

# Identification and Characterization of the *ARIADNE* Gene Family in Arabidopsis. A Group of Putative E3 Ligases<sup>1</sup>

Christina Mladek, Klaus Guger, and Marie-Theres Hauser\*

Center of Applied Genetics, University of Agricultural Sciences Vienna, Austria

ARIADNE (ARI) proteins were recently identified in fruitfly (*Drosophila melanogaster*), mouse, and man because of their specific interaction with the ubiquitin-conjugating (E2) enzymes UbcD10, UbcM4, UbcH7, and UbcH8. They are characterized by specific motifs and protein structures that they share with PARKIN, and there is increasing evidence that ARI/PARKIN proteins function as E2-dependent ubiquitin-protein ligases. On the basis of homology and motif searches, 16 *AtARI* genes were identified in Arabidopsis. Analysis of the position of exons/introns and their chromosomal localization indicates that the *AtARI* gene family expanded via larger and smaller genome duplications. We present evidence that retroposition of processed mRNA may have also contributed to enlarging this gene family. Phylogenetic analyses divides the *AtARI* proteins into three subgroups. Two groups are absent in yeast, invertebrates, and vertebrates and may therefore represent new plant-specific subfamilies. Examination of the predicted protein sequences revealed that the ARI proteins share an additional leucine-rich region at the N terminus that is highly conserved in all phyla analyzed. Furthermore, conserved consensus signals for casein kinase II-dependent phosphorylation and for nuclear localization were identified. The in silico-based analyses were complemented with experimental data to quantify expression levels. Using real-time polymerase chain reaction, we show that the *ARI* genes are differentially transcribed. *AtARI1* is highly expressed in all organs, whereas no transcripts could be detected for *AtARI11*, *AtARI13*, and *AtARI14*. *AtARI12* and *AtARI16* are expressed in an organ-specific manner in the roots and siliques, respectively.

Now that the sequence of the Arabidopsis genome is completed, future challenges lie with the functional characterization of all genes in their cellular, developmental, and evolutionary context. Although extensive bioinformatic efforts at Munich Information Center for Protein Sequences (MIPS), The Arabidopsis Information Resource (TAIR), and The Institute for Genomic Research (TIGR) classified around 69% of the genes, many of them remain unknown or hypothetical (Arabidopsis Genome Initiative [AGI], 2000). One of the most frequently detected domains in the Arabidopsis proteome is the RING-finger, which is a Cys-rich region with a C-X<sub>2</sub>-C-X<sub>9-39</sub>-C-X<sub>1-3</sub>-H-X<sub>2-3</sub>-C/H-X<sub>2</sub>-C-X<sub>4-48</sub>-C-X<sub>2</sub>-C signature coordinating two Zn<sup>2+</sup> ligands (Borden and Freemont, 1996; AGI, 2000). Kosarev et al. (2002) recently reevaluated the RING-finger domains of the Arabidopsis proteome and identified through InterPro searches 387 domains that have the potential to form the typical RING-type cross-brace structure (Borden and Freemont, 1996).

RING-finger domains are regarded as protein interaction domains, involved in diverse cellular functions (Saurin et al., 1996). There is increasing evi-

dence that RING-finger proteins mediate transfer of ubiquitin to proteins targeted for proteolysis via the 26S proteasome pathway (Freemont, 2000; Jackson et al., 2000; Joazeiro and Weissman, 2000). In this pathway, ubiquitin is transferred to a target protein through a cascade of enzymes including ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2 or Ubc), and ubiquitin-protein ligases (E3). E3s, which have the highest diversity, interact directly with the target protein and are responsible for the specificity.

E3s can be divided into two classes, with either a HECT or RING-finger domain(s). Some RING-finger-containing E3s are part of E3 ubiquitin ligase complexes such as the SKP1, Cullin/CDC53, F-box protein, the anaphase-promoting complex, and the VHL-Elongin-C-Elongin-B (for review, see Jackson et al., 2000) others are single-subunit enzymes. PRT1, the first RING-finger protein suggested to be associated with ubiquitination of N-end rule substrates (Potuschak et al., 1998), belongs to the latter subclass of E3 ligases.

There is increasing evidence from recent genetic and biochemical studies that regulated protein degradation via the SKP1, Cullin/CDC53, F-box protein-type E3 ubiquitin ligase pathway is involved in many aspects of plant development. At present, this list comprises regulation of auxin (Gray et al., 2001) and jasmonate (Xie et al., 1998) signaling, flower morphogenesis (Samach et al., 1999), circadian rhythms (Nelson et al., 2000; Somers et al., 2000; Dieterle et al.,

---

<sup>1</sup> This work was supported by the Austrian Science Fund (project nos. P11001 and P14477 to M.T.H.).

\* Corresponding author; e-mail hauser@boku.ac.at; fax 43-1-36006-6392.

Article, publication date, and citation information can be found at [www.plantphysiol.org/cgi/doi/10.1104/pp.012781](http://www.plantphysiol.org/cgi/doi/10.1104/pp.012781).

2001), leaf senescence (Woo et al., 2001), R gene-mediated disease resistance (Austin et al., 2002; Azevedo et al., 2002), and embryogenesis (Shen et al., 2002). Proteins with RING-finger domains are classified based on the presence of additional protein motifs. The ARIADNE (ARI) class of RING-finger proteins are characterized by the presence of an N-terminal acid-rich cluster, followed by a C3HC4 RING-finger motif, a central in between RING-finger (IBR) or B-box, and a second C3HC4 RING-finger structure. At the C terminus, these proteins have a potential coiled-coil domain and a Leu-rich region. Because of their acid-RING-B-box-RING-coiled-coil, they are also designated as the ARBRCC subgroup of RING-finger or R-IBR-R proteins. ARI proteins have been identified by their interaction with the E2 UBCs UbcD10, UbcM4, UbcH7, and UbcH8 of fruitfly (*Drosophila melanogaster*), mouse, and man, respectively (Martinez-Noel et al., 1999; Moynihan et al., 1999; Aguilera et al., 2000; Ardley et al., 2001). ARI proteins share their R-IBR-R domain with PARKIN, a protein involved in autosomal recessive familial Parkinson's disease. PARKIN functions as E2-dependent ubiquitin-protein ligase and promotes the degradation of a synaptic vesicle-associated septin, CDCrel-1 (Zhang et al., 2000; Rankin et al., 2001). Disease-associated mutations within the RING1 of the R-IBR-R domain destroy E3 activity (Shimura et al., 2000). Thus, there is increasing evidence that the ARI/PARKIN proteins define a new class of single-polypeptide RING-finger E3 ligases.

In this study, we classify and characterize ARI proteins of Arabidopsis by reevaluating the predicted exon/intron distribution with computational and experimental analyses. We demonstrate that members of this multigene family are expressed in an organ-specific manner. Sequence comparison and phylogenetic analyses divided the AtARI proteins into three subgroups. Two groups are absent in yeast, invertebrates, and vertebrates including mammals and may therefore represent new subgroups specific to plants.

## RESULTS

### The Arabidopsis Genome Codes for 16 ARI Proteins

To identify the AtARI gene family of Arabidopsis, database homology searches were performed starting with the ARI proteins of fruitfly. The results were evaluated based on the presence of ARI-specific protein domains: the acidic N terminus, the Cys-rich domain of the RING1-IBR-RING2 structure (approximately 200 amino acids), and the Leu-rich domain at the C terminus. On the basis of these criteria, 16 AtARI genes were identified (Fig. 1; Table I). Seven genes are currently annotated as ARI-like proteins (AtARI1/9/10/11/13/14/16), one gene (AtARI6) as an ARI pseudogene, and the remainder as hypothetical, unclassified, putative RING zinc finger or unknown proteins. The recently annotated ARI-like

genes At5g37560 and At3g454800 were excluded from this analysis because they did not fit our criteria and because they lack the acidic N terminus and the C-terminal Leu-rich domain.

Although most of the characteristic domains are present in AtARI proteins, only AtARI4/5/7/8/9/10/11 have a complete RING1-IBR-RING2 domain. The IBR with the signature C-X<sub>4-5</sub>-C-X<sub>15-24</sub>-C-X-C-X<sub>4</sub>-C-X<sub>2</sub>-C-X<sub>4</sub>-H-X<sub>4</sub>-C (C6HC) is conserved in all AtARI but AtARI12 (Fig. 1). AtARI1/2/3 have a mutated RING2 with a Leu instead of the central His, whereas in AtARI12, the central C-X<sub>1-3</sub>-H residues of RING2 are missing. In addition to the truncated RING2, the AtARI14 and AtARI16 of the remaining AtARI13/14/15/16 contain a RING1 domain that has lost the first two Cys residues.

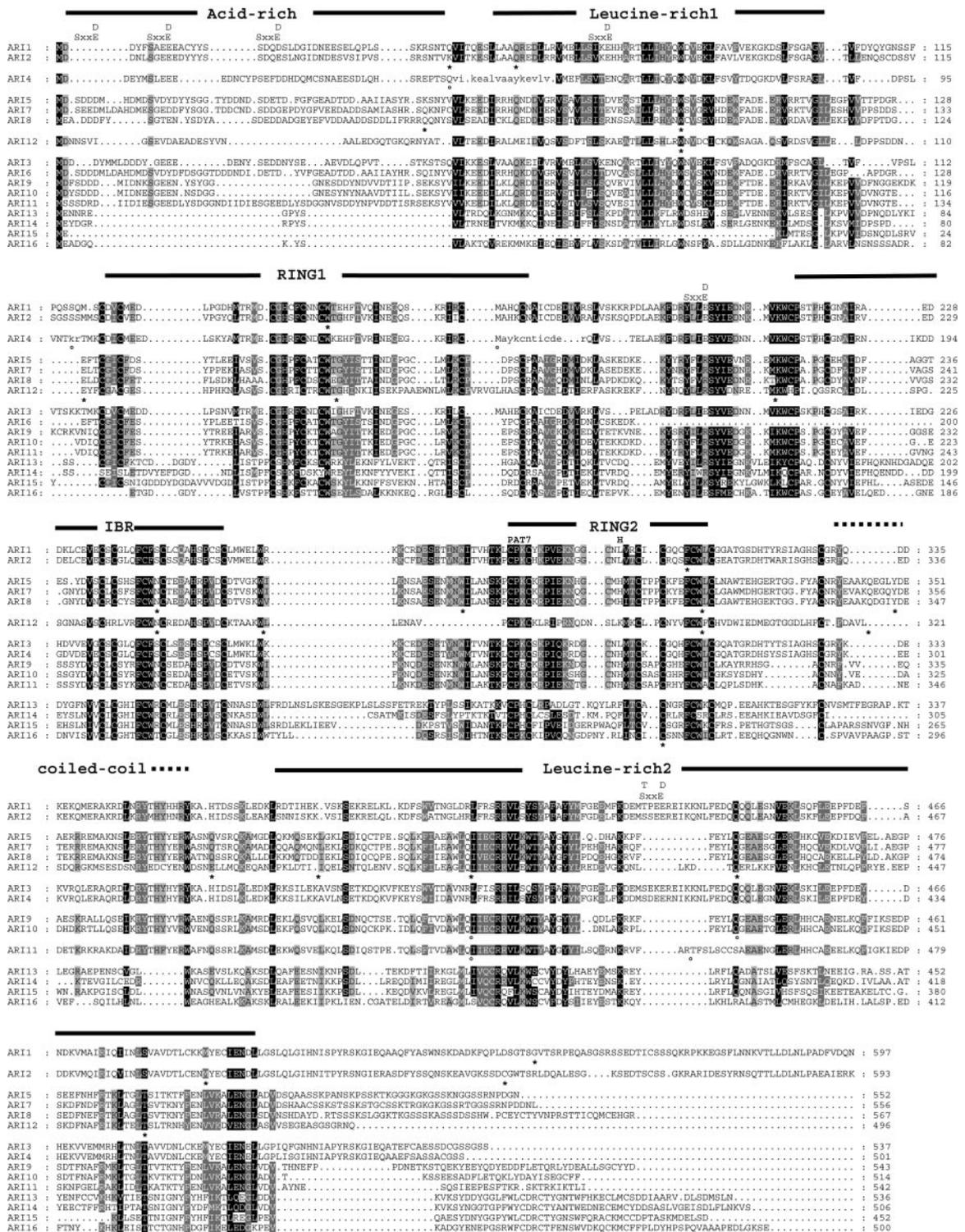
The sequence alignment further reveals highly conserved amino acids such as Val, Leu, Ile, Trp, and Gly between the acidic N terminus and RING1. The positions of Leu/Ile are conserved in ARI proteins of diverse organisms (see below), and this region is subsequently referred to as Leu-rich1 domain.

The acidic N terminus is rich in Ser residues, and several are positioned in consensus phosphorylation sites (S/T XX D/E) of casein kinase II (CKII; Pinna, 1990). In addition, two well-conserved CKII signatures are present in the Cys-rich region between RING1 and IBR domain (Fig. 1). Thus, ARI proteins might be targets for CKII-dependent phosphorylation, which may provide a mechanism for regulating their activity. Protein stability may be another mechanism by which ARI proteins are regulated. In fact, PEST signatures are predicted for AtARI1/2/3/11 in the acidic N terminus.

To uncover the possible subcellular localization of the AtARI proteins, we searched for different signatures specific to cellular compartments. Neither of the AtARI proteins carry a signal peptide or a transmembrane domain. Because of their basic amino acid composition, and as suggested by PSORT, most AtARI family members may be localized in the nucleus (Table I; Reinhardt and Hubbard, 1998). This conjecture is supported by AtARI5/7/8/10/11, which have a classical SV40-type seven-residue pattern (pat7) of P-X<sub>2</sub>-K-X-K-R nuclear localization sequence in the highly conserved region of RING2 (Horton and Nakai, 1997; Fig. 1). Nevertheless, a transient cytoplasmic localization cannot be ruled out.

### Phylogenetic Analysis Subdivides the ARI Protein Family

Visual inspection of the AtARI protein alignment already shows that AtARI proteins fall into three groups: A with AtARI1/2/3/4, B with AtARI5/6/7/8/9/10/11/12, and C with AtARI13/14/15/16. These groups exhibit less than 44% sequence and 33%



**Figure 1.** Protein alignment of the AtARI family. Black, dark gray, and light gray shading indicate amino acid similarities in 80%, 70%, and 60% of the sequences analyzed, respectively. Periods represent gaps, stars below sequences indicate confirmed, and circles postulated intron positions. The conserved acid-rich, Leu-rich1 and 2, RING1-IBR-RING2, and the coiled-coil regions are marked at the top of the alignment with solid and dotted lines, respectively. The CKII phosphorylation signatures, the nuclear localization signal (PAT7), and the conserved His (H) of RING2 are indicated above the alignment. The small letters in the *AtARI4* protein sequence mark the putative amino acids of the putative first and third intron.

**Table 1.** *ARI* gene family members in *Arabidopsis*

The numbering follows the phylogenetic analysis.

Family Name <sup>a</sup>	Gene Structure	Open Reading Frame Name (AGI)		ESTs/Full-Length cDNA TC Sequences (TIGR)	Protein	Amino Acids/kD	Nuclear Localization <sup>b</sup>
		New	Old				
Subgroup A							
<i>AtARI1</i>		At4g34370	F10M10.140	6 ESTs AY050864 AY117264 TC129475	AAK92801	597/68.8	60.9
<i>AtARI2</i>		At2g16090	F7H1.11	3 ESTs AY101519 AY065022 TC147587	AAL57664	593/67.8	52.2
<i>AtARI3</i>		At3g27710	MGF10.11	AY072185 AY096743	BAB02695 AAM20377	537/62.2	52.2
<i>AtARI4<sup>c</sup></i>		At3g27720	MGF10.12		Pseudogene	502/58.5	
Subgroup B							
<i>AtARI5</i>		At1g05890	T20M3.16	AY120783	AAM53341	552/62.5	73.9
<i>AtARI6</i>		At1g63760	F24D7.5		Pseudogene	551/62.6	
<i>AtARI7<sup>c</sup></i>		At2g31510	T9H9.3	AV562500 AFL19/07/H23 RAFL09/44/E08		562/64.0	73.9
<i>AtARI8</i>		At1g65430	T8F5.21	31 ESTs TC133535 TC133534 AY062808	AAL32886	567/64.8	65.2
<i>AtARI9</i>		At2g31770	F20M17.19		AAD32295	543/62.9	69.6
<i>AtARI10</i>		At2g31760	F20M17.20		AAD32296	514/59.4	21.7
<i>AtARI11</i>		At2g31780	F20M17.18		AAD32294	542/62.4	73.9
<i>AtARI12<sup>c</sup></i>		At1g05880	T20M3.15			496/56.4	65.2
Subgroup C							
<i>AtARI13</i>		At5g63750	MBK5.23		BAB10468	536/61.6	73.9
<i>AtARI14</i>		At5g63730	MBK5.21		BAB10466	506/59.0	69.6
<i>AtARI15<sup>d</sup></i>		At5g63760	MBK5.24	6 ESTs TC146548		452/51.8	21.7
<i>AtARI16</i>		At5g08730	T2K12.80		CAC35878	500/56.6	65.2

<sup>a</sup>The numbering of the *AtARI* genes follows the pattern derived from the phylogenetic analysis. <sup>b</sup>Result of the k-nearest neighbors prediction after Horton and Nakai (1997). <sup>c</sup>Reannotated by the authors. <sup>d</sup>Alternatively spliced 3'-UTR intron.

amino acid similarity, whereas the nucleotide and amino acid similarity within groups is higher (Fig. 2).

To determine the evolutionary relationship among the *ARI* proteins, phylogenetic analyses were performed with protein sequences of plant, algae, yeast, invertebrate, and vertebrate origin. Because full-length cDNA information was not available for all organisms, we also constructed the phylogram separately with the Leu-rich1 and -2 and the RING1-IBR-RING2 domain (Fig. 3). In total, we calculated four multiple sequences alignments and neighbor-joining trees with ClustalX v1.8 (Saitou and Nei, 1987). For the RING1-IBR-RING2 domain, we used human and rat PARKIN proteins as the outgroups. The tree topology of the different protein domains was similar to that of the full-length proteins and was largely supported by high bootstrap values (Fig. 3).

The phylogenetic analysis also divides the *AtARI* gene family into three subgroups. Group A radiates closer to *ARI* proteins of yeast, invertebrates, and vertebrates. This suggests that group A includes the closest homologs to the yeast and animal *ARI* proteins. In addition to *AtARI5/7/8/9/10/11/12*, group B also contains *ARI* proteins of barley (*Hv*), rice (*Os*), and *C. reinhardtii* (*Cr*). No other plant proteins cluster into group C with *AtARI13/14/15/16*. Thus, the current lack of *ARI* sequences from other plant families including Brassicaceae is insufficient to determine whether group C proteins are specific to *Arabidopsis* at the level of family, class, or phylum. The findings that only plant *ARI* proteins are present in groups B and C and that several *ARI* proteins of rice group into clade A and B support the existence of plant-specific *ARI* genes,

ARI2	ARI3	ARI4	ARI5	ARI6	ARI7	ARI8	ARI9	ARI10	ARI11	ARI12	ARI13	ARI14	ARI15	ARI16	
81.3/77.9	76.2/68.0	73.8/65.1	43.8/31.7	39.5/25.8	43.8/31.0	42.1/29.2	45.0/31.0	43.8/31.6	43.5/30.5	40.1/28.1	32.2/19.9	35.0/22.3	31.1/19.6	33.9/22.0	ARI1
***	74.6/66.2	72.3/64.4	44.9/31.4	39.0/26.3	44.4/30.5	42.6/28.0	44.0/31.5	43.5/31.7	43.6/31.6	40.4/28.9	31.7/19.4	35.8/23.0	29.8/19.5	33.6/22.2	ARI2
	***	86.8/82.0	43.2/32.0	36.8/27.1	43.8/31.1	42.6/30.0	42.9/31.2	42.2/30.9	43.1/30.3	38.4/26.4	32.5/20.4	35.9/23.5	30.4/20.0	33.7/23.5	ARI3
		***	43.0/30.5	36.7/24.2	44.3/30.1	42.2/30.1	42.8/31.1	41.9/30.3	43.0/30.0	38.6/26.3	32.3/19.6	36.4/23.0	30.1/19.8	33.9/22.6	ARI4
			***	90.7/86.0	80.8/77.2	69.7/66.9	66.1/58.4	66.1/58.7	66.5/58.1	56.8/44.0	37.5/25.6	39.9/28.1	34.3/25.4	40.4/28.7	ARI5
				***	78.8/70.8	63.4/52.5	58.4/48.7	63.2/53.0	65.0/55.1	52.2/38.1	33.3/22.5	35.0/22.5	23.3/16.1	34.9/24.2	ARI6
					***	70.5/69.6	64.9/56.8	65.4/57.1	65.5/56.8	57.0/44.5	36.8/25.4	40.1/29.1	33.4/24.7	39.8/17.7	ARI7
						***	59.0/53.8	58.9/52.8	59.4/54.2	52.5/42.0	36.3/26.3	39.2/29.7	35.1/25.5	39.5/28.9	ARI8
							***	86.5/81.0	78.7/71.4	52.2/39.9	39.1/28.2	41.8/30.2	37.6/28.2	41.5/30.4	ARI9
								***	81.8/74.9	52.5/40.4	40.8/30.3	44.0/33.1	39.3/30.5	43.8/32.4	ARI10
									***	52.0/39.6	37.6/26.6	40.4/29.6	35.3/25.9	40.8/30.5	ARI11
										***	35.8/25.0	39.9/27.3	33.3/22.5	39.5/28.7	ARI12
											***	74.9/61.1	67.0/56.1	60.0/48.5	ARI13
												***	64.8/53.4	61.3/49.9	ARI14
													***	56.7/48.7	ARI15

**Figure 2.** Degree of nucleotide and amino acid similarities between the AtARI family members. Values indicate the percentage of similarity obtained by pair wise comparisons between cDNA (left) and protein sequences (right). The similarities of AtARI members of a particular subgroup are marked light gray and of clustered genes dark gray.

which diverged before the split between mono- and dicotyledons.

### The Distinct ARI Gene Structures

Predicting the sites of splicing correctly is still a challenge for genome annotators. To get a reliable annotation, we combined computational analyses with experimental approaches and confirmed the exon/intron borders by either sequencing reverse transcriptase (RT)-PCR products or by comparing our proposed splicing differences to full-length cDNAs or to the tentative consensus (TC) sequences of TIGR (Seki et al., 1998, 2002; Quackenbush et al., 2000; Table I). With this combined approach, we identified the AtARI genes unambiguously. We stress that one-half of these had previously escaped identification because of erroneous splicing predictions.

No expressed sequence tags (ESTs) have been deposited for AtARI11/13/14 in the publicly available databases, and we were unable to detect expression. Therefore, we predicted their exon/intron splice sites by GENESCAN analyses and based on the most favorable similarity to the characteristic AtARI protein structure.

Despite the conserved nucleotide and amino acid sequence, the number of exons varies between one (AtARI3/6) and 15 (AtARI5/7/8; Figs. 1 and 2). Group C members (AtARI13-16) have the most uniform exon/intron structure with a single intron in the coding region of the RING2 domain (Fig. 1; Table I).

Group B members (AtARI5-12) exhibit the greatest differences. This group includes an intronless pseudogene (AtARI6), and members with the highest number of introns (AtARI5/7/8/12) share the same exon/intron structure (Fig. 1; Table I). One exception is AtARI12, which lacks the first intron and has displaced the seventh and ninth introns (Fig. 1).

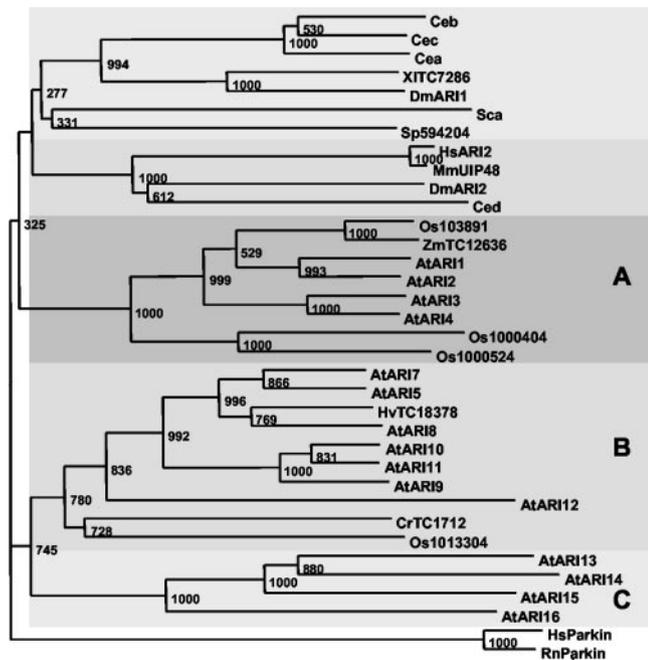
AtARI9/10/11 have only two introns at the same position where AtARI12 and AtARI5/7/8 have their 11th/12th and 12th/13th, respectively. On the basis of protein similarity and conserved exon/intron splice sites, we propose that the second intron of AtARI11 has shifted (Fig. 1).

The group A members do not share any introns with the other classes and can be divided into two subgroups. One subgroup has no introns (AtARI3) or has predicted introns at questionable positions (AtARI4). The other two genes AtARI1 and AtARI2 share their first four introns but not the C-terminal one(s).

### Evidence for AtARI Genes Generated by Retroposition

Retroposition has been suggested in animal systems as one means of generating intronless pseudogenes. Retroposition is thought mechanistically to occur by reverse transcription of mature mRNA after integration in the genome at sites unlinked to the active founder gene (Brosius, 1999). Thus, pseudogenes generated by retroposition may maintain some 5'- and 3'-untranslated region (UTR) sequence similarities to the mRNA of the founder gene. After integration, pseudogenes accumulate various defects, including missense or nonsense substitution, insertions, or deletions that lead to frame-shifts in the coding region.

As indicated above, one intronless pseudogene, AtARI6, was identified with high nucleotide similarity to AtARI5, a gene with 14 introns (Fig. 2; Table I). But there is at least one other gene, AtARI3, which does not contain an intron and exhibits a high-sequence similarity to the intron containing AtARI1. AtARI3 also exhibits a high-sequence similarity to AtARI4, which was predicted to carry three introns.



**Figure 3.** Conservation of the RING1-IBR-RING2 domain and phylogenetic analysis. Phylogenetic tree generated from alignment of the Cys-rich region from the ARI proteins of Arabidopsis (*AtARI1*–*AtARI15*), fruitfly (*DmARI1* and *DmARI2*), Brewer’s yeast (*Saccharomyces cerevisiae*; *Sca*), fission yeast (*Schizosaccharomyces pombe*; *Sp594204*) *Caenorhabditis elegans* (*Cea*, *Ceb*, *Cec*, and *Ced*), mouse (*Mus musculus*; *MmUIP48*), human (*HsARI2*), rice (*Oryza sativa*; *Os103891*, *Os1000404*, *Os1000524*, and *Os1013304*), *Xenopus laevis* (*XITC7286*), barley (*Hordeum vulgare*; *HvTC18378*), *Chlamydomonas reinhardtii* (*CrTC1712*), maize (*Zea mays*; *TC12636*), and two PARKIN proteins of human (*HsParkin*) and rat (*RnParkin*). The tree was calculated using the neighbor-joining method with 1,000 bootstrap replicates of the ClustalX 1.8. The two PARKIN proteins (*HsParkin* and *RnParkin*) were used as outgroup, and the phylogram was drawn using TreeView. The bootstrap values are placed at the nodes. Clades are shaded with different gray levels for better distinction.

Interestingly, one difference between *AtARI3* and *AtARI4* at the sequence level is an insertion of a single nucleotide that introduces a stop codon at codon 61 (Fig. 4). To avoid this stop codon, gene prediction programs propose a 52-bp intron. With the proposed second intron, an exon is generated that codes for a stretch of four and two modified versions of a nona peptide motif of EKTRKRKKE (Fig. 4). A deletion of seven nucleotides is at the 3’ end of the proposed third intron; this destroys the open reading frame coding for highly conserved amino acids between *AtARI3* and *AtARI4* (Fig. 4).

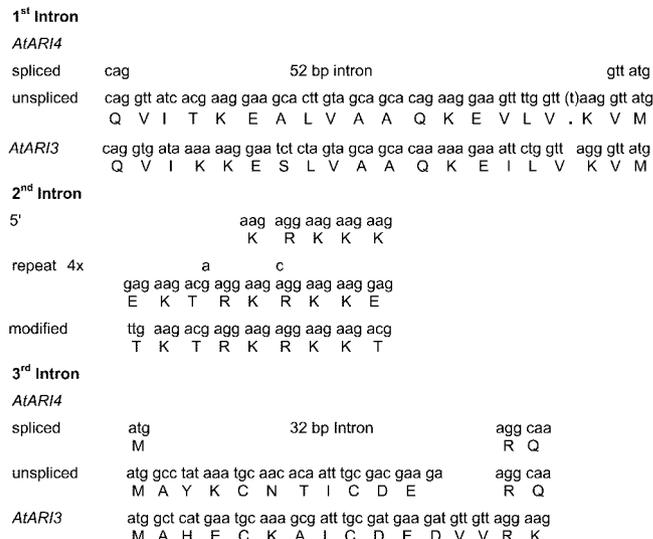
We tried to confirm the exon/intron positions for *AtARI4* but could only amplify fragments without introns. To prevent genomic contaminations we (a) pretreated RNA with DNase before cDNA synthesis and (b) confirmed that other PCR reactions with the same cDNA did not amplify genomic fragments. Thus, the gene either is not expressed or is expressed at very low expression levels irrespective of splicing.

The unspliced transcript codes for a truncated *AtARI4* protein of 60 amino acids.

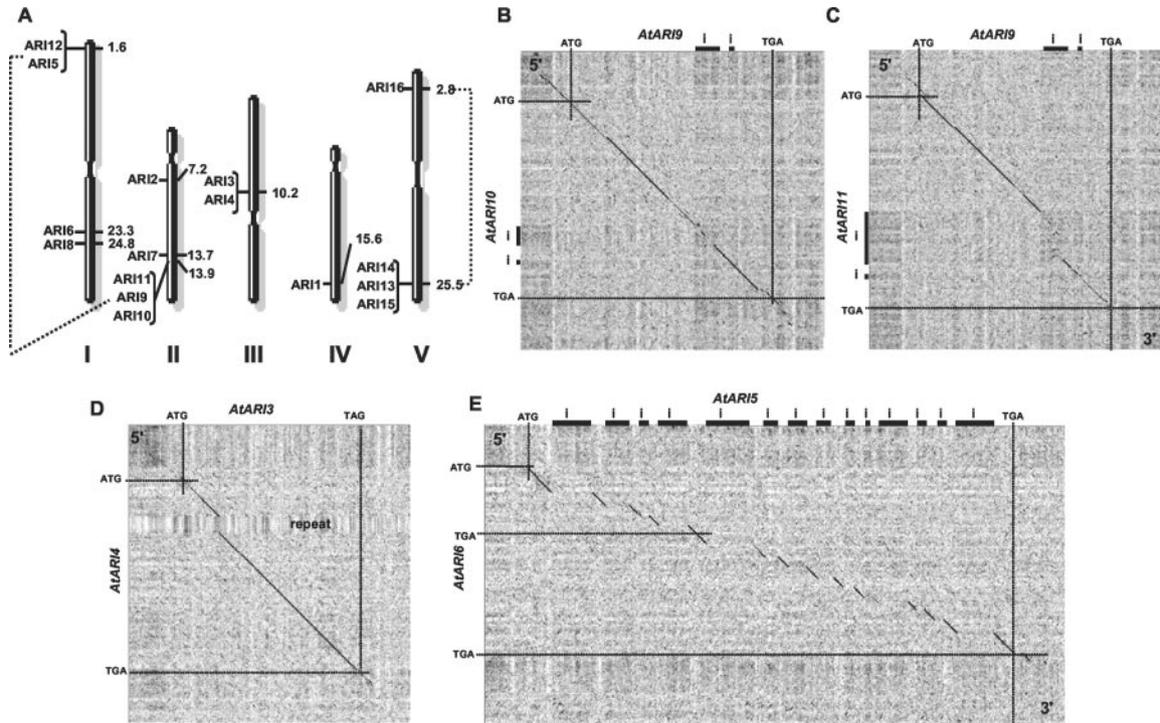
Moreover, the intronless nature of *AtARI3/4* points to a second retroposition event into a locus that supported the transcription of the retroposed gene. A local gene duplication (see below) produced a second copy, which subsequently accumulated mutations. Thus, retroposition might be one of the mechanisms for the expansion of the *AtARI* gene family.

### Chromosomal Distribution and Homology of the *AtARI* Genes

The observation that *AtARI* members differ in their gene structure points to different evolutionary histories of this gene family in Arabidopsis. Systematic analysis of the Arabidopsis genome suggests that at least four large duplications followed by chromosome fusions, inversions, translocations, gene losses, and smaller duplications occurred during evolution/speciation (Koch et al., 1999; Blanc et al., 2000; Vision et al., 2000). To determine whether *AtARI* genes are, in part, the result of genome duplication events, we analyzed the chromosomal distribution of the *AtARI* genes using the Genome and Redundancy Viewer at MIPS ([http://mips.gsf.de/proj/thal/db/gv/gv\\_frame.html](http://mips.gsf.de/proj/thal/db/gv/gv_frame.html)). The 16 *AtARI* genes are distributed on all five chromosomes at 10 loci that code for up to three genes (Fig. 5A). *AtARI5/12* and *AtARI7/9/10/11* are in duplicated blocks estimated to have originated 100 million years ago, long after the divergence of monocots and dicots (Vision et al., 2000). Two loci at the end of published duplication events on chromosome V, *AtARI16* and the cluster of *AtARI13/14/15*, might have originated by a second, younger (approximately 50 million years ago) duplication event.



**Figure 4.** Differences of the coding sequences and the predicted exon/intron borders between *AtARI3* and *AtARI4*.



**Figure 5.** Chromosomal distribution of and homologies between the *AtARI* gene family. A, The 16 *AtARI* genes are localized throughout the Arabidopsis genome as single genes or in clusters, which are specified with brackets. Gene positions on the physical map of AGI at TAIR are indicated in megabases for each gene. The five chromosomes are labeled by roman numerals, and the large genome duplications are represented by dotted lines. B to E, Dot-blot analysis of clustered (B–D) and potentially retroposited (E) *AtARI* genes was done using the identity matrix of the program Dotlet, using a sliding window of 25 nucleotides and gray scale settings of 37% to 78%. The black diagonals denote regions of high nucleotide identities. Gaps in the diagonals represent either mismatches or introns. B compares the genomic sequence between *AtARI9* and *AtARI10*, C between *AtARI9* and *AtARI11*, D between *AtARI3* and *AtARI4*, and E between the 14-intron-containing *AtARI5* and the pseudogene *AtARI6*. Note that the region of sequence identity between *AtARI9/10*, *AtARI9/11*, *AtARI3/4*, and *AtARI5/6* extends into the 5'- and/or 3'-UTR. Start and stop codons, introns, the orientation of the sequences (5'–3'), and special features as the repeat region of *AtARI4* are indicated.

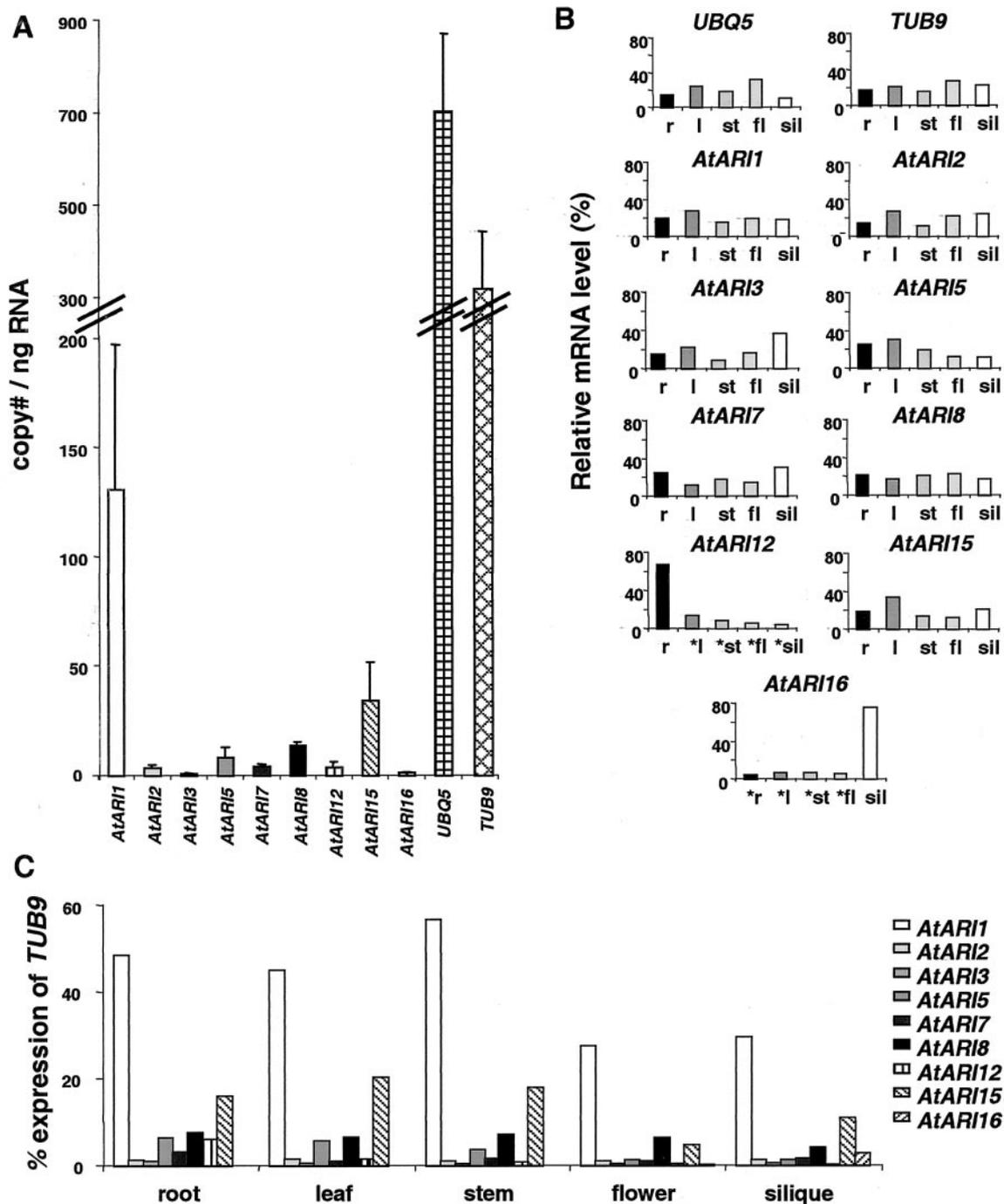
In addition, clustered genes such as *AtARI3/4* and *AtARI9/10/11* share 78% to 86% nucleotide and 71% to 82% amino acid identity (Fig. 2). These high sequence similarities imply that the clusters might have originated from recent smaller duplications, for example because of unequal crossing-over. This mechanism is supported not only by the sequence identity of the coding region but also extends to the 5'- and 3'-UTRs of *AtARI3/4* and *AtARI9/10* (Fig. 5, B and D). In addition, the second intron of *AtARI9/10/11* is highly homologous, whereas the first intron exhibits some similarities between *AtARI9/10* but is larger and has already diverged in *AtARI11* (Fig. 5, B and C).

A different picture emerges for the clusters *AtARI5/12* and *AtARI13/14/15*. Here, sequence identities drop to 56% and 75% at the cDNA and to 44% and 61% at the amino acid level, respectively (Fig. 2). One possibility to explain this divergence is that the *AtARI5/12* and *AtARI13/14/15* are older duplications that have already diversified. Interestingly, *AtARI5* exhibits a higher homology to other *AtARI* members outside the cluster and, in particular, to the unclustered pseudogene *AtARI6*, which does

not localize to a region of larger genome duplication (Figs. 2 and 5, A and E). Instead, *AtARI6* localizes close to two clusters of putative disease resistance protein genes. As pointed out earlier, the *AtARI6* pseudogene does not have any intron and exhibits high homology to *AtARI5*, indicating that it might have been formed by retroposition via a mRNA intermediate. In addition, *AtARI6* possesses features indicative of a relatively recent origin, being almost identical in the “coding” region and exhibiting homologies at the 3'-UTR (Fig. 5E).

#### *AtARI* Genes Are Differentially Expressed and Alternatively Spliced

ESTs or cDNAs are good preliminary sources for information about potential differences in expression, e.g. organ-specific expression or altered patterns of expression during development. Searching different databases, full-length cDNAs were identified for *AtARI1/2/3/5/8* and ESTs for *AtARI7/15* (Table I). For example, six ESTs are deposited for *AtARI1* and for *AtARI15* and 31 for *AtARI8*. No ESTs



**Figure 6.** Real-time PCR expression profiles of individual *AtARI* genes and the housekeeping genes *UBQ5* and *TUB9*. A, The absolute copy number of individual *AtARI*, *UBQ5*, and *TUB9* transcripts per nanogram total RNA of all organs analyzed (r, root; l, leaf; st, stem; fl, flower; sil, silique). B, The relative level of individual *AtARI*, *UBQ5*, and *TUB9* mRNAs in each organ as the percentage of the absolute copy number shown in A. Stars mark data where expression was close to the detection limit. C, The relative quantification of individual *AtARI* genes normalized with the housekeeping gene *TUB9*. Note that *AtARI12* and *AtARI16* exhibit the highest degree of organ specificity. At least three different RNA isolations were analyzed, and each cDNA was measured in triplicate. The expression of the *AtARI4/AtARI9/10* was below the limit for a reliable quantification with real-time PCR. *AtARI11/13/14* could not be amplified at all and are not included.

were identified for *AtARI4/9/10/11/12/13/14/16* or for the pseudogene *AtARI6* (Table I).

To determine the organ-specific expression pattern of each *AtARI* member, real-time PCR was performed

on RNA isolated from roots, leaves, stems, flowers, and green siliques. In addition, we measured expression levels for two housekeeping genes, *UBIQUITIN EXTENSION PROTEIN5 (UBQ5)* and  $\beta$ -*TUBULIN9*

(*TUB9*), which are genes transcribed at intermediate levels (Fig. 6A). Surprisingly, *UBQ5* is expressed at significantly lower levels in siliques, whereas *TUB9* is expressed more uniformly in the organs tested (Fig. 6B).

Although low amounts of amplification products were generated of *AtARI4/9/10* cDNAs, they were below the detection limit of the real-time PCR which is about 5 cDNA copies  $\text{ng}^{-1}$  RNA. No cDNA could be amplified for *AtARI11/13/14*. For these genes, ESTs have also not been deposited in public databases. Thus, either these *AtARI* members are expressed at extremely low levels or in specific tissues and at distinct developmental stages that were underrepresented in our organ samples.

Organ-specific transcription, i.e. in roots and green siliques, was detected for *AtARI12* and *AtARI16*, respectively. In contrast, *AtARI1/2/3/7/8/15* are expressed to a similar level in all organs (Fig. 6, B and C). The highest expression level was detected for *AtARI1* followed by *AtARI15* (Fig. 6A). The other *AtARI* members are expressed at low levels with less than 20 cDNA copies  $\text{ng}^{-1}$  RNA. The expression of *AtARI3* and *AtARI16* was close to the detection limit. The results obtained by measuring the absolute levels of transcript copy number were consistent with the relative quantification based on the ratio of *AtARI* transcript number to the transcript number of the housekeeping genes *TUB9* and *UBQ5* (Fig. 6, B and C).

Alternative splicing is one possibility to increase the number of proteins and their translation efficiency. We detected such an alternative splicing event for *AtARI15* because two different mRNA forms exist at the 3'-UTR of the *AtARI15* (Fig. 7A). The smaller transcript lost the 3'-UTR intron during pre-mRNA splicing. The larger transcript retains this intron (Fig. 7B). We consider genomic contamination unlikely to account for this observation because we have rigorously attempted to minimize genomic contamination. In addition, and more importantly, our

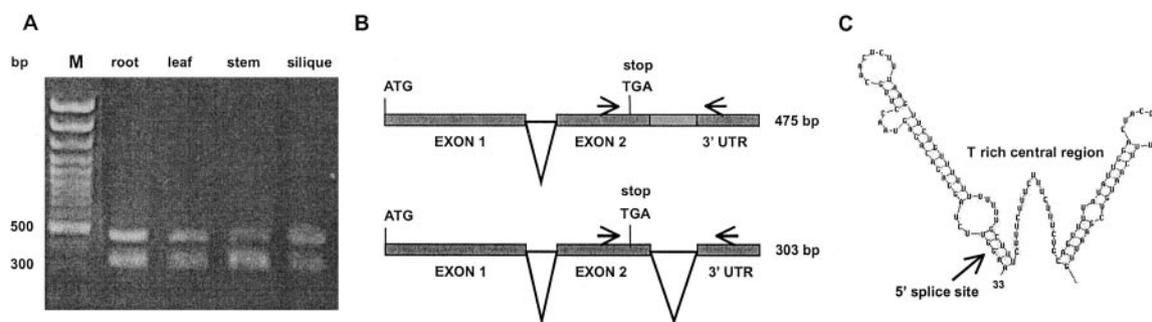
finding of alternative splicing is further substantiated by the existence of a cDNA clone (GenBank accession no. BE520707) that maintained the 3'-UTR intron. This intron can form two hairpin structures at its 5' and 3' end that flank a highly T-rich central sequence (Fig. 7C) and thus may have functional relevance for the stability and/or localization of the mRNA or the translational efficiency.

## DISCUSSION

### Retroposition and Large and Small Duplications Are Three Modes of *AtARI* Gene Family Expansion

The completion of the Arabidopsis genome allows the analysis of gene families, their evolution, functional redundancy, and divergence. It is a striking feature of the Arabidopsis genome that the proportion of proteins belonging to families of more than five members is higher (37.4%) than in other model organisms as fruitfly (12.1%) or *C. elegans* (24.0%; AGI, 2000). The *AtARI* gene family also conforms to this empirical observation. Although our database searches identified 16 *AtARI* genes in Arabidopsis, only two *ARI* genes are present in fruitfly and four in *C. elegans*.

Vision et al. (2000) explained this pronounced redundancy in the Arabidopsis genome through at least four different large-scale duplication events. In fact, at least two *AtARI* gene clusters are located in such large duplicated stretches between the top and the bottom of chromosomes 1 and 2, respectively (Fig. 5A). In addition, the sequence and structural homology of *AtARI13/14/15/16* and their vicinity to an intrachromosomal duplication on chromosome 5 suggest that they also originated through one of the large-scale duplications. Thus, part of the expansion of the *AtARI* gene family can be traced back to the large-scale duplication events that have already been described.



**Figure 7.** Alternative splicing of a 3'-UTR intron of *AtARI15*. A, The two real-time PCR amplicons of *AtARI15* were separated on a 1.5% (w/v) agarose gel and stained with ethidium bromide. Sequencing of the two amplicons revealed that the 303-bp fragment is spliced, whereas the 475-bp fragment corresponds to the unspliced form of the *AtARI15* transcript. Lane M was loaded with a 100-bp DNA size marker (MBI Fermentas, Vilnius, Lithuania). B, Schematic diagram of *AtARI15* with the alternatively processed 3'-UTR intron. C, The 3'-UTR of the *AtARI15* mRNA can form two potential hairpin structures. The  $\Delta G$  value of the hairpin structure of  $-19.9 \text{ kcal mol}^{-1}$  was calculated according to Mathews et al. (1999). The hairpin structures start 33 nucleotides downstream of the stop codon, include 117 nucleotides, and stop 30 nucleotides before the 3' site of the alternatively spliced intron.

**Table II.** Primer sequences and amplicon sizes for real-time PCR analysis and exon/intron border confirmation

Gene	AGI Gene Name	Primer Pairs	Fragment Length	
			Genomic	cDNA
				<i>bp</i>
<i>UBQ5</i>	At3g62250	5'-AACCTTGAGGTTGAATCATCC-3' 5'-GTCCTTCTTCTGGTAAACGT-3'	426	426
<i>TUB9</i>	At4g20890	5'-GTACCTGAAGCTTGCTAATCTA-3' 5'-GTTCTGGACGTTTCATCATCTGTTTC-3'	470	360
<i>AtARI 1</i>	At4g34369	5'-CCCACCTCCCTTGCTCTTG-3' 5'-TTCGCTCTTCCATCTGCTTCT-3'	410	240
<i>AtARI 2</i>	At2g16090	5'-AATGGTGAAGTGGTGTCCGAGT-3' 5'-ACCCGAGATTCTAGCCCAAGTG-3'	425	340
<i>AtARI 3</i>	At3g27710	5'-CATTGCGGGAGTGCGATACG-3' 5'-CAAAAATGCTGCCACACTTACAA-3'	280	280
<i>AtARI 4</i>	At3g27720	5'-GGAAGGAACATTTTACCGTGAGG-3' 5'-GAACATTCCACCTCATCAACATCG-3'	255	128 predicted
<i>AtARI 5</i>	At1g05890	5'-AACGCGGAAGAGTCCGATTG-3' 5'-ACCCTTCTGTCTGCTGCTATT-3'	1100	390
<i>AtARI 7</i>	At2g31510	5'-AGAAAAGTTAACCGGCGAATGTG-3' 5'-GCATATGGATACTACCTACCTGAGC-3'	650	285
<i>AtARI 8</i>	At1g65430	5'-GCAAGCCGCATGTGAGTCTCA-3' 5'-ATGGATTCTGAAGAAGACATGC-3'	1,950	400
<i>AtARI 9</i>	At2g31770	5'-AACGCCGAGAGATGGCAAAGAAGTCT-3' 5'-TCCTGCCAGTTTTGTGCGGAATTCA-3'	430	430
<i>AtARI 10</i>	At2g31760	5'-GCAAGCTCTATCGAGAAGAGC-3' 5'-GATAGGACGCTTGCCTTGCATTCAGGAC-3'	680	415
<i>AtARI 11</i>	At2g31780	5'-TGCTCGCTTGGCGAGATTATC-3' 5'-AGCGCAACAAGCGAGTATTGTC-3'	450	No expression
<i>AtARI 12</i>	At1g05880	5'-CTACAGTCCTTCGAGCAAACAC-3' 5'-ACGTATCGCAGCTTAAGTTCATCCT-3'	400	230
		5'-TAGGAACTGCCAAGGAGGAGC-3' 5'-GTAGAACTTAAGCTGCGATACGT-3'		For sequencing
		5'-TGCGCCATTGATCTTAGTCCA-3' 5'-ATCCTTTGACGGTTCCTCCTCG-3'		
		5'-AAGGTAAGCATTTTCATCTCAGTG-3' 5'-GCCACACTTCTGCTTCTCAC-3'		
<i>AtARI 13</i>	At5g63750	5'-CCGGTCTTGACTTGTGCCT-3' 5'-CTCTCGACATCTCATACTCTGC-3'	465	No expression
<i>AtARI 14</i>	At5g63730	5'-CGACAAAACCAAGACAGTGAC-3' 5'-CACGTTCCAGCGATCCTCAC-3'	300	No expression
<i>AtARI 15</i>	At5g63760	5'-GATTGATGCCAATACCAAGCCT-3' 5'-CCTTAGTACCTTAACGTCTTGC-3'	450	340
		5'-CGAGAAGGTTTGCCGAGGTA-3' 5'-CAGGGCTTTAGTTTCTGAGA-3'		For sequencing
<i>AtARI 16</i>	At5g08730	5'-ACGTACCTGTTGGATCAGTCA-3' 5'-GATTCAGGTGGAGAATCTGTG-3'	300	240

More than one-half of the *AtARI* genes cluster as tandems with up to three copies. This observation is again consistent with published data showing that 17% of all Arabidopsis genes are organized in tandem arrays (AGI, 2000). Most of the clustered *AtARI* genes exhibit a high sequence similarity and conserved intron positions. This indicates that they may have originated recently. The fact that the sequence identity extends to non-coding regions points to a duplication through an unequal crossing-over mechanism (Fig. 5, B–D).

However, duplication events do not suffice to explain the abundance of the *AtARI* family members. Interestingly we identified three intronless *AtARI*

genes that may have derived from a mRNA-mediated mechanism. One of them, *AtARI6*, is annotated as a pseudogene with short 5'- and 3'-UTR sequences of 380 and 179 bp, respectively, and the putative open reading frame codes for a premature stop codon. *AtARI6* shares 90.7% nucleotide identity to *AtARI5* cDNA, indicating that this *AtARI* member may be the founder gene. *AtARI3/4* may have similarly arisen from reinsertion of processed mRNA into the genome. Although, the annotation of *AtARI4* in the public databases proposes three introns, our analyses do not support these predictions. Moreover, we propose that *AtARI4* is another pseudogene and that the insertion of a single nucleotide introduced a stop

codon at the 5' end of the gene. In addition, *AtARI4* is interrupted by a large repeat located after the first stop codon. Thus, one plausible scenario for the development of the *AtARI3/4* gene cluster would involve a retroposition of a founder gene, probably *AtARI1*, through a processed mRNA intermediate, a duplication through unequal crossing-over leading to the tandemly clustered *AtARI3/4*, and the accumulation of mutations in *AtARI4*. Whereas *AtARI3* is expressed and retains its function, *AtARI4* is a pseudogene expressed at very low levels. Similar scenarios are suggested to be a major driving force in autosomal gene decay in fruitfly (Steinemann and Steinemann, 1997). Finally, there is increasing evidence that retroposition also contributes to the emergence of new functional and nonfunctional gene copies in primates and in other mammalian genomes (for review, see Brosius, 1999). Thus, it is reasonable to speculate that retroposition may have had an important role in Arabidopsis as well. Obviously, retroposition provides a mechanism for the expansion of gene families. We are aware that our analysis may not show the complete picture because only the accession Columbia was used and because there is evidence that the pattern of duplications can be different in other accessions (M.-T. Hauser, unpublished data).

#### Phylogenetic Analysis Points to Plant-Specific Subgroups of ARI Proteins

Gene evolution and divergence can be inferred by comparing the degree of conservation of intron positions, protein structures, and domains and the homology of coding DNA and encoded amino acid sequences. We have used these criteria to further subdivide the *AtARI* family into three subgroups. By including ESTs and full-length sequences of algae, yeast, plants, invertebrates, and vertebrates, we inferred the orthology of the Arabidopsis paralogs. The phylogenetic analyses were done with full-length protein alignments and of conserved domains. Regardless of the size of the compared sequences, all calculations revealed that vertebrate and invertebrate sequences form two separate clades each with one of the fruitfly DmARI1 and DmARI2 paralogs. Three clades contain only plant sequences, and the *AtARI* subgroups match to these clades. In addition, the existence of *ARI* orthologous EST and gene sequences from rice, maize, and barley suggests that at least clade A and B arose before the split of the monocot and dicot phyla. Furthermore, *ARI* ESTs were identified from cauliflower (*Brassica oleracea*), soybean (*Glycine max*), potato (*Solanum tuberosum*), tomato (*Lycopersicon esculentum*), wheat (*Triticum aestivum*), and poplar, but the sequences were too small to be included in our analysis (data not shown). We identified four *ARI* genes in rice. Three cluster in clade A and one in clade B, indicating that different *ARI*

paralogs of this gene family exist also in other phyla and have diverged to a higher degree in the plant kingdom.

The proximity of plant clade A to the "animal" clades suggests that class A members may be their orthologs. It remains to be demonstrated whether their functions and the function of the paralogous classes are similar and to what extent they have diverged.

#### Protein Domains of ARI-Like Proteins: First Indications for Function and Regulation?

One objective of these detailed analyses was to gather information that allows the prediction of the biochemical or biological functions of the *AtARI* gene family. To investigate the biological significance of the *AtARI* gene family, we identified conserved domains, subcellular localization signatures, and other features that are the basis for further experimental investigations on the *AtARI* proteins.

It has been demonstrated that the most characteristic signature of the ARI proteins, the RING1-IBR-RING2 domain, interacts specifically with the E2 UBCs UbcD10, UbcM4, UbcH7, and UbcH8 of fruitfly, mouse, and man, respectively (Martinez-Noel et al., 1999; Moynihan et al., 1999; Aguilera et al., 2000). These data suggest that the ARI proteins are functionally linked to the ubiquitin/proteasome pathway. Interaction studies of the human ARI protein (HHARI) revealed that the RING1 and part of the IBR motif is necessary for the binding to UbcH7 (Moynihan et al., 1999). The importance of the RING1 domain for interaction with specific UBCs was confirmed with DmARI1 and DmARI2 of fruitfly and UbcD10 but also with the mouse homolog UbcM4 (Aguilera et al., 2000). Thus, the interaction of ARI proteins with specific UBCs is conserved between insects and mammals. Ardley et al. (2001) further showed that the distance between RING1 and the IBR is crucial for maintaining the binding to UBCs. The RING1 domain and the distance to the IBR is highly conserved in the *AtARI* protein family with the exception of some subgroup C members. Thus, it is intriguing to argue that the *AtARI* proteins might interact with specific UBCs. At least 37 potential UBC-like genes were recently identified in the Arabidopsis genome that are candidates for interaction with the *AtARI* proteins (Bachmair et al., 2001).

Given that the ARI proteins and PARKIN have the RING1-IBR-RING2 domain in common, we anticipate that the proteins share functional similarities. It has recently been shown that PARKIN acts as an E2-dependent ubiquitin-protein ligase (Shimura et al., 2000; Zhang et al., 2000; Rankin et al., 2001). However, we are aware that this analogy may be misleading because Zhang et al. (2000) mapped the interaction domain of PARKIN with UbcH8 to RING2. Second, Shimura et al. (2000) suggested that

the ubiquitin-like (Ubl) domain of PARKIN functions as a module necessary for binding the ubiquitinated proteins. Because ARI proteins do not have a Ubl domain, the binding domain for the ubiquitinated substrates may be different in PARKIN. Third, Zhang et al. (2000) identified CDCrel-1 to be a member of the septin family as a target and showed that PARKIN ubiquitinates and regulates CDCrel-1 degradation. Septin proteins are well-conserved small GTP-binding proteins of animals and fungi. Members of this family are involved in vesicle transport or fusion and cell division. In the Arabidopsis genome, they have not been annotated, although several small GTP-binding proteins with weak similarities are present (M.-T. Hauser, unpublished data).

PARKIN ubiquitinates itself and thereby promotes its own degradation (Zhang et al., 2000). On the basis of this precedent, we searched for degradation signals. For AtARI1/2/3/5/7/11, potential PEST sequences were identified at the N terminus and would support the hypothesis that ARI proteins might be short-lived. Other indications for potential mechanisms of regulation are the presence of several highly conserved consensus sites for CKII-dependent phosphorylation in the acidic N terminus and in the two Leu-rich domains.

Our results provide a framework for future analyses that focus on demonstrating that ARI proteins act as E3 ubiquitin ligases or part of E3 ubiquitin ligase complexes and that investigate their regulation by phosphorylation and degradation. Further objectives lie in the identification of the interacting E2 UBC and of their substrates, the degradation of which may be promoted by the AtARIs.

## MATERIALS AND METHODS

### Sequence Analysis and Bioinformatic Methods

Sequence information of genes, proteins, ESTs, and cDNAs was retrieved by searching public databases with the BLAST algorithm (Altschul and Lipman, 1990; Altschul et al., 1997) at EMBL-European Bioinformatics Institute (<http://www.ebi.ac.uk/index.html>), National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>), TAIR (<http://www.ebi.ac.uk/index.html>), TIGR (<http://www.tigr.org/>), and MIPS (<http://mips.gsf.de/>; Schoof et al., 2002). Exon/intron splice sites were analyzed and confirmed by comparing the genomic with EST and cDNA sequences, by RT-PCR and sequencing (see below), and with GENSCAN (<http://genes.mit.edu/GENSCAN.html>; Burge and Karlin, 1997, 1998). 5'- and 3'-UTRs were analyzed using PLACE (<http://www.dna.affrc.go.jp/htdocs/PLACE/>; Higo et al., 1999). For the secondary structure prediction of the alternative spliced intron, we used the MFOLD program of Mathews et al. (1999) (<http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi>). Dotlet (<http://www.isrec.isb-sib.ch/java/dotlet/Dotlet.html>) was used to identify similarities in introns and 5' and 3' sequences of AtARI genes.

The protein sequence analysis was done with PSORT (<http://psort.nibb.ac.jp/>), which identifies sorting signals and predicts the subcellular localization and coiled-coil structures. PROSITE ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_prosite.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html); Bairoch et al., 1997) was used to detect protein modification signatures as consensus sequences for protein kinases and structural domains as the RING-fingers. With the program PESTfind (<http://www.at.emblnet.org/embnet/tools/bio/PESTfind/>; Rogers et al., 1986; Rechsteiner and Rogers, 1996), we searched for proteo-

lytic signals. Most of these on-line analyses programs can be found under the ExPASy proteomics tools at <http://us.expasy.org/tools/>.

Nucleotide and protein sequences were aligned with the ClustalV method developed by Higgins and Sharp (1989) of the MegAlign sequence analysis software in the DNASTAR program package (DNASTAR, Inc., Madison, WI). The protein alignments were decorated with the GeneDoc (v2.5.000; <http://www.psc.edu/biomed/genedoc/>; Nicholas et al., 1997).

### Phylogenetic Analysis

For the phylogenetic reconstruction, we included ARI proteins of fruitfly (*Drosophila melanogaster*; DmARI1, Q94981; and DmARI2, AAF46823), Brewer's yeast (*Saccharomyces cerevisiae*; Sca, CAA82089), fission yeast (*Schizosaccharomyces pombe*; Sp594204, NP\_594204), *Caenorhabditis elegans* (Cea, AAB93643; Ceb, AAB93644; Cec, AAB93645; and Ced, AAB03121), mouse (*Mus musculus*; MmUIP48, AF124664), human (*Homo sapiens*; HsARI2, CAA10276), and rice (*Oryza sativa*; Os103891, AC103891; Os1000404, AAAA0100404; Os1000524, AAA01000524; and Os1013304, AAAA01013304). The ARI protein sequences of *Xenopus laevis* (XITC7286, TC72869), barley (*Hordeum vulgare*; HvTC18378, TC18378), *Chlamydomonas reinhardtii* (CrTC1712, TC1712), and maize (*Zea mays*; TC12636, TC126367) were translated cDNA sequences, retrieved from the TC sequences of the eukaryotic gene orthologs database at TIGR Gene Indices (<http://www.tigr.org/tdb/tgi/>). The RING1-IBR-RING2 domain of human (HsParkin, NP\_004553) and rat (RnParkin, NP\_064478) PARKIN proteins were included as outgroup.

Alignments for the phylogenetic analysis were calculated with ClustalX 1.8 (<http://inn-prot.weizmann.ac.il/software/ClustalX.html>; Thompson et al., 1997; Jeanmougin et al., 1998) and default parameter settings using full-length and domain sequences of the Leu-rich region 1 and 2 and the RING1-IBR-RING2 regions. The N terminus of the full-length multiple protein alignment was adjusted manually. Neighbor-joining tree calculations were done with ClustalX v1.8 and default setting with 1,000 bootstrap replicas (Saitou and Nei, 1987). The phylograms were drawn using TreeView (<http://darwin.zoology.gla.ac.uk/~rpage/treeview/>).

### Plant Material and Growth Conditions

Arabidopsis accession Columbia was cultivated on sterile nutrient agar plates and soil as previously described (Hauser and Benfey, 1994). Roots and leaves were harvested from 21-d-old seedlings cultured under slight agitation of 60 rpm at 22°C and 16-h light (80  $\mu\text{mol m}^{-2} \text{s}^{-1}$ )/8-h dark cycles in Murashige and Skoog (1962) liquid medium supplemented with 2% (w/v) Suc. Stems, flowers, and siliques were collected from plants growing on soil at 20°C under similar light conditions.

### Quantitative Real-Time PCR Expression Analysis and Splice Site Confirmation

Total RNA was isolated with RNeasy Plant Kit (Qiagen USA, Valencia, CA) or with TRI REAGENT (MRC, Cincinnati) according to the manufacturers' protocols. cDNA was synthesized essentially as described by Karsai et al. (2002). In brief, 2.5  $\mu\text{g}$  total RNA was pretreated with DNase I (Roche Diagnostics, Indianapolis) and reversed transcribed with Moloney murine leukemia virus reverse transcriptase (M-MuLV RT, Invitrogen, Carlsbad, CA) and d(T)<sub>18</sub>. The cDNA reaction was diluted 1:10 with water, and 1  $\mu\text{L}$  of the diluted cDNA (corresponding to 17 ng total RNA) was used as template for real-time PCR analysis with the Rotor-Gene 2000 (Corbett Research, Sydney) and our homemade SYBR green-I reaction mixture containing 1:20,000 diluted SYBR green-I (Roche Diagnostics), 200  $\mu\text{M}$  dNTPs, 10 mM Tris-HCl, pH 8.5, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 0.15% (v/v) Triton X-100, 5 pmol of each primer, and 0.5 units of *Taq* DNA polymerase (Qbiogene-Appigene, Heidelberg). Primers, AGI gene name, and the size of genomic and cDNA amplicons are listed in Table II. After a first denaturation step for 90 s/94°C, 35 cycles followed with 10 s/94°C, 5 s/55°C, and 10 s/72°C. The fluorescence was measured three times at the end of the extension step at 72°C, 81°C, and 84°C. The identities of the amplicons and the specificity of the reaction were verified by agarose gel electrophoresis and melting curve analyses. Absolute and relative copy numbers of individual mRNA species was calculated with standard curves of known molar concentrations for each of the 15 AtARI genes and the housekeeping genes *TUB9* and *UBQ5* (Karsai et al., 2002). At least three different RNA isolations and cDNA

syntheses were used for quantification and each cDNA was measured in triplicate.

Ambiguous exon/intron borders were confirmed by PCR (for primers see Table II) of cDNAs and sequencing with the BigDye terminator cycle sequencing chemistry (Applied Biosystems, Foster City, CA) and the ABI-Prism 310 genetic analyzer.

## ACKNOWLEDGMENTS

We thank E. Dornstauder and A. Pircher for their technical assistance. We thank A. Karsai for sequencing some of the RT-PCR products and F. Adhami for her help in the initial phase of this project. We are grateful to G. Muir, C. Schlötterer, and M. Freissmuth for critical reading the manuscript, and we thank the Salk Institute Genomic Analysis Laboratory for information on the full-length cDNA clones of the Salk, Stanford, Plant Gene Expression Center Consortium in collaboration with the RIKEN Genome Science Center and Ceres Inc.

Received August 9, 2002; returned for revision September 3, 2002; accepted September 26, 2002.

## LITERATURE CITED

- Aguilera M, Oliveros M, Martinez-Padron M, Barbas JA, Ferrus A (2000) Ariadne-1: A vital *Drosophila* gene is required in development and defines a new conserved family of ring-finger proteins. *Genetics* **155**: 1231–1244
- Altschul SF, Lipman DJ (1990) Protein database searches for multiple alignments. *Proc Natl Acad Sci USA* **87**: 5509–5513
- Altschul SF, Madden TL, Schaffer AA, Zhang J, 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
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815
- Ardley HC, Tan NG, Rose SA, Markham AF, Robinson PA (2001) Features of the parkin/ariadne-like ubiquitin ligase, HHARI, that regulate its interaction with the ubiquitin-conjugating enzyme, Ubc7. *J Biol Chem* **276**: 19640–19647
- Austin MJ, Muskett P, Kahn K, Feys BJ, Jones JD, Parker JE (2002) Regulatory role of SGT1 in early R gene-mediated plant defenses. *Science* **295**: 2077–2080
- Azevedo C, Sadanandom A, Kitagawa K, Freialdenhoven A, Shirasu K, Schulze-Lefert P (2002) The RAR1 interactor SGT1, an essential component of R gene-triggered disease resistance. *Science* **295**: 2073–2076
- Bachmair A, Novatchkova M, Potuschak T, Eisenhaber F (2001) Ubiquitylation in plants: a post-genomic look at a post-translational modification. *Trends Plant Sci* **6**: 463–470
- Bairoch A, Bucher P, Hofmann K (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res* **25**: 217–221
- Blanc G, Barakat A, Guyot R, Cooke R, Delseny M (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. *Plant Cell* **12**: 1093–1101
- Borden KL, Freemont PS (1996) The RING finger domain: a recent example of a sequence-structure family. *Curr Opin Struct Biol* **6**: 395–401
- Brosius J (1999) RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene* **238**: 115–134
- Burge CB, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**: 78–94
- Burge CB, Karlin S (1998) Finding the genes in genomic DNA. *Curr Opin Struct Biol* **8**: 346–354
- Dieterle M, Zhou YC, Schafer E, Funk M, Kretsch T (2001) EID1, an F-box protein involved in phytochrome A-specific light signaling. *Genes Dev* **15**: 939–944
- Freemont PS (2000) RING for destruction? *Curr Biol* **10**: R84–R87
- Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M (2001) Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature* **414**: 271–276
- Hauser MT, Benfey PN (1994) Genetic regulation of root expansion in *Arabidopsis thaliana*. In P Puigdomenech, G Coruzzi, eds, NATO-ASI Plant Molecular Biology Series, Vol 81. Springer-Verlag, New York, pp 31–40
- Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *Comput Appl Biosci* **5**: 151–153
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* **27**: 297–300
- Horton P, Nakai K (1997) Better prediction of protein cellular localization sites with the k nearest neighbors classifier. *Proc Int Conf Intell Syst Mol Biol* **5**: 147–152
- Jackson PK, Eldridge AG, Freed E, Furstenthal L, Hsu JY, Kaiser BK, Reimann JD (2000) The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. *Trends Cell Biol* **10**: 429–439
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Clustal X is a powerful multiple sequence alignment program with a user-friendly graphical interface. *Trends Biochem Sci* **23**: 403–405
- Joazeiro CA, Weissman AM (2000) RING finger proteins: mediators of ubiquitin ligase activity. *Cell* **102**: 549–552
- Karsai A, Müller S, Platz S, Hauser MT (2002) Evaluation of a homemade SYBR greenI reaction mixture for real-time PCR quantification of gene expression. *BioTechniques* **32**: 790–796
- Koch M, Bishop J, Mitchell-Olds T (1999) Molecular systematics of *Arabidopsis* and *Arabis*. *Plant Biol* **1**: 529–537
- Kosarev P, Mayer KF, Hardtke CS (2002) Evaluation and classification of RING-finger domains encoded by the *Arabidopsis* genome. *Genome Biol* **3**: RESEARCH0016.1–RESEARCH0016.12
- Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* **252**: 1162–1164
- Martinez-Noel G, Niedenthal R, Tamura T, Harbers K (1999) A family of structurally related RING finger proteins interacts specifically with the ubiquitin-conjugating enzyme UbcM4. *FEBS* **454**: 257–261
- Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911–940
- Moynihan TP, Ardley HC, Nuber U, Rose SA, Jones PF, Markham AF, Scheffner M, Robinson PA (1999) The ubiquitin-conjugating enzymes UbcH7 and UbcH8 interact with RING finger/IBR motif-containing domains of HHARI and H7-AP1. *J Biol Chem* **274**: 30963–30968
- Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol Plant* **15**: 473–497
- Nelson DC, Lasswell J, Rogg LE, Cohen MA, Bartel B (2000) FKF1, a clock-controlled gene that regulates the transition to flowering in *Arabidopsis*. *Cell* **101**: 331–340
- Nicholas KB, Nicholas HB Jr, Deerfield DW II (1997) GeneDoc: analysis and visualization of genetic variation. *EMBNEW*. **NEWS** **4**: 14
- Pinna LA (1990) Casein kinase 2: an “eminence grise” in cellular regulation? *Biochim Biophys Acta* **1054**: 267–284
- Potuschak T, Stary S, Schlögelhofer P, Becker F, Nejinikaia V, Bachmair A (1998) PRT1 of *Arabidopsis thaliana* encodes a component of plant N-end rule pathway. *Proc Natl Acad Sci USA* **95**: 7904–7908
- Quackenbush J, Liang F, Holt I, Perteau G, Upton J (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Res* **28**: 141–145
- Rankin CA, Joazeiro CA, Floor E, Hunter T (2001) E3 ubiquitin-protein ligase activity of Parkin is dependent on cooperative interaction of RING finger (TRIA) elements. *J Biomed Sci* **8**: 421–429
- Rechsteiner M, Rogers SW (1996) PEST sequences and regulation by proteolysis. *Trends Biochem Sci* **21**: 267–271
- Reinhardt A, Hubbard T (1998) Using neural networks for prediction of the subcellular location of proteins. *Nucleic Acids Res* **26**: 2230–2236
- Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234**: 364–368
- Saitou RK, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425
- Samach A, Klenz JE, Kohalmi SE, Risseuw E, Haughn GW, Crosby WL (1999) The UNUSUAL FLORAL ORGANS gene of *Arabidopsis thaliana* is an F-box protein required for normal patterning and growth in the floral meristem. *Plant J* **20**: 433–445
- Saurin AJ, Borden KL, Boddy MN, Freemont PS (1996) Does this have a familiar RING? *Trends Biochem Sci* **21**: 208–214
- 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
- Seki M, Carninci P, Nishiyama Y, Hayashizaki Y, Shinozaki K (1998) High-efficiency cloning of *Arabidopsis* full-length cDNA by biotinylated CAP trapper. *Plant J* **15**: 707–720

- Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y et al.** (2002) Functional annotation of a full-length *Arabidopsis* cDNA collection. *Science* **296**: 141–145
- Shen WH, Parmentier Y, Hellmann H, Lechner E, Dong A, Masson J, Granier F, Lepiniec L, Estelle M, Genschik P** (2002) Null mutation of AtCUL1 causes arrest in early embryogenesis in *Arabidopsis*. *Mol Biol Cell* **13**: 1916–1928
- Shimura H, Hattori N, Kubo S, Mizuno Y, Asakawa S, Minoshima S, Shimizu N, Iwai K, Chiba T, Tanaka K, Suzuki T** (2000) Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nat Genet* **25**: 302–305
- Somers DE, Schultz TF, Milnamow M, Kay SA** (2000) ZEITLUPE encodes a novel clock-associated PAS protein from *Arabidopsis*. *Cell* **101**: 319–329
- Steinemann M, Steinemann S** (1997) The enigma of Y chromosome degeneration: TRAM, a novel retrotransposon is preferentially located on the Neo-Y chromosome of *Drosophila miranda*. *Genetics* **145**: 261–266
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG** (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **24**: 4876–4882
- Vision TJ, Brown DG, Tanksley SD** (2000) The origins of genomic duplications in *Arabidopsis*. *Science* **290**: 2114–2117
- Woo HR, Chung KM, Park JH, Oh SA, Ahn T, Hong SH, Jang SK, Nam HG** (2001) ORE9, an F-box protein that regulates leaf senescence in *Arabidopsis*. *Plant Cell* **13**: 1779–1790
- Xie DX, Feys BF, James S, Nieto-Rostro M, Turner JG** (1998) COI1: an *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280**: 1091–1094
- Zhang Y, Gao J, Chung KK, Huang H, Dawson VL, Dawson TM** (2000) Parkin functions as an E2-dependent ubiquitin-protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. *Proc Natl Acad Sci USA* **97**: 13354–13359