown in Fig. 3. We initially identified lines containing prospective promoter insertions for PIRL1, but no strong candidate knock-outs. However, as larger collections of sequence-indexed T-DNA insertion mutant collections became available (Alonso et al. 2003), we were able to identify a prospective pirl1 knockout allele. Thus, mutant alleles for all genes except PIRL6 have been identified. These lines will provide a foundation for in-depth phenotypic analyses to investigate PIRL functions at the organismal level, and for the construction of double or triple mutant lines that may be necessary to overcome potential genetic redundancy between closely related PIRL genes.

---

**Fig. 4** Evolutionary relationships in the PIRL gene family. (A) Phenogram generated from ClustalW alignment of amino acid sequences. Sequences were predicted from cDNAs. Closely related gene pairs are indicated by parentheses, with percentage amino acid identity provided on the right. Subfamilies sharing sequence similarity and gene exon/intron structures are indicated by larger brackets. (B) Chromosomal positions of PIRL loci, showing dispersed locations and lack of tandem gene clusters. Roman numerals indicate chromosome identities; Arabic numbers mark the positions of the respective PIRL genes. The closest loci, PIRL5 and PIRL6 on chromosome 2, are from different subfamilies, and are separated by 189 annotated genes.
PIRLs represent a novel and plant-specific class of LRR proteins

Previous studies have classified members of the LRR superfamily based on motif sequence (Claudianos and Campbell 1995, Kajava 1998). Those studies defined a ‘plant-specific’ consensus motif found on secreted anti-fungal PGIPs, transmembrane proteins with extracellular LRR domains, and on an intracellular LRR that interacts with the transcription factor AGAMOUS (Di Matteo et al. 2003, Acevedo et al. 2004, Guyon et al. 2004). Here, we have described the PIRLs, a distinct class of plant proteins more closely related to intracellular animal and fungal Ras-interacting LRR proteins than to previously defined plant LRR proteins. The PIRLs’ evolutionary relationship to Ras group proteins was evident in detailed comparisons of LRR consensus motifs, BLAST alignments and the identification of a GGYW-like motif at a precisely conserved position downstream of the final leucine-rich repeat motif.

PIRL gene family evolution

Aspects of PIRL gene family evolution can be surmised from sequence relationships and chromosomal positions. Sequences and exon/intron structures suggest PIRL family expansion through gene duplication. However, PIRL genes are not found in clusters, as are some members of larger plant LRR gene families that have expanded in part by tandem duplications, such as LRR-kinases and NBS-LRR resistance genes (Fluhr 2001, Shiu and Bleecker 2001, Shiu and Bleecker 2003). Instead, closely related related pairs of PIRL genes are in dispersed chromosomal locations. This suggests that the different PIRL subfamilies were represented in the ancestral genome and subsequently expanded by ancient genome or segmental duplications, either of which would be expected to give rise to chromosomally dispersed gene pairs. Both whole-genome and segmental duplications have occurred during Arabidopsis evolution and contributed to gene family expansion (Arabidopsis Genome Initiative 2000, Vision et al. 2000). The divergent intron sizes and limited amino acid identities (<80%) between related PIRLs are consistent with this scenario, suggesting that the family did not expand by recent localized gene duplications.

PIRL8 was the single ‘orphan’ PIRL gene without a duplicate counterpart within its subfamily. This can be explained either by loss of one gene after two rounds of ancient duplication in that subfamily (resulting in a subfamily of three related genes, rather than four), or by early duplication of the progeni-
tor of that family, followed by subsequent segmental duplication of just one member to create a third gene more closely related to its immediate progenitor. Multiple rounds of duplication, as well as gene loss following duplication, were both common events during Arabidopsis genome evolution (Vision et al. 2000, Simillion et al. 2002, Shiu and Bleecker 2003).

**PIRLs are broadly expressed in adult plants**

Using RT–PCR, followed by sequencing to confirm PCR product specificity, we have demonstrated that all nine PIRL genes are actively transcribed. mRNA expression for PIRL genes as a group was widespread in organs of adult plants and in seeds. However, we observed differential expression for some of the genes. PIRL2 and PIRL5 transcripts were qualitatively detectable by RT–PCR in all organs examined, but Northern blots demonstrated substantially higher expression in leaves and roots than in flowers for these two genes. The related PIRL7 and PIRL8 transcripts were also differentially regulated in these organs: PIRL7 was only detectable in flowers, while PIRL8 was detected solely in leaves and roots. PIRL6 RT–PCR yielded multiple products in leaf and root tissue. Sequencing of the most abundant and shortest product unambiguously identified it as the fully processed PIRL6 mRNA (Fig. 5). The nature of the higher molecular weight bands is not clear, but preliminary analysis suggest they may be incompletely or aberrantly spliced PIRL6 transcripts (N. Forsthoefel and D. Vernon, unpublished).

Further work will be needed to examine PIRL expression in more detail. However, such studies would best be carried out in conjunction with reverse genetics, such that phenotypes can provide clues to where to focus those efforts, using approaches such as in situ hybridization.

**A general hypothesis for PIRL function**

The sequence features of the PIRLs suggest a two-part hypothesis regarding their function. First, PIRLs are likely to take part in functionally significant protein–protein interactions, mediated by their LRR domains. Secondly, PIRLs may do so as components of intracellular signaling networks.

The first aspect of this hypothesis is based on the well-established role of LRR domains in mediating specific protein interactions (Kobe and Deisenhofer 1994, Kobe and Deisenhofer 1995, Kobe and Kajava 2001, Vernon and Forsthoefel 2002, and references therein). Low complexity regions in the PIRL N-terminal domains are also consistent with a role in protein binding, because such sequences can form non-globular surfaces favorable for protein–protein interactions. Thus, it is possible that PIRLs interact with other factors via their N-terminal domains, as well as with their LRR domains.

The second aspect of this hypothesis—a role for PIRLs in signal transduction—is speculative, but is suggested by the established roles of other LRR proteins in plant signaling, and by the relationship of PIRLs to RAS group LRR proteins and to Scribble and Densin, which all take part in developmental signaling in animal systems. Two of the best characterized animal Ras group LRR proteins, SUR8 and RSP-1, serve as ‘ adaptor proteins’, linking Ras into larger complexes with other signal transduction components (Sternberg and Alberola-Ila 1998, Li et al. 2000). It is possible that PIRLs could play roughly analogous roles in plants, bridging or coordinating multiple component signaling complexes (although probably not ones involving Ras; see below).

Alternatively, it is possible that PIRLs do not act directly in signal transduction: for example, they could have novel roles in intracellular transport, or in regulation of their binding targets. There is some precedent for some LRR proteins being involved in such processes. For example, RanGAPs and related LRR proteins, which have an overall domain organization similar to the PIRLs (but different LRR motif sequences), take part in nuclear envelope assembly, nucleo-cytoplasmic transport and possibly chromatin regulation (Rose and Meier 2001, Acker et al. 2002, Kusano et al. 2004). Also, FLOR1, another intracellular plant LRR protein, interacts with the transcription factor AGAMOUS (Acevedo et al. 2004). Thus, while we hypothesize functions in signal transduction based on PIRL structural observations, we cannot entirely rule out other important intracellular roles.

Elucidation of PIRL biochemical function will require identification of binding targets. LRR motif sequences are important determinants of LRR binding specificity, so it is tempting to speculate, based on PIRL’s similarity to RAS group LRR proteins, that they may interact with small GTPases. While plants apparently lack direct orthologs to Ras (Arabidopsis Genome Initiative 2000), they do contain small GTPases of the Ras superfamily, such as the plant-specific Rops, which have diverse roles in developmental signal transduction (Zheng and Yang 2000, Li et al. 2001). Thus, it is reasonable to propose an analogous binding role for PIRLs. However, while related to Ras group LRR proteins, PIRLs are somewhat divergent, and such functional parallels must be considered tentative. As a novel class of LRR proteins, PIRLs may interact with plant-specific targets distinct from those identified for any previously characterized LRR proteins.

Both components of this general hypothesis are testable and it sets up a framework for a systematic dissection of PIRL functions. An important first step is to define the biological contexts in which PIRLs operate. To this end, we have taken a reverse-genetic strategy and identified a set of T-DNA knockout mutants representing eight of the nine PIRL loci. Characterization of these mutants will involve in-depth phenotypic characterization, as well as construction of double or multiple mutant lines to circumvent potential functional redundancy between closely related PIRLs. Ultimately, these mutants will provide a starting point for in-depth dissection of PIRL functions.
Materials and Methods

Plant materials

Plants used for DNA and RNA isolation (ecotype WS) were grown in soil from seed obtained from the Arabidopsis Biological Resource Center (ABRC; Ohio State University). Growth conditions were as previously described (Tax and Vernon 2001). Seedlings from knock-out lines grown for T-DNA insert analysis were surface sterilized and germinated in culture as described by Vernon and Meinke (1995) in the presence of 50 μg ml⁻¹ kanamycin.

Gene identification and sequence analyses

A portion of PIRL1 was identified during investigation of a region of chromosome V, described by Tax and Vernon (2001). Full-length PIRL1 and additional PIRL sequences were identified by BLASTN searches (Altschul et al. 1997) at GenBank [http://www.ncbi.nlm.nih.gov] and TAIR [http://www.Arabidopsis.org]. Gene boundaries and exon/intron junctions were determined experimentally by comparison between annotated gene sequences and cDNA sequences obtained by RT–PCR (see below) or expressed sequence tag (EST) sequences identified with BLASTN searches.

All sequencing was done by the University of Arizona DNA sequencing core facility (Tucson, AZ, USA). Sequence alignments, open reading frame analysis and protein predictions were done with experimentally determined cDNA sequences, using MacVector 6.0 software (Accelrys, San Diego, CA, USA).

Protein comparisons were carried out using both computer alignments and manual analysis. BLASTP alignments (Altschul et al. 1997) were used to identify related proteins, using PIRL query sequences and default parameters at the NCBI site (http://www.ncbi.nlm.nih.gov/).

Repeat units and boundaries of PIRL LRR domains were identified by further manual analysis of PIRL sequences, to compare LRR motifs with those of previously characterized LRR proteins, and to identify potential partial or degenerate motifs. The PIRL LRR consensus sequence was determined manually. First, individual LRR unit motifs within each LRR domain were identified. Then, all 85 such motifs from the nine PIRLs were aligned using ClustalW, and subjected to further detailed analysis to determine amino acid frequency at specific positions. Positions conserved at a frequency of >49%, or between 25 and 49%, were identified and designated as consensus positions.

RNA expression and cDNA amplification

Plant RNA isolation, RT–PCR and genomic PCR were carried out as described by Cushing et al. (2005). Reactions were carried out for 35 cycles (94°C for 45 s, 58°C for 60 s, 72°C for 120 s). Products were isolated with QIAquick Gel Extraction reagents (Qiagen, Valencia, CA, USA) and sequenced to verify product identity and determine correct intron/exon boundaries. Sequencing of purified products demonstrated primer specificity.

For Northern blots, RNA isolation, probe generation and blotting were carried out as described in Chauhan et al. (2000), using the GeneTargeting and sequencing strategies dictated by that facility (Krysan et al. 1999, Sussman et al. 2000). T-DNA-specific primers were those dictated by the Knockout facility protocols: JL202 (left-border specific), CATTITATAATACGCTGCGGACATCTAC; and XR2 (right-border specific), TGGGAAAACCTGGCGTTACCCAACTTAAT. The PIRL1 allele was identified in the sequence-indexed Salk insertion mutant collection (Alonso et al. 2003; searched at http://www.arabidopsis.org) and seeds were obtained from the ABRC (Ohio State University). T-DNA insert locations were confirmed by PCR of T-DNA/gene junction fragments using genomic DNA preparation and PCR methods described by Tax and Vernon (2001) and Cushing et al. (2005), respectively.

Identification of knock-out mutants

Prospective knock-out alleles for PIRL genes 2, 3, 4, 5, 7, 8 and 9 were identified among pooled lines provided by the University of Wisconsin Knockout facility, using PCR screening and Southern blotting, and sequencing strategies dictated by that facility (Krysan et al. 1999, Sussman et al. 2000). T-DNA-specific primers were those dictated by the Knockout facility protocols: JL202 (left-border specific), CATTITATAATACGCTGCGGACATCTAC; and XR2 (right-border specific), TGGGAAAACCTGGCGTTACCCAACTTAAT. The PIRL1 allele was identified in the sequence-indexed Salk insertion mutant collection (Alonso et al. 2003; searched at http://www.arabidopsis.org) and seeds were obtained from the ABRC (Ohio State University). T-DNA insert locations were confirmed by PCR of T-DNA/gene junction fragments using genomic DNA preparation and PCR methods described by Tax and Vernon (2001) and Cushing et al. (2005), respectively.

Acknowledgments

We thank Dr. Heidi Sofia (Pacific Northwest National Laboratory) for valuable discussions on sequence comparisons, and Brady Hamilton, Marianne (Brady) Bryan and Jennifer Eastberg (Whitman College) for assistance. This work was initiated with award NSF-96-04344 and supported by USDA 02-35304-12304 and a grant from the M. J. Murdock Charitable Trust (all to D.M.V.).

References


